Smoothness-Driven Consensus Based on Compact Representation for Robust Feature Matching

Aoxiang Fan, Xingyu Jiang, Yong Ma, Xiaoguang Mei, and Jiayi Ma

***Abstract*—For robust feature matching, a popular and particularly effective method is to recover smooth functions from the data to differentiate the true correspondences (inliers) from false correspondences (outliers). In the existing works, the well-established regularization theory has been extensively studied and exploited to estimate the functions while controlling its complexity to enforce the smoothness constraint, which has shown prominent advantages in this task. However, despite of the theoretical optimality properties, the high complexities in both time and space are induced and become the main obstacle of their application. In this paper, we propose a novel method for multivariate regression and point matching, which exploits the sparsity structure of smooth functions. Specifically, we use compact Fourier bases for constructing the function, which inherently allows a coarse-to-fine representation. The smoothness constraint can be explicitly imposed by adopting a few low-frequency bases for representation, resulting in reduced computational complexities of the induced multivariate regression algorithm. To cope with potential gross outliers, we formulate the learning problem into a Bayesian framework with latent variables indicating the inliers and outliers and a mixture model accounting for the distribution of data, where a fast Expectation-Maximization solution can be derived. Extensive experiments are conducted on synthetic data and real-world image matching and point set registration datasets, which demonstrate the advantages of our method against the current state-of-the-art methods in terms of both scalability and robustness.**

***Index Terms*—Feature matching, regularization, compact representation, outlier, mismatch removal.**

I. Introduction

IN computer vision, establishing reliable correspondences between two feature sets is a fundamental problem that typically arises from image matching or point set processing tasks [1]. It is the critical prerequisite in a wide spectrum of applications such as panoramic stitching [2], image and point set registration [3], 3D reconstruction [4], and simultaneous localization and mapping [5]. Traditional solutions for the matching problem include directly dealing two discrete point sets [3], [6], [7], which are very sophisticated but less attractive because of the high computational complexity. Fortunately, for the processed data such as images or point sets, the local features can be utilized and has been extensively studied to simplify the problem [8], [9]. By evaluating the similarity of the feature vectors (or descriptors) associated with each two points, most of the possible matches can be rejected, and only a limited number of matches are kept as the putative correspondence set. However, due to the inherent ambiguities of local feature representation, the putative correspondence set is typically contaminated by a large number of outliers. Consequently, the problem we are faced with boils down to identifying the inliers, or mismatch removal. In this paper, we focus on investigating such algorithms to filter out the outliers and (possibly) recover the underlying transformation model of inliers from the contaminated data.

The central issue for the robust matching task is the exploitation of geometric constraint. For rigidly moving objects, it has been revealed by the study of camera models that the geometric constraint relating two image scenes can be exactly modeled by a fundamental matrix, known as the epipolar geometry, with the Degree of Freedom (DoF) of 8 [10]. For certain special cases, the DoF can be even further reduced, with models such as homography and affine. This simple fact has been the inspiration of a large group of resampling methods. Examples include the long established RAndom SAmple Consensus (RANSAC) algorithm [11] and its numerous variants that cover all phases of the resampling scheme. Despite their successes, some fundamental setbacks exist in this framework. First, the required runtime grows exponentially with the outlier ratio increasing, making it impractical for severely contaminated data. Second, the parametric model has its own restrictions and cannot address more general scenarios, *e.g.*, non-rigid transformation.

The publication of the seminal work Vector Field Consensus (VFC) [12] has encouraged another line of work. The geometric constraint used is more general, *i.e.*, motion coherence or smoothness. The transformation of points is modeled by a more flexible vector field function, which admits a higher DoF. The smoothness constraint is imposed by using the well-established regularization theory [13], [14]. The VFC algorithm has been demonstrated to be a general philosophy to handle the robust matching task, applicable to both simple cases controlled by a low-DoF parametric model and complex cases involving a high-DoF non-rigid transformation. However, the main drawback of VFC is its high computational complexity, *i.e.* $O(N^3)$ in time and $O(N^2)$ in space, where N denotes the number of correspondences. In efforts to remedy this issue, FastVFC [12] and SparseVFC [15] have been proposed. Specifically, for FastVFC, a low-rank approximation of the kernel matrix is used, which reduces the complexity to $O(N)$ in time in the main iterations. However, the computation of the low-rank approximation still requires time in $O(N^3)$. For SparseVFC, a sparse random

This work was supported in part by the National Natural Science Foundation of China under Grant nos. 62061160370, 61903279 and 61773295, and in part by the Natural Science Fund of Hubei Province under Grant no. 2019CFA037. *(Corresponding author: Jiayi Ma.)*

The authors are with the Electronic Information School, Wuhan University, Wuhan, 430072, China (e-mail: fanaoxiang@whu.edu.cn, jiangx.y@whu.edu.cn, mayong@whu.edu.cn, meixiaoguang@gmail.com, jyma2010@gmail.com).

basis technique is used to compute the vector field function, which reduces the complexity to $O(N)$ in both space and time. However, this strategy lacks sufficient theoretical justifications and typically leads to numerical instability in practice. As will be shown, its computational complexity does not conform to $O(N)$ in general.

In this paper, analogous to SparseVFC, we also take advantage of the sparsity structure to develop a computationally efficient method. The difference is that we explore a different way to model the underlying transformation instead of it that the classical regularization theory has suggested. The main observation is that smooth functions actually possess a sparse structure in a different representation, which has not been exploited in classical regularization theory. Different from the Reproducing Kernel Hilbert Space (RKHS), we use Fourier bases to construct the function space, which allows a coarse-to-fine representation directly linked to the concept of frequency, or smoothness. As we will demonstrate, this leads to a much more efficient algorithm that recovers smooth functions from contaminated data, without sacrificing the robustness.

Our contributions in this paper include the following three aspects. Firstly, we introduce an alternative of the classical regularization theory to learn a smooth function from sparse samples. The method exploits the sparsity structure and admits a compact Fourier representation to model the function to be learnt, resulting in computationally efficient algorithms without sacrificing the accuracy. Secondly, we incorporate the compact representation technique into a Bayesian framework to cope with potential gross outliers in the sample set for robust image feature matching, which significantly generalizes the practicability of our method. Thirdly, we demonstrate the superiority of our method in terms of both efficiency and robustness in various tasks, including multivariate regression, robust feature matching and point set registration.

The rest of the paper is organized as follows. Sec. II describes background material and related work. Sec. III introduces the classical regularization theory and describes the proposed method to learn a smooth multivariate function from sparse data. In Sec. IV, we consider the image feature matching task and discuss an outlier-robust algorithm with our compact representation technique to address it. Sec. V illustrates the experimental results on the tasks of multivariate regression, robust feature matching and point set registration to demonstrate our method. The concluding remarks are presented in Sec. VI.

II. Related Work

In this section, we briefly review the background literature that is closely related to our work. This includes the methods to create a putative correspondence set for matching, and robust matching methods with different geometric constraints to remove outliers. Some solutions that directly establish correspondences from two sparse point sets are also discussed.

For data instances such as images and point sets, local features are of great interest and have been extensively studied for matching. Usually at first interest points are detected, and then a feature vector, *i.e.* descriptor, is generated for each point for local feature representation. For image matching, this includes some long known methods such as Scale-Invariant Feature Transform (SIFT) [8], Speeded-Up Robust Features (SURF) [8], Oriented FAST and Rotated Brief (ORB) [16], as well as some recently developed methods using deep learning technique [17], [18]. The situation is similar for point sets, both hand-crafted methods, such as Shape Context [9] and Fast Point Feature Histogram (FPFH) [19], and deep learning-based solutions have been well-studied [19], [20]. However, due to the inherent ambiguities of local features, the putative correspondence set usually contains a large number of outliers.

In response to the outlier issue, a myriad of methods using different geometric constraints have been proposed. The resampling methods, represented by the well-known RANSAC algorithm [11], has been a standard solution for decades. Assuming rigid motion of objects, the geometric constraints of correspondences can be described by low-DoF parametric models, such as fundamental matrix, homography transformation or affine transformation. This establishes the theoretical foundation of the resampling methods, which iteratively draws a minimal subset of samples from the contaminated data, in the hope of finding an outlier-free subset to compute the true geometric model. The inliers can then be identified accordingly. Improvements of RANSAC cover almost all phases of the resampling scheme, including model quality evaluation [21], guided sampling [22], [23], [24], fast verification [25] and local optimization [26], [27]. Notice that Universal SAmple Consensus (USAC) [28] is an acknowledged representative of the RANSAC family, which unifies the most meaningful improvements. Recently, Maginalizing SAmple Consensus (MAGSAC) [29] has been proposed to get rid of the cumbersome requirements for setting an inlier-outlier threshold.

Generalized geometric constraints have also been extensively studied, which are based on the pursuit of smooth transformations. This first includes methods that aim to find a smooth function for the transformation of points. In Identifying point correspondences by Correspondence Function (ICF) [30], the so-called correspondence function is defined to model the bidirectional transformation and estimated by the robust Support Vector Regression technique to reject mismatches. In Bounded Distortion (BD) [31], the piecewise affine deformation model with bounded distortion is considered, and it has been shown that such a map can be found by solving a constrained optimization problem. A practically more powerful method is VFC [12], which explicitly imposes smoothness constraint. The nonparametric model is expressed as a vector field, which is assumed to be smooth in the RKHS, using the regularization theory. The whole estimation procedure is achieved in a Bayesian framework under the consideration of the outlier distribution. The VFC algorithm is efficient and general, and has encouraged many follow-up works [32], [33], [34], [35]. Recently, the geometric constraint has also been considered without specifying a transformation. In COherence based DEcision boundaries (CODE) [36], the true matches are identified by using likelihood functions, which are determined by nonlinear regression technique in a specially designed domain for correspondences to enforce local motion coherence. A clustering view can also be used to resolve the feature matching problem, by seeing each correspondence as a data point in a more general sense to separate inliers from outliers [37], [38], [39]. In Locality Preserving Matching (LPM) [40], a locality preserving matching method is proposed, where a (relaxed) local geometric distortion functional is defined. The credibility of each correspondence is directly given as the closed-form solution. This method also leads to some variants [41], [42], and a similar idea based on local supporting matches is proposed in Grid-based Motion Statistics (GMS) [43]. Learning-based matching methods have been given increasing attention in recent years. For instance, Ma *et al.* proposed a general framework to learn a two-class classifier for mismatch removal [44]. Yi *et al.* presented a first attempt to use deep learning techniques for the robust matching to aid the wide-baseline stereo task [45], followed by a number of more recent works [46], [47].

In addition to the two-stage strategy, which first uses local features to create putative correspondence set in the first stage and remove outliers with geometric constraint in the second stage, there are also a group of methods aiming to directly establish correspondence between two point sets. The point set registration methods and the majority of graph matching methods follow this idea. The former category, represented by Iterative Closest Point (ICP) [48], Coherent Point Drift (CPD) [3] and Thin-Plate Splines Robust Point Matching (TPS-RPM) [6], recovers a transformation to align the point sets. For the latter category, such as Spectral Matching with Affine Constraint (SMAC) [49], Integer Projected Fixed Point (IPFP) [50], Factorized Graph Matching (FGM) [7] and Composition based Affinity Optimization (CAO) [51], the problem is formulated as a combinatorial quadratic assignment problem. However, directly matching two point sets is a much harder problem, and these methods typically suffer from high computational complexity. Notice that due to the relaxation of constraints, some graph matching methods such as Spectral Matching (SM) [52] and Graph Shift (GS) [53] are applicable to solving the mismatch removal problem and quite efficient.

Aiming at recovering a smooth function, our method is firstly closely related to a number of outlier-robust methods, which is represented by the seminal work VFC [12], built on classical regularization theory. In comparison to VFC and its variants, the proposed method is differently established and built on a novel compact representation framework to exploit sparsity structure of smooth functions. The advantages of our formulation will be shown in the remainder of this paper. Our method is also closely related to certain point set registration methods, represented by CPD [3], which also adopts a probabilistic formulation. The differences are two-fold. First, the formulation of CPD is intended for directly establishing correspondences between two point sets, in contrast to our method which operates on putative correspondences. This drastically increases the computational complexity of CPD. Second, in a similar manner to VFC, CPD also utilizes the classical regularization theory to model deformations, which is different from our compact representation framework. These two factors render the proposed method a much more efficient choice for the matching problem, with an inherently sparse and more flexible deformation model.

III. Smoothness-Driven Multivariate Regression from Sparse Data

This section describes the proposed method to learn a smooth multivariate function from sparse data. We start by introducing the classical regularization theory, and then present our method using a compact representation.

A. Regularization Theory

Suppose we have obtained a set of sparse samples $S = \{(\mathbf{x}_n, y_n)\}_{n=1}^N \subset \mathcal{X} \times \mathcal{Y}$ sampled *i.i.d.* from an unknown probability distribution P on $\mathcal{X} \times \mathcal{Y}$. Typically, the input space $\mathcal{X}$ is a subset of $\mathbb{R}^D$, and the output space $\mathcal{Y}$ is a subset of $\mathbb{R}$. The goal is to learn a function f with small expected error $E[V(y, f(\mathbf{x})]$, in which the expectation is taken *w.r.t.* P and V is a prescribed loss function such as the square error $(y - f(\mathbf{x}))^2$. To recover the function from S is clearly ill-posed with no further restrictions on f, since it has an infinite number of solutions. A classical way to solve it is to use the regularization theory [13], [14], *i.e.* learning f as the minimizer of a regularized risk functional:

$$\min_{\mathbf{f} \in \mathcal{H}} \sum_{n=1}^{N} V(y_n, f(\mathbf{x}_n)) + \lambda \|f\|_{\mathcal{H}}^2, \tag{1}$$

where $\|f\|_{\mathcal{H}}$ is a norm in an RKHS $\mathcal{H}$ defined by the positive definite function K, λ is the regularization parameter that controls the tradeoff between the empirical risk and the complexity (smoothness) of the solution.

It can be shown that the optimal solution to (1) must have the form:

$$f(\mathbf{x}) = \sum_{n=1}^{N} c_n K(\mathbf{x}, \mathbf{x}_n), \tag{2}$$

where $\{c_n\}_{n=1}^N$ is a set of real parameters. The kernel K has the property that for $\mathbf{x} \in \mathcal{X}$, $K(\mathbf{x}, \cdot) \in \mathcal{H}$, and for $f \in \mathcal{H}$, $\langle f, K(\mathbf{x}, \cdot)\rangle_{\mathcal{H}} = f(\mathbf{x})$. Hence by (2):

$$\|f\|_{\mathcal{H}}^2 = \sum_{m=1}^{N} \sum_{n=1}^{N} c_m c_n K(\mathbf{x}_m, \mathbf{x}_n). \tag{3}$$

In light of (2) and (3), (1) reduces to

$$\min_{\mathbf{c}} \sum_{n=1}^{N} V(\mathbf{y}, \mathbf{Kc}) + \lambda \mathbf{c}^T \mathbf{Kc}, \tag{4}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_N]^T$, $\mathbf{c} = [c_1, c_2, \ldots, c_N]^T$ and $\mathbf{K}$ is an $N \times N$ matrix with mn-th entry as $K(\mathbf{x}_m, \mathbf{x}_n)$. The result in (2) is known as the *representer theorem*, which is remarkably important as it makes the variational problem (1) amenable for computations. In fact, as long as the loss function V is convex, a unique minimizer can be found by simple numerical algorithms for (4). Considering the simplest case with square error, we can see the coefficients can be obtained by solving the following linear system:

$$(\mathbf{K} + \lambda \mathbf{I})\mathbf{c} = \mathbf{y}. \tag{5}$$

Clearly, the linear system (5) incurs $O(N^3)$ time complexity and this cannot be reduced by using a different loss function since $\mathbf{K}$ is of size $N \times N$. The high computational complexity is one of the major restrictions for real-world applications. As in VFC, a straightforward idea is to apply the low rank approximation to the positive definite matrix $\mathbf{K}$, which reduces the time complexity of (5) to $O(N)$. However, the procedure itself involves the singular value decomposition (SVD) decomposition, which is still of $O(N^3)$ time complexity.

B. Learning Smooth Functions with Compact Representation

In this paper, we attempt to resolve this issue by exploiting the sparsity structure. Note that in classical regularization theory, the function is constructed in a highly expressive function space, *i.e.* the RKHS, and a regularization term is used to control its complexity. Our key observation is that, for a smooth function, there is a more compact representation, which can be fruitfully leveraged to design a computationally more efficient algorithm. Next, we discuss how to construct such a compact representation.

Without loss of generality, we consider the domain as a D-dimensional cube $\Omega := [0,1]^D$. In practice, this can be accomplished by a simple normalization step beforehand. We start by considering the eigenfunctions $\{\phi_1, \phi_2, \ldots\}$ and eigenvalues of $\{\mu_1, \mu_2, \ldots\}$ of the scalar Laplacian Δ on Ω:

$$-\Delta\phi_k = \mu_k \phi_k. \tag{6}$$

To complete the specification, we need to determine a boundary condition. Naturally, there are three different choices:

$$\phi_k(\mathbf{x}) = 0 \quad (\mathbf{x} \in \partial\Omega) \quad \text{(Dirichlet)}, \tag{7a}$$

$$\frac{\partial}{\partial n}\phi_k(\mathbf{x}) = 0 \quad (\mathbf{x} \in \partial\Omega) \quad \text{(Neumann)}, \tag{7b}$$

$$\frac{\partial}{\partial n}\phi_k(\mathbf{x}) + h\phi_k(\mathbf{x}) = 0 \quad (\mathbf{x} \in \partial\Omega) \quad \text{(Robin)}, \tag{7c}$$

where $\frac{\partial}{\partial n}$ is the normal derivative pointed outwards the domain, and h is a positive constant. For image matching and point set registration, the Dirichlet condition may not be a good choice since the intensity of the vector field can be very large near the boundary. Meanwhile, the Robin condition may be unnecessarily complex, thus we adopt a Neumann condition in our formulation. These ϕ_i can be then computed analytically, which are exactly the cosine elements of the Fourier basis [54]:

$$\mathcal{B}_\phi := \{\phi : [0,1]^D \to \mathbb{R}, \mathbf{x} \to \prod_{d=1}^{D} cos(x_d \pi j_d) | \mathbf{j} \in \mathbb{N}^D\}, \tag{8}$$

where x_d denotes the d-th component of $\mathbf{x} \in \Omega$, j_d denotes the d-th component of $\mathbf{j}$. The corresponding eigenvalue for each basis function is $\pi^2\|\mathbf{j}\|^2$. It has been proved that $\mathcal{B}_\phi$ forms a complete basis set in the function space $L^2(\Omega)$ of measurable and square-integrable functions on Ω [55], [56]. Some basis functions in the 2D case are visualized in Fig. 1.

It becomes much easier to impose smooth constraint with $\mathcal{B}_\phi$ since the eigenvalue is conceptually related to frequency. This means the number of basis functions can be adjusted for either speed or expressiveness. Specifically, to represent a smooth function, the Fourier bases are first rearranged in ascending order of their eigenvalues $\pi^2\|\mathbf{j}\|^2$, and $\mathcal{B}_\phi$ interchangeably becomes $\{\phi_1, \phi_2, \ldots\}$. That is to say, denoting the eigenvalue of function ϕ_i as μ_i, the condition $0 \le \mu_1 \le \mu_2 \le \ldots \nearrow \infty$ is satisfied. Then we can define and use a restricted set of T bases:

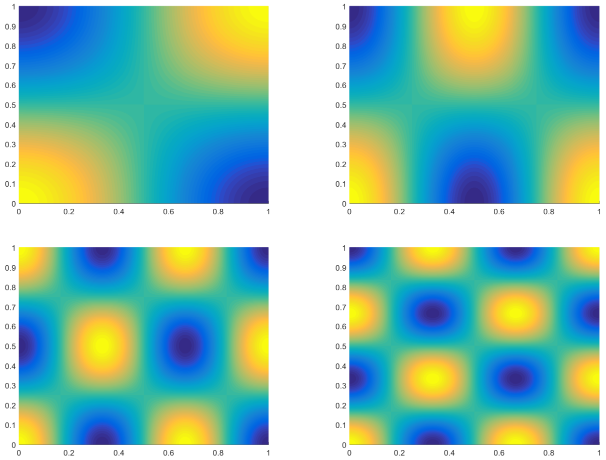


Fig. 1. Visualization of some Fourier basis functions in $[0,1]^2$ with Neumann boundary condition. From left to right, top to bottom, the frequency increases and the corresponding $\mathbf{j}$ that determines the functions are $[1,1]$, $[2,1]$, $[3,2]$ and $[3,3]$, respectively. The continuous transformation of color from blue to yellow indicates the values changing from small to large.

$$\mathcal{B}_T := \{\phi_1, \phi_2, \ldots, \phi_T\}, \tag{9}$$

which implies a compact representation:

$$f(\mathbf{x}) = \sum_{n=1}^{T} a_n \phi_n(\mathbf{x}). \tag{10}$$

To further regularize the function, we interpret the coefficients $\mathbf{a} = [a_1, a_2, \ldots, a_T]^T$ as random variables with a normal distribution $\mathbf{a} \sim \mathcal{N}(0, \frac{1}{\lambda}\mathbf{R})$, with $\mathbf{R} := diag(\omega_1, \omega_2, \ldots, \omega_T)$. The weights ω_k are constructed from the eigenvalues as follows:

$$\omega_k = \mu_k^{-\frac{D}{2}}. \tag{11}$$

The mathematical background of this choice for the weights follows the Karhunen-Loeve expansion [57], which promotes a damping of the high frequency components and thereby smoothness of the function. Consequently, with the compact representation (10), the variational multivariate regression problem has the following finite-dimensional form:

$$\min_{\mathbf{a}} \sum_{n=1}^{N} V(\mathbf{y}, \mathbf{\Gamma a}) + \lambda \mathbf{a}^T \mathbf{R}^{-1} \mathbf{a}, \tag{12}$$

where $\mathbf{\Gamma}$ is an $N \times T$ matrix with each entry $\mathbf{\Gamma}_{mn} = \phi_n(\mathbf{x}_m)$.

Analogously, for the simplest case with square error, the optimal solution can be obtained by solving the following linear system:

$$(\mathbf{\Gamma}^T\mathbf{\Gamma} + \lambda\mathbf{R}^{-1})\mathbf{a} = \mathbf{\Gamma}^T\mathbf{y}. \tag{13}$$

Clearly, (13) is computationally much more efficient to solve compared to (5), with $O(N)$ time complexity.

IV. Smooth Functions for Robust Feature Matching

In this section, we focus on the fundamental problem in computer vision, *i.e.* robust feature matching, as an application for the proposed compact representation technique. As will be shown, the problem can be translated to learning several multivariate functions, which are required to be smooth but less demanding in accuracy. This is exactly the scenario that our compact representation can excel in. Moreover, in the last section, we mainly discuss the multivariate regression problem with a clean sample set, *i.e.* no outliers. However, for robust feature matching, a large number of outliers may be included in the data, thus we start by introducing a Bayesian formulation to cope with outliers.

A. Problem Formulation

Suppose we have obtained a set of putative correspondences $S = \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^N$, where $\mathbf{x}_n, \mathbf{y}_n \in \mathbb{R}^D$ are D-dimensional column vectors and typically $D = 2$ or 3. To identify the inlier set $\mathcal{I} \subseteq S$, we aim to recover from S the underlying transformation, *i.e.* a smooth vector field $\mathbf{f} : \mathbb{R}^D \to \mathbb{R}^D$ in our context, such that $\mathbf{y}_n = \mathbf{f}(\mathbf{x}_n)$ for $(\mathbf{x}_n, \mathbf{y}_n) \in \mathcal{I}$. The inlier set $\mathcal{I}$ can be readily found as the correspondences consistent with the transformation.

To this end, for each correspondence $(\mathbf{x}_n, \mathbf{y}_n)$, we consider it as a measurement $\mathbf{y}_n$ at position $\mathbf{x}_n$. We assume that the correspondences are independent and identically distributed. Moreover, for the inliers, we assume that the noise is Gaussian on each component with zero mean and uniform standard deviation σ. For the outliers, since the measurement $\mathbf{y}_n$ lies randomly in a bounded region in $\mathbb{R}^D$, we assume the distribution to be uniform $1/a$ with a denoting the volume of this region. We then associate the n-th correspondence with a latent variable $z_n \in \{0, 1\}$, where $z_n = 1$ indicates the correspondence $(\mathbf{x}_n, \mathbf{y}_n)$ being an inlier and $z_n = 0$ indicates otherwise. Each latent variable z_n follows a discrete distribution, *i.e.* $p(z_n = 1) = \gamma$, and $p(z_n = 0) = 1-\gamma$, where $\gamma \in [0, 1]$. Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)^T$ be the position data, and let $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N)^T$ be the measurements. Under the *i.i.d.* assumption of data, the joint likelihood function takes the form:

$$\begin{aligned} p(\mathbf{Y}|\mathbf{X}, \theta) &= \prod_{n=1}^N \sum_{z_n} p(\mathbf{y}_n, z_n|\mathbf{x}_n, \theta) \\ &= \prod_{n=1}^N \left(\frac{\gamma}{(2\pi\sigma^2)^{D/2}} e^{-\frac{\|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2}{2\sigma^2}} + \frac{1-\gamma}{a}\right), \end{aligned} \tag{14}$$

where $\theta = \{\mathbf{f}, \sigma, \gamma\}$ represents the unknown parameters. In a Bayesian view, we also have prior information for θ to regularize the estimation process, expressed as a prior distribution $p(\theta)$.

Using Bayes rule, the maximum *a posteriori* (MAP) estimation can be expressed as:

$$\theta^* = \arg\max_\theta p(\theta|\mathbf{X}, \mathbf{Y}) = \arg\max_\theta p(\mathbf{Y}|\mathbf{X}, \theta) p(\theta). \tag{15}$$

This is equivalent to finding the parameters that minimize the following energy:

$$E(\theta) = -\ln p(\theta) - \prod_{n=1}^N \ln \sum_{z_n} p(\mathbf{y}_n, z_n|\mathbf{x}_n, \theta). \tag{16}$$

B. An Expectation-Maximization Based Solution

To solve (16), we consider the EM algorithm, which is a general technique for learning and inference with the existence of latent variables and very efficient. Basically, it is an iterative algorithm that alternates between two steps: the expectation step (E-step) and the maximization step (M-step). Considering the negative log posterior function (16), the expectation of the complete-data log likelihood is:

$$\begin{aligned} \mathcal{L}(\theta, \theta^{old}) = &-\frac{1}{2\sigma^2} \sum_{n=1}^N p_n \|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2 - \frac{D}{2} \ln \sigma^2 \sum_{n=1}^N p_n \\ &+ \ln(1-\gamma) \sum_{n=1}^N (1 - p_n) + \ln \gamma \sum_{n=1}^N p_n - \ln p(\theta), \end{aligned} \tag{17}$$

where $p_n = P(z_n = 1|\mathbf{x}_n, \mathbf{y}_n, \theta^{old})$ denotes the posterior probability of z_n. The E-step and M-step are accordingly outlined below:

E-step: In this step, the posterior probabilities are evaluated based on the current parameter values. Due to the *i.i.d.* assumption, it can be achieved separately for each correspondence:

$$p_n = \frac{\gamma e^{-\frac{\|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2}{2\sigma^2}}}{\gamma e^{-\frac{\|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2}{2\sigma^2}} + (1-\gamma)\frac{(2\pi\sigma^2)^{D/2}}{a}}. \tag{18}$$

M-step: This step determines the revised parameter estimation θ^{new} using (17): $\theta^{new} = \arg\max_\theta \mathcal{L}(\theta, \theta^{old})$. For σ and γ, the updating rules can be derived by taking the derivatives of (17) and setting them to 0. Let $\mathbf{P} = diag(p_1, p_2, \ldots, p_N)$ be a diagonal matrix, we have:

$$\sigma^2 = \frac{tr((\mathbf{Y} - \mathbf{T})^T \mathbf{P} (\mathbf{Y} - \mathbf{T}))}{D \cdot tr(\mathbf{P})}, \tag{19}$$

$$\gamma = \frac{tr(\mathbf{P})}{N}, \tag{20}$$

where $\mathbf{T} = (\mathbf{f}(\mathbf{x}_1), \mathbf{f}(\mathbf{x}_2), \ldots, \mathbf{f}(\mathbf{x}_n))^T$ and $tr(\cdot)$ is the trace.

Next we consider to update the vector field function $\mathbf{f}$. We assume flat priors for σ and γ, thus $p(\theta)$ degrades into $p(\mathbf{f})$. Abstracting the terms related to $\mathbf{f}$, we obtain the following functional:

$$\varepsilon(\mathbf{f}) = \frac{1}{2\sigma^2} \sum_{n=1}^N p_n \|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2 - \ln p(\mathbf{f}). \tag{21}$$

This is the core step for vector field learning, and in this paper, we use the compact representation technique for the best efficiency.

Remark. Note that (21) requires to compute a vector-valued function $\mathbf{f}$. Typically, this relates the task to multi-task or multi-output learning, which aims to leverage useful information contained in each task to help improve the overall

Algorithm 1 Compact Representation Consensus

```
Input: The correspondence set $\mathcal{S}$, basis functions $\mathcal{B}_T$, constant $\lambda$ and inlier threshold $\tau$
Output: Inlier set $\mathcal{I}$
Initialize $\gamma$, $a$, $\mathbf{T} = \mathbf{0}_{N\times 2}$, $\mathbf{P} = \mathbf{I}_{N\times N}$;
Initialize $\sigma^2$ by (19);
Construct $\mathbf{\Gamma}$ using $\mathcal{B}_T$;
while $\mathcal{L}$ not converge do
    E-step:
        Update $\mathbf{P}$ by (18);
    M-step:
        Update $\mathbf{a}_i$ by (24) for $i = 1, 2, \ldots, D$;
        Update $\mathbf{T}$ by (22);
        Update $\sigma^2$ and $\gamma$ by (19) and (20);
end while
Determine the inlier set by (25).
```

performance. In this context, different tasks correspond to learning the mappings from $\mathbf{x}_n$ to each component of $\mathbf{y}_n$. We denote each mapping as $\mathbf{f}_i : \mathbb{R}^D \to \mathbb{R}$ for the i-th component. However, in most previous works for robust matching [3], [12], the implicit assumption that $\mathbf{f}_i$ is independent is used. This is mainly because that it is unclear how the mappings are related, and that the independent assumption works well in practice with preferred efficiency. Thus, we also take this assumption, which allows the use of multivariate regression techniques.

Let each mapping $\mathbf{f}_i$ be a compact representation to the underlying smooth function:

$$\mathbf{f}_i(\mathbf{x}) = \sum_{n=1}^{T} a_n^i \phi_n(\mathbf{x}). \tag{22}$$

As aforementioned, we can see $\mathbf{a}_i = [a_1^i, a_2^i, \ldots, a_T^i]^T$ as random variables, and the term $-\ln p(\mathbf{f}_i)$ translates to $\mathbf{a}_i^T \mathbf{R}^{-1} \mathbf{a}_i$. Consequently, (21) can be decomposed into solving the following problem:

$$\begin{aligned} \min_{\mathbf{f}_i \in span(\mathcal{B}_T)} \; & \frac{1}{2\sigma^2} \sum_{n=1}^{N} p_n \|y_n^i - \mathbf{f}_i(\mathbf{x}_n)\|^2 - \ln p(\mathbf{f}_i) \\ & = \frac{1}{2\sigma^2} \|\mathbf{P}^{1/2}(\mathbf{Y}_i - \mathbf{\Gamma}\mathbf{a}_i)\|^2 + \lambda \mathbf{a}_i^T \mathbf{R}^{-1} \mathbf{a}_i, \end{aligned} \tag{23}$$

where y_n^i denotes the i-th component of $\mathbf{y}_n$, $\mathbf{Y}_i$ is the i-th column of $\mathbf{Y}$ and $\mathbf{\Gamma}$ is an $N \times T$ matrix with each entry $\mathbf{\Gamma}_{mn} = \phi_n(\mathbf{x}_m)$.

It can be seen that (23) is convex and the optimal solution can be obtained by solving the following linear system:

$$(\mathbf{\Gamma}^T \mathbf{P} \mathbf{\Gamma} + \lambda \sigma^2 \mathbf{R}^{-1}) \mathbf{a}_i = \mathbf{\Gamma}^T \mathbf{P} \mathbf{Y}_i. \tag{24}$$

After convergence of the EM algorithm, we can obtain the inlier set with a predefined threshold τ by evaluating the posterior probability:

$$\mathcal{I} = \{(\mathbf{x}_n, \mathbf{y}_n) : p_n > \tau, n \in \mathbb{N}_N\}. \tag{25}$$

We name our algorithm as Compact Representation Consensus (CRC) and summarize it in Alg. 1.

C. Computational Complexity and Implementation Details

The main computational cost for our CRC is to solve the linear system (24), which requires runtime in $O(N)$ since $\mathbf{\Gamma}$ is of $N \times T$ and $N \gg T$. Thus the time complexity of our CRC is $O(cN)$ where c denotes the number of iterations for EM. The space complexity is $O(N)$ to store $\mathbf{\Gamma}$. Although our CRC has theoretically the same time complexity as a sparse random basis approximation, it is much faster due to the numerical instability of the latter. This can be seen in Sec. V-A.

There are several parameters to be set for our CRC, *i.e.* T, λ, γ, and τ. Parameter T represents the number of adopted basis functions, we empirically set it to 15 as suggested in Sec. V-A. Parameter λ is used to control the magnitude of the regularization for $\mathbf{a}$, we empirically set it to 1. Parameter γ reflects our initial assumption on the amount of inliers in the correspondence sets, we empirically set it to 0.95. Parameter τ is a threshold to decide the correctness of a match, we empirically set it to 0.75. The data is normalized to $[0, 1]^2$ before processing.

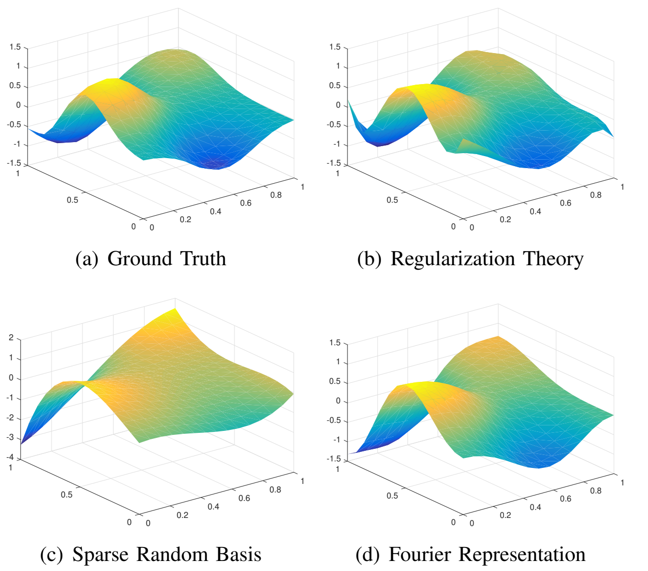
V. Experimental Results

In this section, numerical experiments are conducted to demonstrate the superiority of our method. First, synthetic data is used to compare the proposed method with classical regularization theory and its several variants for the task of multivariate regression under outliers. Then, the application to image feature matching is considered and real image datasets are used. Additionally, we explore the generality of the proposed method on point set registration. The experiments are performed on a laptop with 2.3 GHz Intel Core CPU, 16 GB memory and Matlab Code.

A. Multivariate Regression with Synthetic Data

In a general point of view, our method can be seen as an alternative of the classical regularization framework in learning smooth functions from sparse samples. We use synthetic data to provide an ideal environment for testing the performances of our method and the regularization theory. We constructed the ground truth smooth function as a mixture of C Gaussian functions, which have the same covariance $0.04\mathbf{I}$ and center at random positions in $[0, 1]^2$ with different random amplitudes restricted to $[0, 1]$. In this way, the function is guaranteed to be smooth and its complexity grows with C. An example is shown in Fig. 2 with $C = 8$, which is shown as a smooth surface in the 3D space. We then draw random samples from this function and manually add a number of outliers to the data as the input to each method. The performance is evaluated by the discrepancy between the recovered function and the ground truth. The discrepancy is calculated as the mean absolute error *w.r.t.* a number of uniform samples in $[0, 1]^2$.

Three competitors are adopted for comparison, including the classical regularization theory [13], [14] and two approximation schemes based on low-rank approximation [12] and sparse random basis [15]. For low-rank approximation, we approximate the kernel matrix using the first 30 largest eigenvalues and the corresponding eigenvectors. For sparse random basis, we use 80 bases as suggested in [15]. All



(a) Ground Truth (b) Regularization Theory

(c) Sparse Random Basis (d) Fourier Representation

Fig. 2. A testing example of the algorithms for learning smooth functions from sparse outlier-contaminated samples. The function is visualized as a surface embedded in 3D space.

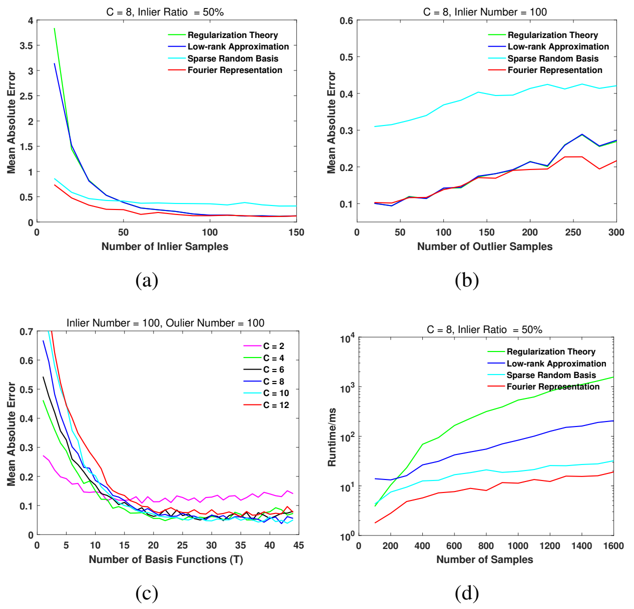


Fig. 3. The summary statistics of each method's performance on synthetic data. Each data point is averaged over 50 instances of randomly generated smooth surfaces.

methods are implemented using the same Bayesian framework to cope with outliers. Fig. 3 presents all the testing results, where each datum is a summary statistic based on 50 randomly generated synthetic instances. Fig. 3(a) and Fig. 3(b) show the performance of each method against the number of inlier samples and outlier samples, respectively. It can be seen that in our settings, 100 samples are sufficient to recover the smooth function. Our method based on compact Fourier Representation is clearly advantaged since it requires fewer inlier samples for regression, and is more robust to outliers. The regularization theory and the low-rank approximation has similar performances. However, the sparse random basis strategy seems to fail to give accurate results. As shown in Fig. 2, it generates unduly smooth surface. In Fig. 3(c), the performance of our method *w.r.t.* the number of basis functions (T) is investigated. We can see that generally 15 basis functions are already sufficient for regression of smooth functions, which clearly indicates the compactness of the Fourier representation. Fig. 3(d) shows the runtime statistics. It can be observed that our method has the lowest computational complexity, and can be orders of magnitude faster that the competitors. Although the sparse random basis method has the same computational complexity as our method, it is significantly slower than the proposed CRC due to its numerical instability.

B. Robust Image Feature Matching

In this subsection, we focus on the application of robust feature matching for image datasets. The aim is to differentiate the inliers from the contaminated correspondence set. As aforementioned, our method has the same rationale as VFC [12] and its variants FastVFC and SparseVFC [15], *i.e.*, fitting smooth functions to reject the outliers. Three widely used datasets in the robust feature matching literature are adopted in our evaluation:

-*DAISY*. The dataset is used in [40], which consists of wide baseline image pairs with ground truth depth maps, including two short image sequences and several individual image pairs. In total 52 image pairs are created for evaluation. Determined by the imaging scenes, the correspondences are related by epipolar constraint.

-*CRS*. The dataset is a composition of the rigid and projective remote sensing datasets used in [42], which consists of 161 pairs of remote sensing images, including color images captured by a UAV, images obtained by synthetic-aperture radars, panchromatic aerial photographs and color infrared aerial photographs. The relation between the image pairs can be described by homography transformation.

-*VGG+*. The image pairs in this dataset are a collective of the data used in VFC [12], including image pairs related by homography and non-rigid transformation, and image pairs of wide baseline. We use SIFT to establish the putative correspondence set for each pair, and manually annotate the inliers as ground truth. This dataset is quite challenging due to the low inlier ratio.

Some representative image matching examples from the adopted datasets are used for testing of the proposed CRC, as shown in Fig. 4. The data cover different types of feature matching scenarios including image pairs related by homography, fundamental matrix and non-rigid transformation. For each group of results, the left image pair schematically shows the matching result, and the right motion field provides the decision correctness of each correspondence in the putative set. It can be seen that our CRC can always produce satisfying results, regardless of the complex relations of the image pairs, and even in the presence of a high outlier ratio.

Since our method utilizes EM algorithm for optimization, it essentially gives only a locally optimal solution, thus the problem of sensitivity to initial guess also needs to be addressed. We use the combination of all the three datasets,

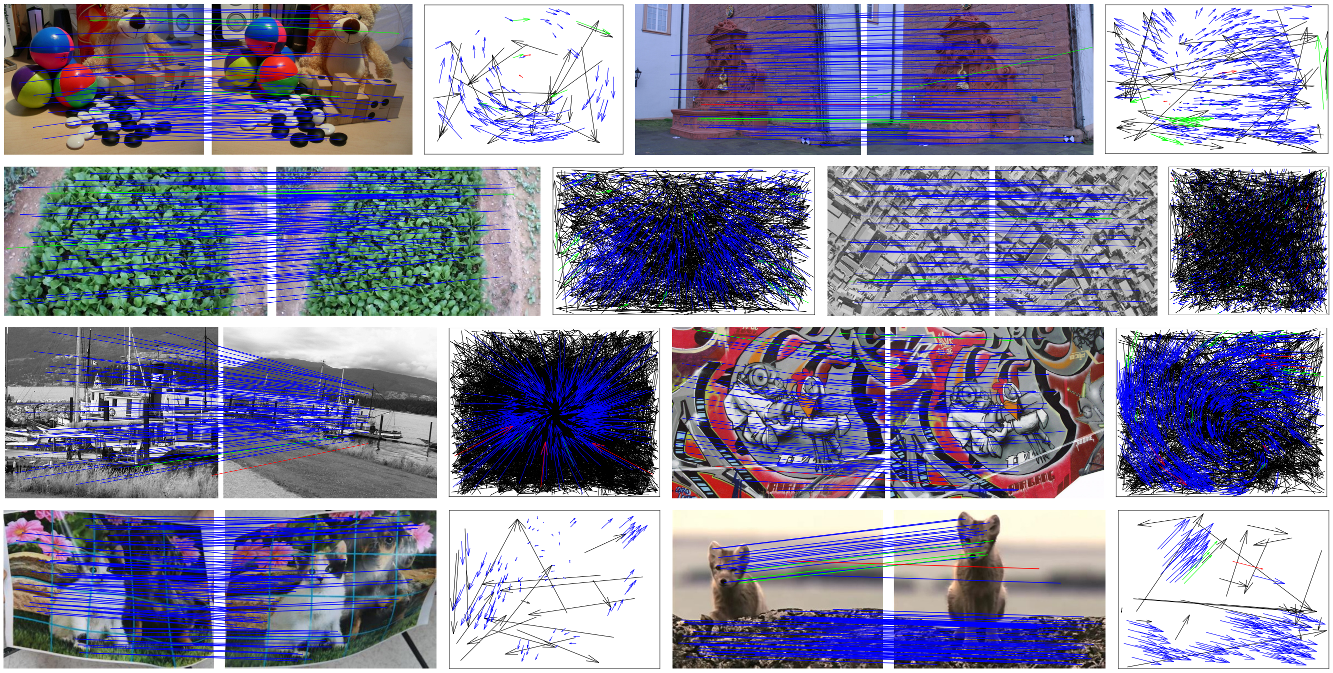


Fig. 4. Feature matching results of our CRC on eight representative image pairs from *DAISY*, *CRS*, and *VGG+*. The head and tail of each arrow in the motion field correspond to the positions of feature points in two images (blue = true positive, black = true negative, green = false negative, red = false positive). For visibility, in the image pairs, at most 100 randomly selected matches are presented, and the true negatives are not shown. Best viewed in color.

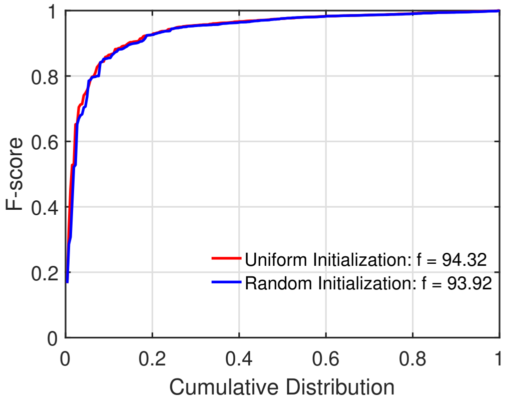


Fig. 5. Testing for the influences of different initial guesses to our CRC on the combination of *DAISY*, *CRS* and *VGG+* datasets. The performance is evaluated by F-score, and the red curve represents the result of a uniform initialization, the blue curve represents a random initialization. The average F-score for the two settings are given in the legend.

namely *DAISY*, *CRS* and *VGG+* to test the robustness of our method to different initial guesses. In our method, the initial guess is determined by the initialization of $\mathbf{P}$, which encodes the probabilities of each correspondence being an inlier. We construct two different initializations, *i.e.* uniform initialization which gives a uniform probability, and random initializations which gives a random probability for each correspondence. The performance is evaluated by F-score and the results are given in Fig. 5. It can be observed that the uniform initialization and the random initialization give very close results, which clearly demonstrates the robustness of our method to initial guesses.

Nine state-of-the-art methods for robust feature matching are adopted for comparison, namely RANSAC [11], ICF [30], SM [52], VFC [12], FastVFC [12], SparseVFC [15], LPM [40], GMS [43], and the recent deep learning technique ACNe [47]. All these algorithms are implemented based on publicly available codes. The parameters are fixed after being carefully tuned. The quantitative results are presented in Fig. 6. The cumulative distribution of initial inlier ratios on the three data sets is provided in the first row. We can see that the *VGG+* dataset is the most challenging one with quite low inlier ratio. The statistic results on the three data sets, *i.e.* precision, recall, F-score and runtime, are summarized in the second, third, fourth and fifth row, respectively. The classical RANSAC algorithm has a varying performance in the three datasets. This is due to its randomized nature, which is disadvantaged in the presence of severe outliers. The ICF algorithm also attempts to recover a smooth function, however, it is very sensitive to parameter settings, which leads to its inferior performances. The SM algorithm utilizes pairwise relations to obtain the matches with high credibility, which has better generality but less accuracy, thus its performances are generally moderate. The VFC family, *i.e.* VFC, FastVFC and SparseVFC, have very close performances, and outperform all the other methods despite their high runtime. The LPM and GMS algorithm are much more efficient, yet they can not achieve on par performance with the VFC family. In contrast, our CRC is very competitive in both accuracy and computational complexity. Based on a similar rationale, it has slightly inferior performance with the VFC family, but is more than an order of magnitude faster. Different from SparseVFC, it achieves true linear complexity in our experiments, and the runtime is less than 10 milliseconds in general. This demonstrates the superiority of our method. Also, our method has shown a

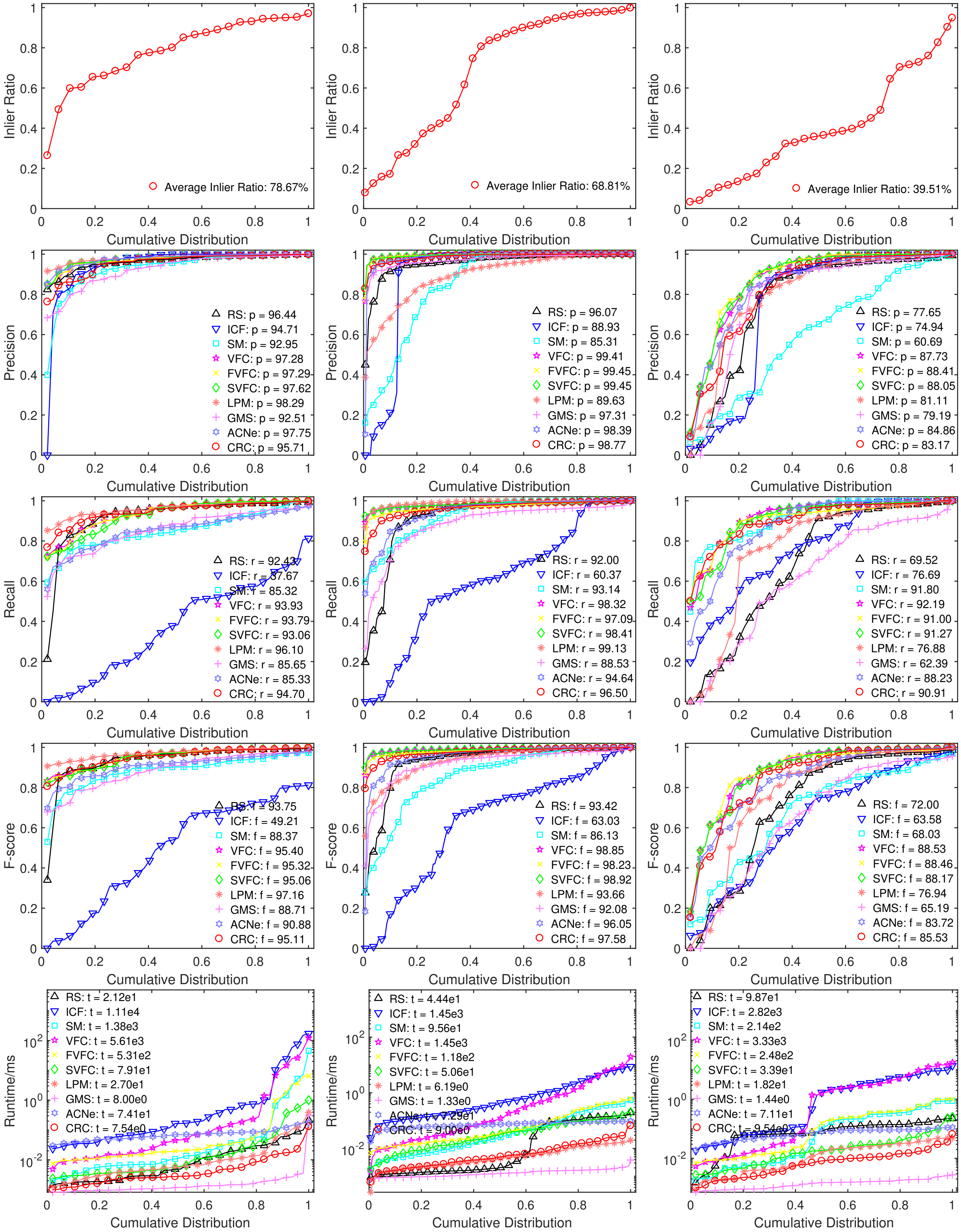


Fig. 6. Quantitative comparisons of our CRC and nine competitors for robust image feature matching on three datasets, *i.e.* (left to right) DAISY, CRS, VGG+. From top to bottom, the initial inlier ratio, precision, recall, F-score and runtime are presented with respect to the cumulative distribution. The average statistics are reported in the legend for each method. FastVFC and SparseVFC are abbreviated as FVFC and SVFC, respectively.

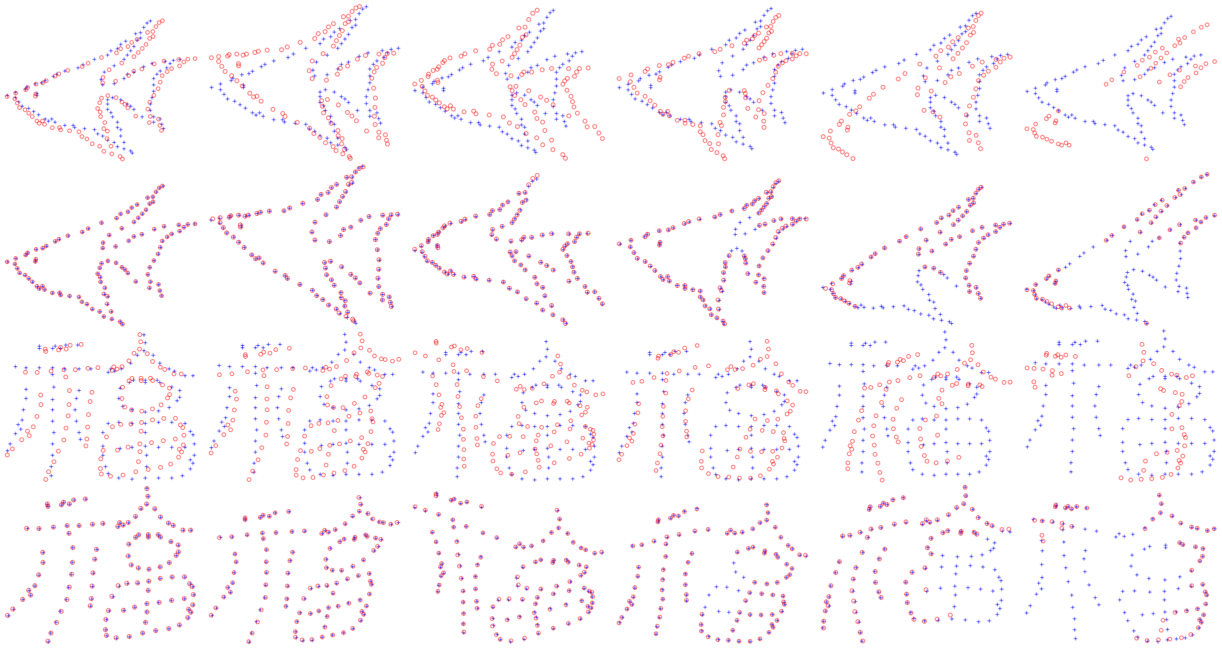


Fig. 7. Registration results of our method on the *fish* and *Chinese character* patterns. The goal is to align the model point sets (blue pluses) onto the target point sets (red circles). In the first three columns, the point sets suffer from different levels of deformation, and in the last three columns, the point sets suffer from different levels of occlusion. The first and third rows are the initial point sets, and the second and fourth rows present the registration results.

clear advantage compared to the state-of-the-art deep learning technique ACNe, in terms of both accuracy and efficiency.

C. Non-rigid Point Set Registration

Point set registration is a classical problem in computer vision, which aims to align one point set to another by recovering the transformation. For the non-rigid case, the most popular solutions are typically based on regularization theory. Similar to robust feature matching, the transformation is generally simplified in practice, as the composition of two separate multivariate functions [3]. Thus we can apply our CRC to robustly recover the transformation and register the point sets. As in VFC, our algorithm follows an ICP-like procedure. In each iteration, we use Shape Context [9] to establish correspondences between point sets and then use CRC to robustly estimate the transformation function. We empirically use the transformed point set after 10 iterations as the final registration result.

We adopt both synthetic data and real-world data to demonstrate the efficacy of the proposed method. For synthetic data, the *fish* and *Chinese character* patterns [58] with different degrees of deformation and occlusion are used in our experiment. Some qualitative results of our method on the two shapes are depicted in Fig. 7. We organize the results in every two rows: the first row is the initial point sets, the second row is the corresponding registration results. For real-world data, we adopt the publicly available benchmark IMM Face Database [59][1] for evaluation. The database consists of 240 images of 40 human faces with resolution 640 × 480, and for each face there are 6 samples of different expressions, poses and illuminations. The facial structures such as eyebrows, eyes, nose, mouth and jaw are annotated using 58 landmarks. The ground-truth correspondences between the landmark sets are supplied with the database. In our experiment, we aim at aligning the landmark sets extracted from different samples of an individual. For each individual, we construct the registration pairs using the first sample and each of the rest five samples, thus creating 5 groups (40 pairs for each group) for registration. Some qualitative results of face landmark registration with comparison to other methods are presented in Fig. 8.

For comparison, we adopt the well-known CPD method [3] and VFC [12], which are in essence developed based on regularization theory, and also Gaussian Mixture Model Registration (GMMR) [60] as the competitors. The registration error between two point sets is characterized by the average Euclidean distance of the ground truth correspondences between the warped model set and the target set. For each degradation level in synthetic data, we compute the mean and standard deviation of the registration errors on all 100 instances to derive the summary statistics. We also compute these metrics for each group in the real-world data. The quantitative results are reported in Fig. 9 for synthetic data and Fig. 10 for real-world data. From the results, we can see that our method is able to produce accurate alignments even in the presence of severe degradations such as non-rigid deformation and occlusion. In addition, since CPD and GMMR do not involve local features in the registration process, they are generally outperformed by VFC and our CRC. The main difference between VFC and our CRC is the modeling of transformation. It can be seen that with our compact representation, the registration error is consistently reduced, as manifested both in synthetic data and real-world data. This demonstrates the superiority of our method as an alternative of the methods driven by classical regularization theory.

[1] http://www.imm.dtu.dk/ aam/datasets/datasets.html

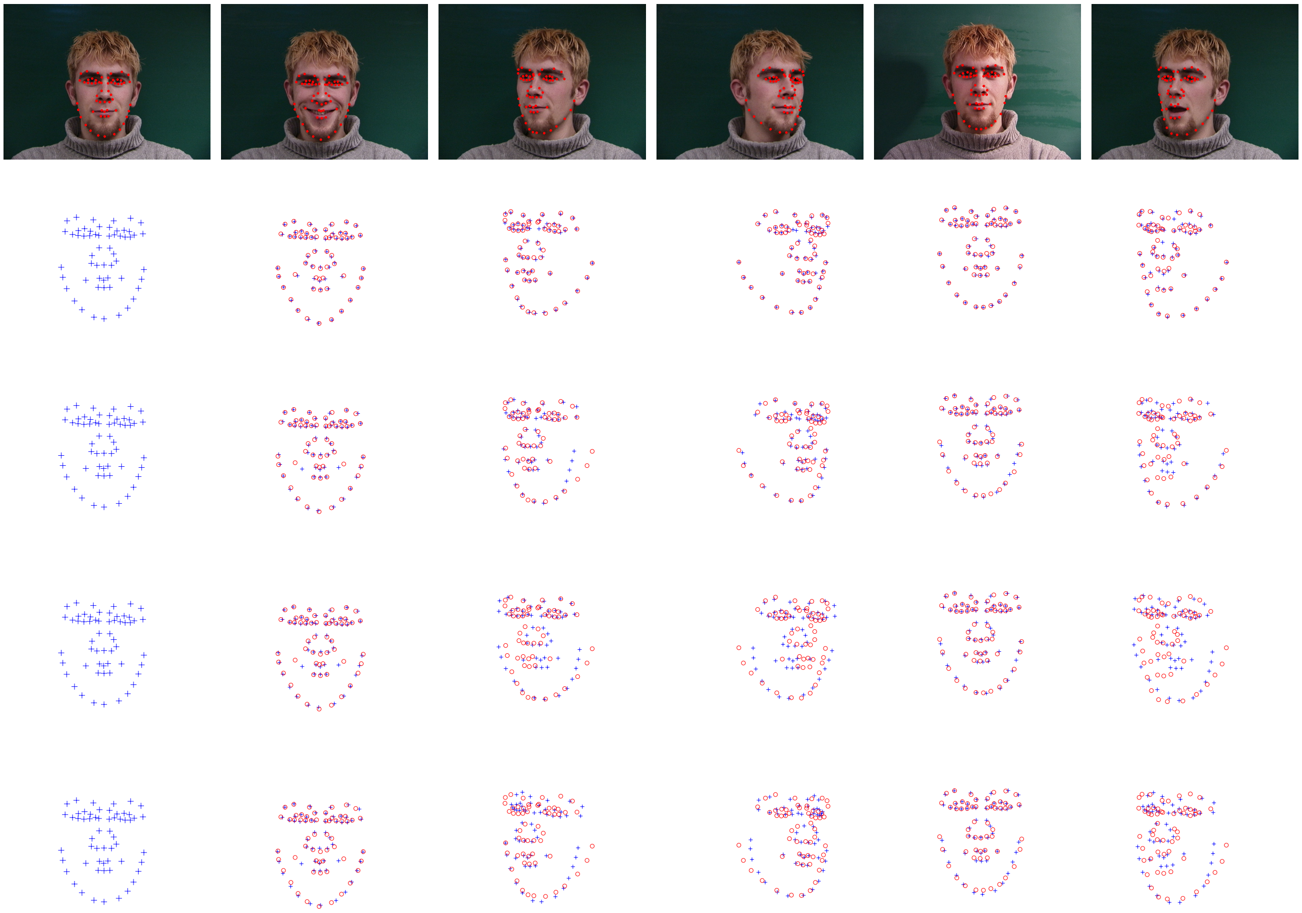


Fig. 8. Some qualitative examples for face landmark registration. The first row showcases an example face group in the IMM Face Database with annotations. The second row showcases the registration result of our CRC method, and the third, fourth and fifth row showcase the results of VFC, CPD and GMMR, respectively.

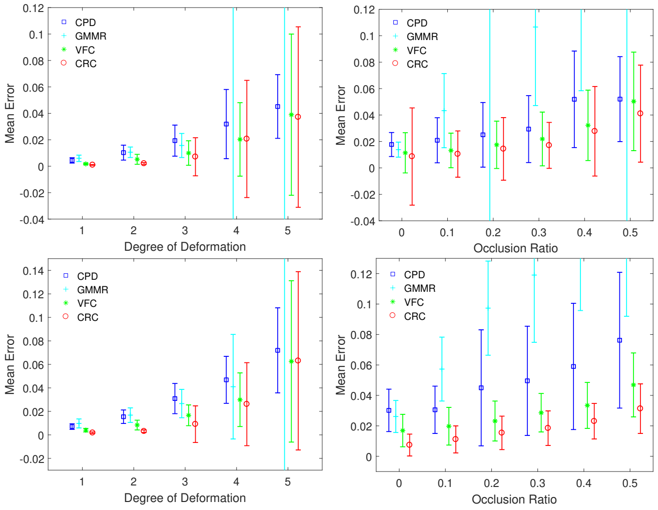


Fig. 9. Quantitative comparison of CPD, VFC, GMMR and our CRC on the *fish* (top) and *Chinese character* (bottom) patterns. The error bars indicate the registration error means and standard deviations over 100 instances.

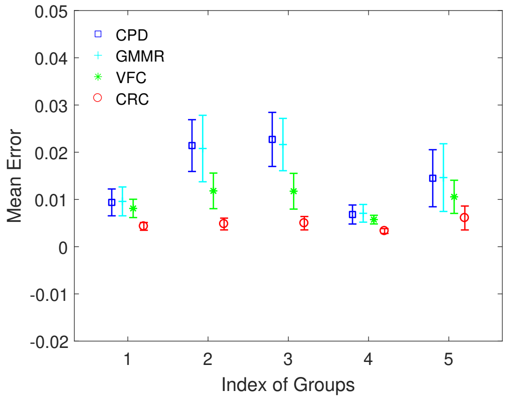


Fig. 10. Quantitative comparison of CPD, VFC, GMMR and our CRC on the IMM Face Database. The error bars indicate the registration error means and standard deviations over each group in the database.

VI. Conclusion

In this paper, we propose a novel method for learning smooth functions from sparse data based on a compact Fourier representation, and extend it with outlier-robust property using a Bayesian framework. The theoretical performance has been tested with synthetic data, in addition, practical applications such as robust image feature matching and point set registration are also investigated. Experimental results have demonstrated that due to the exploitation of sparsity structure, our method is both robust and efficient, being orders of magnitude faster compared to classical regularization theory without degrading the accuracy.

Although the proposed method has shown remarkable performance in our evaluation, there still exist some limitations for our CRC. For one, the deformation with the proposed Fourier basis representation is somewhat restricted due to the boundary condition. This means that the method may fail to accurately model a general smooth function near the boundary. Also, for high-dimensional data, the proposed method may also be impotent since the number of required low-frequency Fourier basis functions grows exponentially with dimension. In the future, we will look more carefully at the problem of boundary problem to improve the representation power of the proposed CRC method, thus extending the scope of CRC to more general deformations. Additionally, we also plan to explore more practical scenarios to generalize CRC for a broader range of applications.

References

[1] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 23–79, 2021.

[2] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, 2007.

[3] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, 2010.

[4] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4104–4113.

[5] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *IEEE Trans. Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[6] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Understand.*, vol. 89, no. 2-3, pp. 114–141, 2003.

[7] F. Zhou and F. De la Torre, "Factorized graph matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 127–134.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[9] S. Belongie, J. Malik, and J. Puzicha, "Shape context: A new descriptor for shape matching and object recognition," in *Adv. Neural Inf. Process. Syst.*, 2001, pp. 831–837.

[10] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

[11] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[12] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1706–1721, 2014.

[13] F. Girosi, M. Jones, and T. Poggio, "Regularization theory and neural networks architectures," *Neural Comput.*, vol. 7, no. 2, pp. 219–269, 1995.

[14] T. Evgeniou, M. Pontil, and T. Poggio, "Regularization networks and support vector machines," *Advances in Computational Mathematics*, vol. 13, no. 1, p. 1, 2000.

[15] J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu, "Regularized vector field learning with sparse approximation for mismatch removal," *Pattern Recognit.*, vol. 46, no. 12, pp. 3519–3532, 2013.

[16] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 2564–2571.

[17] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 224–236.

[18] Y. Tian, B. Fan, and F. Wu, "L2-net: Deep learning of discriminative patch descriptor in euclidean space," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 661–669.

[19] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (fpfh) for 3d registration," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2009, pp. 3212–3217.

[20] H. Deng, T. Birdal, and S. Ilic, "Ppfnet: Global context aware local features for robust 3d point matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 195–205.

[21] P. H. Torr and A. Zisserman, "Mlesac: A new robust estimator with application to estimating image geometry," *Comput. Vis. Image Understand.*, vol. 78, no. 1, pp. 138–156, 2000.

[22] O. Chum and J. Matas, "Matching with prosac-progressive sample consensus," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 220–226.

[23] V. Fragoso, P. Sen, S. Rodriguez, and M. Turk, "Evsac: accelerating hypotheses generation by modeling matching scores with extreme value theory," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2472–2479.

[24] K. Ni, H. Jin, and F. Dellaert, "Groupsac: Efficient consensus in the presence of groupings," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 2193–2200.

[25] O. Chum and J. Matas, "Optimal randomized ransac," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 8, pp. 1472–1482, 2008.

[26] O. Chum, J. Matas, and J. Kittler, "Locally optimized ransac," in *Proc. Joint Pattern Recognit. Symp.*, 2003, pp. 236–243.

[27] K. Lebeda, J. Matas, and O. Chum, "Fixing the locally optimized ransac–full experimental evaluation," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 1–11.

[28] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J.-M. Frahm, "Usac: a universal framework for random sample consensus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 2022–2038, 2012.

[29] D. Barath, J. Matas, and J. Noskova, "Magsac: marginalizing sample consensus," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10 197–10 205.

[30] X. Li and Z. Hu, "Rejecting mismatches by correspondence function," *Int. J. Comput. Vis.*, vol. 89, no. 1, pp. 1–17, 2010.

[31] Y. Lipman, S. Yagev, R. Poranne, D. W. Jacobs, and R. Basri, "Feature matching with bounded distortion," *ACM Trans. Graph.*, vol. 33, no. 3, pp. 1–14, 2014.

[32] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, and J. Tian, "Robust feature matching for remote sensing image registration via locally linear transforming," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6469–6481, 2015.

[33] J. Ma, J. Zhao, J. Jiang, and H. Zhou, "Non-rigid point set registration with robust transformation estimation under manifold regularization," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 4218–4224.

[34] J. Ma, J. Wu, J. Zhao, J. Jiang, H. Zhou, and Q. Z. Sheng, "Nonrigid point set registration with robust transformation learning under manifold regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 12, pp. 3584–3597, 2019.

[35] G. Wang and Y. Chen, "Scm: Spatially coherent matching with gaussian field learning for nonrigid point set registration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 203–213, 2021.

[36] W.-Y. Lin, F. Wang, M.-M. Cheng, S.-K. Yeung, P. H. Torr, M. N. Do, and J. Lu, "Code: Coherence based decision boundaries for feature correspondence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 34–47, 2017.

[37] X. Jiang, J. Ma, J. Jiang, and X. Guo, "Robust feature matching using spatial clustering with heavy outliers," *IEEE Trans. Image Process.*, vol. 29, pp. 736–746, 2020.

[38] X. Peng, Z. Huang, J. Lv, H. Zhu, and J. T. Zhou, "Comic: Multi-view clustering without parameter selection," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 5092–5101.

[39] X. Peng, H. Zhu, J. Feng, C. Shen, H. Zhang, and J. T. Zhou, "Deep clustering with sample-assignment invariance prior," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4857–4868, 2020.

[40] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, 2019.

[41] J. Ma, J. Jiang, H. Zhou, J. Zhao, and X. Guo, "Guided locality preserving feature matching for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4435–4447, 2018.

[42] X. Jiang, J. Jiang, A. Fan, Z. Wang, and J. Ma, "Multiscale locality and rank preservation for robust feature matching of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6462–6472, 2019.

[43] J.-W. Bian, W.-Y. Lin, Y. Liu, L. Zhang, S.-K. Yeung, M.-M. Cheng, and I. Reid, "GMS: Grid-based motion statistics for fast, ultra-robust feature correspondence," *Int. J. Comput. Vis.*, vol. 128, no. 6, pp. 1580–1593, 2020.

[44] J. Ma, X. Jiang, J. Jiang, J. Zhao, and X. Guo, “Lmr: Learning a two-class classifier for mismatch removal,” *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4045–4059, 2019.

[45] K. Moo Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, and P. Fua, “Learning to find good correspondences,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2666–2674.

[46] J. Zhang, D. Sun, Z. Luo, A. Yao, L. Zhou, T. Shen, Y. Chen, L. Quan, and H. Liao, “Learning two-view correspondences and geometry using order-aware network,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 5845–5854.

[47] W. Sun, W. Jiang, E. Trulls, A. Tagliasacchi, and K. M. Yi, “Acne: Attentive context normalization for robust permutation-equivariant learning,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11 286–11 295.

[48] P. J. Besl and N. D. McKay, “Method for registration of 3-d shapes,” in *Sensor Fusion IV: Control Paradigms and Data Structures*, 1992, pp. 586–606.

[49] T. Cour, P. Srinivasan, and J. Shi, “Balanced graph matching,” in *Adv. Neural Inf. Process. Syst.*, 2007, pp. 313–320.

[50] M. Leordeanu, M. Hebert, and R. Sukthankar, “An integer projected fixed point method for graph matching and map inference,” in *Adv. Neural Inf. Process. Syst.*, 2009, pp. 1114–1122.

[51] J. Yan, M. Cho, H. Zha, X. Yang, and S. M. Chu, “Multi-graph matching via affinity optimization with graduated consistency regularization,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 6, pp. 1228–1242, 2015.

[52] M. Leordeanu and M. Hebert, “A spectral technique for correspondence problems using pairwise constraints,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, pp. 1482–1489.

[53] H. Liu and S. Yan, “Common visual pattern discovery via spatially coherent correspondences,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 1609–1616.

[54] D. S. Grebenkov and B.-T. Nguyen, “Geometrical structure of laplacian eigenfunctions,” *SIAM Review*, vol. 55, no. 4, pp. 601–667, 2013.

[55] R. Courant and D. Hilbert, *Methods of Mathematical Physics: Partial Differential Equations*. John Wiley & Sons, 2008.

[56] M. Reed, *Methods of modern mathematical physics: Functional analysis*. Elsevier, 2012.

[57] T. J. Sullivan, *Introduction to uncertainty quantification*. Springer, 2015, vol. 63.

[58] J. Ma, J. Zhao, and A. L. Yuille, “Non-rigid point set registration by preserving global and local structures,” *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 53–64, 2016.

[59] M. B. Stegmann, B. K. Ersboll, and R. Larsen, “Fame-a flexible appearance modeling environment,” *IEEE Trans. Medical Imaging*, vol. 22, no. 10, pp. 1319–1331, 2003.

[60] B. Jian and B. C. Vemuri, “Robust point set registration using gaussian mixture models,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1633–1645, 2010.

Aoxiang Fan received the B.S. degree from the Electronic Information School, Wuhan University, Wuhan, China, in 2018. He is currently pursuing the master’s degree with the Multi-Spectral Vision Processing Lab, Wuhan University. His current research interests include computer vision and pattern recognition.

Xingyu Jiang received the B.E. degree from the Department of Mechanical and Electronic Engineering, Huazhong Agricultural University, Wuhan, China, in 2017, and the M.S. degree from the Electronic Information School, Wuhan University, Wuhan, China, in 2019. He is currently a Ph.D. student with the Electronic Information School, Wuhan University. His research interests include computer vision, machine learning, and pattern recognition.

Yong Ma graduated from the Department of Automatic Control, Beijing Institute of Technology, Beijing, China, in 1997. He received the Ph.D. degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2003. His general field of research is in signal and systems. His current research projects include remote sensing of the Lidar and infrared, as well as Infrared image processing, pattern recognition, interface circuits to sensors and actuators. Between 2004 and 2006, he was a Lecturer at the University of the West of England, Bristol, U.K. Between 2006 and 2014, he was with the Wuhan National Laboratory for Optoelectronics, HUST, Wuhan, where he was a Professor of electronics. He is now a Professor with the Electronic Information School, Wuhan University.

Xiaoguang Mei received the B.S. degree in communication engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2007, the M.S. degree in communications and information systems from Central China Normal University, Wuhan, in 2011, and the Ph.D. degree in circuits and systems from the HUST, in 2016. From 2010 to 2012, he was a Software Engineer with the 722 Research Institute, China Shipbuilding Industry Corporation, Wuhan. He is currently an associate professor with Wuhan University. His research interests include hyperspectral image processing, image fusion and machine learning.

Jiayi Ma received the B.S. degree in information and computing science and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively. He is currently a Professor with the Electronic Information School, Wuhan University. He has authored or co-authored more than 200 refereed journal and conference papers, including IEEE TPAMI/TIP, IJCV, CVPR, ICCV, ECCV, *etc*. His research interests include computer vision, machine learning, and remote sensing. Dr. Ma has been identified in the 2020 and 2019 Highly Cited Researcher lists from the Web of Science Group. He is an Area Editor of *Information Fusion*, an Associate Editor of *Neurocomputing*, *Sensors* and *Entropy*, and a Guest Editor of *Remote Sensing*.