Robust Point Matching via Vector Field Consensus

Jiayi Ma, Ji Zhao, Jinwen Tian, Alan L. Yuille, and Zhuowen Tu

***Abstract*— In this paper, we propose an efficient algorithm, called vector field consensus, for establishing robust point correspondences between two sets of points. Our algorithm starts by creating a set of putative correspondences which can contain a very large number of false correspondences, or outliers, in addition to a limited number of true correspondences (inliers). Next, we solve for correspondence by interpolating a vector field between the two point sets, which involves estimating a consensus of inlier points whose matching follows a nonparametric geometrical constraint. We formulate this a maximum *a posteriori* (MAP) estimation of a Bayesian model with hidden/latent variables indicating whether matches in the putative set are outliers or inliers. We impose nonparametric geometrical constraints on the correspondence, as a prior distribution, using Tikhonov regularizers in a reproducing kernel Hilbert space. MAP estimation is performed by the EM algorithm which by also estimating the variance of the prior model (initialized to a large value) is able to obtain good estimates very quickly (e.g., avoiding many of the local minima inherent in this formulation). We illustrate this method on data sets in 2D and 3D and demonstrate that it is robust to a very large number of outliers (even up to 90%). We also show that in the special case where there is an underlying parametric geometrical model (e.g., the epipolar line constraint) that we obtain better results than standard alternatives like RANSAC if a large number of outliers are present. This suggests a two-stage strategy, where we use our nonparametric model to reduce the size of the putative set and then apply a parametric variant of our approach to estimate the geometric parameters. Our algorithm is computationally efficient and we provide code for others to use it. In addition, our approach is general and can be applied to other problems, such as learning with a badly corrupted training data set.**

***Index Terms*— Point correspondence, outlier removal, matching, regularization.**

I. Introduction

ESTABLISHING reliable correspondence between two images is a fundamental problem in computer vision and it is a critical prerequisite in a wide range of applications including structure-from-motion, 3D reconstruction, tracking, image retrieval, registration, and object recognition [22], [28], [34], [42], [47], [64]. In this paper, we formulate it as a matching problem between two sets of discrete points where each point is an image feature, extracted by a feature detector, and has a local image descriptor (e.g., SIFT [31] or shape context [4]). The matching problem is ill-posed and is typically regularized by imposing two types of constraints: (i) a descriptor *similarity constraint*, which requires that points can only match points with similar descriptors, and (ii) *geometric constraint*, which requires that the matches satisfy an underlying geometrical requirement, which can be either parametric (e.g., rigid transformations) or non-parametric (e.g., non-rigid). Even after regularization there remain an exponential number of possible matches between the two sets and efficient algorithms are required to obtain the best solution by removing the false matches. The difficulty of the matching problem is typically made harder by the presence of unmatched points in the two images (due to occlusion or failures of the feature detectors).

A popular strategy for solving the matching problem is to use a two stage process. In the first stage, a set of *putative correspondences* are computed by using a similarity constraint to reduce the set of possible matches. This putative correspondence set typically includes most of the true matches, the *inliers*, but also a large number of false matches, or *outliers*, due to ambiguities in the similarity constraints (particularly if the images contain repetitive patterns). The second stage is designed to remove the outliers and estimate the inliers and the geometric parameters [18], [26], [35], [49]. This strategy is commonly used for situations where the geometrical constraints are parametric, such as requiring that corresponding points lie on epipolar lines [22]. Examples of this strategy include the RANSAC algorithm [18] and analogous robust hypothesize-and-verify methods [13], [42], [49]. Although these methods are very successful in many situations they have had limited success if the geometrical constraints are non-parametric, for example if the real correspondence is non-rigid, and they also tend to degrade badly if the proportion of outliers in the putative correspondence set becomes large [26].

In this paper we address these limitations by formulating the point matching problem as robust vector field interpolation using a non-parametric geometrical constraint. As discussed in the background section, vector flow interpolation arises frequently in computer vision and machine learning. Regularization theory [57] provides a framework for estimating vector fields when the problems are ill-posed. Yuille and Grzywacz [64], [65] formulated the discrete motion matching task in terms of finding those matches which give rise to the best interpolated vector field and subsequent work by Rangarajan and colleagues [12], [21] applied this to shape matching. Poggio and his collaborators [41] formulated learning in terms

Manuscript received July 12, 2013; revised October 31, 2013; accepted January 14, 2014. Date of publication February 20, 2014; date of current version March 4, 2014. This work was supported in part by the National Natural Science Foundation of China under Grant 61273279, in part by NSF under Grant 0917141, and in part by NIH under Grant 5R01EY022247-03. The work of Z. Tu was supported in part by NSF under Grant IIS-1216528 and in part by NSF CAREER under Award IIS-0844566. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Anthony Vetro.

J. Ma and J. Tian are with the National Key Laboratory of Science and Technology on Multi-Spectral Information Processing, School of Automation, Huazhong University of Science and Technology, Hubei 430074, China (e-mail: jyma2010@gmail.com; jwtian@hust.edu.cn).

J. Zhao is with Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: zhaoji84@gmail.com).

A. L. Yuille is with the Department of Statistics, University of California at Los Angeles, Los Angeles, CA 90095 USA (e-mail: yuille@stat.ucla.edu).

Z. Tu is with the Department of Cognitive Science, University of California at San Diego, La Jolla, CA 92697 USA (e-mail: zhuowen.tu@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2014.2307478

of interpolating a vector field from a discrete set of training samples (see also [37]), and other related machine learning work includes Gaussian processes [1], [8], [43].

Vector field interpolation assigns each position $\mathbf{x} \in \mathbb{R}^P$ (e.g., in one image) to a vector $\mathbf{y} \in \mathbb{R}^D$ defined by a vector-valued function $\mathbf{f}$, hence specifying a mapping $\mathbf{x} \mapsto \mathbf{f}(\mathbf{x})$ between two images. The problem of vector field interpolation is to fit a vector field $\mathbf{f}$ which interpolates a given sparse sample set $S = \{(\mathbf{x}_n, \mathbf{y}_n) : n \in \mathbb{N}_N\}$, i.e., $\forall n \in \mathbb{N}_N$, $\mathbf{y}_n = \mathbf{f}(\mathbf{x}_n)$. In this paper, we define *robust vector field* interpolation to be the spacial case where the sparse sample set S contains a large number of outliers which must be removed. We formulate this by a mixture model by introducing explicit latent/hidden variables for all members of the sample set which identifies/rejects the outliers and imposing a prior on the geometry which imposes a non-parametric smoothness constraint on the vector fields [64]. This leads to a maximum a posteriori (MAP) estimation problem which risks having many local minima which an algorithm may get trapped in. To address this issue, we use the EM algorithm [17] to estimate the variance of the prior, while simultaneously estimating the outliers, and give the variance a large initial value. This is conceptually similar to deterministic annealing [63], which uses the solution of an easy (e.g., smoothed) problem to recursively give initial conditions to increasingly harder problems, but differs in several respects (e.g., by not requiring any annealing schedule). Our method is computationally attractive and able to deal with a significant amount (up to 90%) of outliers.

To illustrate the main ideas of this paper, we show a simple example in Fig. 1. Given two sets of interest points extracted from an image pair, we want to match them to establish their point-wise correspondence. We first compute a set of putative correspondences based on their SIFT features as shown in Fig. 1(a), where the blue and red lines denote inliers and outliers respectively (we only show a subset of 50 members of the putative set). The input $\mathbf{x} \in \mathcal{R}^2$ denotes the location vectors, and the output $\mathbf{y} \in \mathcal{R}^2$ is the displacement vectors (or disparity) at that location; then the putative correspondences are displayed by motion field samples, as shown in Fig. 1(b). The inliers are shown in Fig. 1(c). If we use a recent interpolation method [37], which does not use an outlier process, we obtain the motion fields in Fig. 1(d) and (e) from the correspondences shown in Fig. 1(b) and (c) respectively. Clearly, the motion field in Fig. 1(d) is inaccurate because it is contaminated by the outliers in the putative set. Hence our task is to recover the motion field from the samples by removing the outliers – in other words to get Fig. 1(b)–(e). We note that our method can fail if the inliers of the putative set do not obey the smoothness assumption we impose [59], [64] (e.g., suppose the "true matches" are not indicated by the blue arrows and instead correspond to a subset of the red arrows). Interestingly, we demonstrate that we obtain very good results using our method even for cases where the underlying motion is rigid and parametric (e.g., cases addressed by RANSAC) and, in particular, we perform better than RANSAC if the putative set contains a large proportion of outliers.

Our contributions in this paper include the following. Firstly, we present an algorithm for determining point correspondences between a pair of 2D or 3D images. Unlike some standard methods, for example RANSAC, we do not assume an underlying parametric geometrical constraint (e.g., epipolar lines) but instead use a more flexible, non-rigid, non-parametric constraint. This greatly increases the generality of our approach and makes it robust to an extremely large amount of outliers – up to 90% of the putative set. Secondly, we also study a variant of our model which uses parametric constraints which we show is also more robust than RANSAC when the proportion of outliers is not so large (e.g., less than 50%). This parametric variant can be used in a two-stage process where we first use our non-parametric method to estimate an *reduced putative set*, by first removing many/most of the outliers in the putative set and then run our parameter method on this reduced set to directly output the parameters (assuming we want to estimate them). Thirdly, our approach can be used for robust learning where the dataset of training examples is contaminated by outliers. We illustrate this by a simple toy example, because it gives some insight in our approach, but we do not explore this application in this paper. This article is an extension of our earlier work [67], and the primary new contributions are an expanded derivation as well as comprehensive evaluations with more discussions and analysis.

The rest of the paper is organized as follows. Section II describes background material and related work. Section III

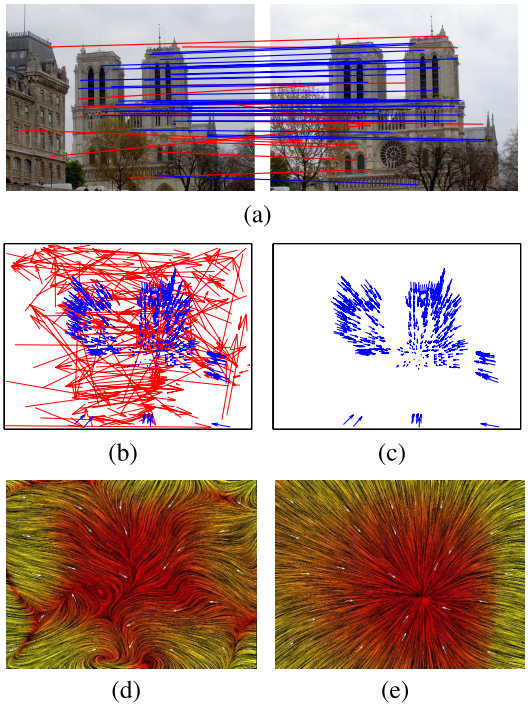

Fig. 1. Robust Vector field interpolation. (**a**) An image pair and its putative correspondences. Blue and red lines represent inliers and outliers respectively. For visibility, only 50 randomly selected elements of the putative set are shown. (**b**) and (**c**) Motion field samples generated by all putative correspondences and only inliers respectively. The head and tail of each arrow correspond to the positions of feature points in two images. (**d**) and (**e**) The interpolated vector field using samples from (**b**) and (**c**) respectively. The visualization method is line integral convolution (LIC) [9], color indicates the magnitude of the displacement at each point. Best viewed in color.

describes our vector field consensus algorithm for interpolation which is robust to a high proportion of outliers. In Section IV, we discuss how to apply our algorithm to the point matching problem with either non-parametric or parametric geometric constraints. Section V illustrates our algorithm on a synthetic learning task and then tests it for motion correspondence on several datasets with comparisons to other approaches, followed by some concluding remarks in Section VI.

II. Related Work

This section briefly reviews the background material that our work is based on. This includes methods for establishing a set of putative correspondences and geometric constraints. Next we discuss approaches for solving matching problems which solve for a correspondence matrix between point sets. Then we discuss vector field interpolation.

A. Establishing Point Correspondences Using Putative Sets and Geometric Constraints

A popular strategy [42] for establishing reliable point correspondences between image pairs involves two steps: (i) computing a set of putative correspondences, and (ii) then removing the outliers using geometrical constraints. In the first step, putative correspondences are obtained by pruning the set of all possible point correspondence by computing feature descriptors at the points and removing the matches between points whose descriptors are too dissimilar. The types of descriptors used include as SIFT [31] and shape contexts [4] in the 2D case, and spin image [24], MeshDOG and MeshHOG [66] in the 3D case. In the second step, robust estimators typically based on parametric geometrical models (e.g., rigidity constraints) are used to detect and remove the outliers.

There has been considerable study of robust estimation in the statistics literature [23], [46]. This work shows, for example, that maximum likelihood estimator of parameters using quadratic L_2 norms are not-robust and highly sensitive to outliers. By contrast, methods which minimize L_1 norm are more robust and capable of resisting a larger proportion of outliers. A particularly robust method is the redescending M-estimator [23]. It can be shown that this estimator results from using an explicit variable to indicate whether data is an outlier or an inlier (this indicator variable must be estimated) [19]. The use of explicit variable to indicate outliers has a long history in computer vision [7], [19], [20] and for signal processing methods like robust PCA [61]. We return to the use of explicit outlier variables in the next section.

The RANSAC algorithm matches two point sets by first computing a putative set and then using robust methods to impose parametric geometric constraints [18]. RANSAC uses a hypothesize-and-verify framework. It proceeds by repeatedly generating solutions estimated from a small set of correspondences randomly selected from the data, and then tests each solution for support from the complete set of putative correspondences. RANSAC has several variants such as MLESAC [49], LO-RANSAC [14] and PROSAC [13]. MLESAC adopts a new cost function using a weighted voting strategy based on M-estimation and chooses the solution that maximizes the likelihood rather than the inlier count. RANSAC is also enhanced in LO-RANSAC with a local optimization step based on how well the measurements satisfy the current best hypothesis. Alternatively, prior beliefs are assumed in PROSAC about the probability of a point being an inlier to modify the random sampling step of the RANSAC. A detailed comparative analysis of RANSAC techniques can be found in [42]. Recently, some new non-parametric model-based methods have also been developed, such as identifying point correspondences by correspondence function (ICF) [26]. It uses support vector regression to learn a correspondence function pair which maps points in one image to their corresponding points in another, and then rejects the outliers by checking whether they are consistent with the estimated correspondence functions.

Another strategy for point correspondences is to formulate this problem in terms of a correspondence matrix between points together with a parametric, or non-parametric, geometric constraint [5], [12], [21], [39]. These approaches relate closely to earlier work on mathematical models of human perception of long-range motion. This includes Ullman's minimal mapping theory [53] and Yuille and Grzywacz's motion coherence theory [65] which formulate correspondence in terms of vector field interpolation and use Gaussian kernels. We note that these types of models give accurate prediction for human perception of long range motion [32].

These methods typically involve a two step update process which alternates between the correspondence and the (rigid/non-rigid) transformation estimation. The iterated closest point (ICP) algorithm [5] is one of the best known point registration approaches. It exploits nearest-neighbor relationships to assign a binary correspondence, and then uses estimated correspondence to refine the transformation. Rangarajan and colleagues [12], [21] established a general framework for estimating correspondence and transformations for point matching, building on Yuille and Grzywacz's work [65]. Specifically, for the non-rigid case, they modeled the transformation as a thin-plate spline and did robust point matching by an algorithm (TRS-RPM) which involves deterministic annealing and soft-assignment. Alternatively, the coherence point drift (CPD) algorithm [39] uses Gaussian radial basis functions instead of thin-plate splines (this corresponds to a different type of regularizer, see next section). In these formulations, both the rigid and non-rigid cases can be dealt with, but these methods usually cannot tolerate large numbers of outliers and searching over all possible matches is in general NP-hard. Some robustness can be achieved by paying a penalty for unmatched points.

Point correspondence has also been formulated as a graph matching problem, such as the dual decomposition (DD) [50], Spectral Matching (SM) [25], and graph shift (GS) [29], [30]. The DD approach formulates the matching task as an energy minimization problem by defining a complex objective function of the appearance and the spatial arrangement of the features, and then minimizes this function based on the dual decomposition approach. The SM method uses an efficient spectral method for finding consistent correspondences

between two sets of features. Based on the SM method, the GS method constructs an affinity graph for the correspondences, and the maximal clique of the graph is viewed as spatially coherent correspondences. Besides, Cho and Lee [11] introduced novel progressive framework which combines probabilistic progression of graphs with matching of graphs. The SIFT-flow algorithm [28] builds a dense correspondence map between two arbitrary images with a particular advantage for matching two scenes; it does not explicitly deal with the outliers and may not be able to produce the accuracy for the precise matching for problems like structure-from-motion. Note that this type of graph matching formulation can in some cases be mathematically equivalent to the methods with correspondence variables and geometric constraints [63], [65].

B. Vector Field Interpolation

A classical problem of vector field interpolation is to measure dense motion (velocity) fields. In this specific context, a number of methods have been developed based on regularization theory [16], [65]. Corpetti *et al*. [16] proposed an optical-flow technique specifically dedicated to estimating fluid flows from image sequences. Yuille and Grzywacz [65] introduced the motion coherence theory for computing a velocity field defined in an image. They used a quadratic regularizer to impose geometric constraints on the correspondences, and showed that this was equivalent to formulating the problem in terms of a space of kernels. These methods usually do not consider the interactions among the $\mathbf{x}$ and $\mathbf{y}$ components of the fields. But recently Micchelli and Pontil [37] developed a framework of regularization in the RKHS of vector-valued functions, which can directly encode relationships between the components of vector fields by choosing a suitable matrix-valued kernel. Based on their work, Baldassarre *et al*. [3] investigated a spectral regularization scheme to interpolate vector fields. See also Wu *et al*. [60] who developed different regularizers and kernels for different types of motion fields and investigated their relation to human perception of motion flow. In addition, Lin *et al*. [27] proposed a novel semi-supervised multi-task learning formulation using vector fields. These methods however ignore the robustness issue, in which the presence of outliers in the dataset may greatly degrade the performance.

The technique of robust vector field interpolation has been adopted in Gaussian processes, basically by using the so-called t-processes [62], [68]. The t-processes are inherently robust to outliers and can be seen as a robust extension of Gaussian processes, in which the priors of the function values are sampled from a (heavy-tailed) multivariate t distribution. In this work, we introduce a novel robust vector field interpolation method; our approach tries to associate each sample with a latent variable indicating whether it is an inlier for purpose of robust estimation.

III. The Vector Field Consensus Algoroithm

This section describes the vector field consensus algorithm (the next section discusses how to apply it to point matching). We start by briefly introducing the interpolation problem, and then lay out the formulation of our robust vector field interpolation and derive an EM solution by using a regularized kernel method. We subsequently discuss some potentially useful matrix-valued kernels for vector field interpolation, and followed by the fast implementation. Finally, we analyze the computational complexity of the proposed approach.

A. Interpolation by Regularization

Assume a set of observed input-output pairs $S = \{(\mathbf{x}_n, \mathbf{y}_n) \in \mathcal{X} \times \mathcal{Y} : n \in \mathbb{N}_N\}$ that are samples randomly drawn from a vector field, where $\mathcal{X} \subseteq \mathbb{R}^P$ and $\mathcal{Y} \subseteq \mathbb{R}^D$ are input and output space respectively. The goal is to fit a mapping $\mathbf{f}$ interpolating the sample set, i.e., $\forall n \in \mathbb{N}_N, \mathbf{y}_n = \mathbf{f}(\mathbf{x}_n)$. This problem is in general ill-posed since it has an infinite number of solutions. To obtain a meaningful solution, it can one way be formulated into an optimization problem with a certain choice of regularization [3], [41], which typically operates in a vector-valued Reproducing Kernel Hilbert Space (RKHS) [2] (associated with a particular kernel), as described in detail in Appendix A. Specifically, the Tikhonov regularization [48] in an RKHS $\mathcal{H}$ defined by a matrix-valued kernel $\Gamma : \mathbb{R}^P \times \mathbb{R}^P \to \mathbb{R}^{D\times D}$ minimizes a regularized risk functional

$$\mathcal{E}(\mathbf{f}) = \min_{\mathbf{f}\in\mathcal{H}} \left\{ \sum_{n=1}^{N} \|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2 + \lambda \|\mathbf{f}\|_{\mathcal{H}}^2 \right\}, \tag{1}$$

where the first term is the empirical error (risk) which enforces closeness to the data, the second term is a stabilizer which enforces smoothness to the vector field $\mathbf{f}$, λ is a regularization constant controlling the trade-off between these two terms, and $\|\cdot\|_{\mathcal{H}}$ denotes the norm of $\mathcal{H}$.

According to the representer theorem [37], the solution of the regularized risk functional (1) is given by

$$\mathbf{f}(\mathbf{x}) = \sum_{n=1}^{N} \Gamma(\mathbf{x}, \mathbf{x}_n)\mathbf{c}_n, \quad \mathbf{c}_n \in \mathcal{Y}, \tag{2}$$

with the coefficient set $\{\mathbf{c}_n : n \in \mathbb{N}_N\}$ determined by a linear system

$$(\widetilde{\Gamma} + \lambda \mathbf{I})\widetilde{\mathbf{C}} = \widetilde{\mathbf{Y}}, \tag{3}$$

where the Gram matrix $\widetilde{\Gamma} \in \mathbb{R}^{DN\times DN}$ is an $N \times N$ block matrix with the (i, j)-th block $\Gamma(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{I}$ is an identity matrix, $\widetilde{\mathbf{Y}} = (\mathbf{y}_1^{\mathrm{T}}, \ldots, \mathbf{y}_N^{\mathrm{T}})^{\mathrm{T}}$ and $\widetilde{\mathbf{C}} = (\mathbf{c}_1^{\mathrm{T}}, \ldots, \mathbf{c}_N^{\mathrm{T}})^{\mathrm{T}}$ are column vectors.

B. Problem Formulation

The Tikhonov regularization treats all samples as inliers, which may be problematic if there are outliers. Hence we assume that the given sample set S may contain some amount of unknown outliers. Our purpose is to fit a vector field $\mathbf{f} : \mathcal{X} \to \mathcal{Y}$ interpolating the inliers, and consequently distinguish inliers from the outliers.

Due to the existence of outliers, it is desirable to have a robust estimation of $\mathbf{f}$. To this end, we make the assumption that, for the inliers, the noise is Gaussian on each component

with zero mean and uniform standard deviation σ; for the outliers, the output space is a bounded region of $\mathbb{R}^D$, and the distribution is assumed to be uniform $\frac{1}{a}$, where a is just a constant (the volume of this region). We then associate the n-th sample with a latent variable $z_n \in \{0, 1\}$, where $z_n = 1$ indicates a Gaussian distribution and $z_n = 0$ points to a uniform distribution. Let $\mathbf{X}$ and $\mathbf{Y}$ be the set of observed input and output data, in which the n-th rows represent $\mathbf{x}_n^{\mathrm{T}}$ and $\mathbf{y}_n^{\mathrm{T}}$. Thus, the likelihood is a mixture model given by

$$p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) = \prod_{n=1}^{N} \sum_{z_n} p(\mathbf{y}_n, z_n|\mathbf{x}_n, \boldsymbol{\theta}) = \prod_{n=1}^{N} \left(\frac{\gamma}{(2\pi\sigma^2)^{D/2}} e^{-\frac{\|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2}{2\sigma^2}} + \frac{1-\gamma}{a} \right), \quad (4)$$

where $\boldsymbol{\theta} = \{\mathbf{f}, \sigma^2, \gamma\}$ includes a set of unknown parameters, γ is the mixing coefficient specifying the marginal distribution over the latent variable, i.e., $\forall z_n$, $p(z_n = 1) = \gamma$. Note that the uniform distribution function is nonzero only in a bounded region (here we omit the indicator function for clarity).

We want to recover the vector field $\mathbf{f}$ from the data S. Taking a probabilistic approach, we assume $\mathbf{f}$ to be a realization of a random field with a known prior probability distribution $p(\mathbf{f})$. The prior is used to impose constraints on $\mathbf{f}$, assigning significant probability only to those functions that satisfy those constraints. We consider the slow-and-smooth model [59], [64] which has been shown to account for a range of motion phenomena, the prior of $\mathbf{f}$ then has the form:

$$p(\mathbf{f}) \propto e^{-\frac{\lambda}{2}\phi(\mathbf{f})}, \quad (5)$$

where $\phi(\mathbf{f})$ is a smoothness functional and λ a positive real number (we will discuss the details of $\mathbf{f}$ later).

Note that flat priors are also implicitly assumed on σ^2 and γ. Using Bayes rule, we estimate a MAP solution of $\boldsymbol{\theta}$, i.e., $\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Y}) = \arg\max_{\boldsymbol{\theta}} p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) p(\mathbf{f})$. This is equivalent to seeking the minimal energy

$$E(\boldsymbol{\theta}) = -\ln p(\mathbf{f}) - \sum_{n=1}^{N} \ln \sum_{z_n} p(\mathbf{y}_n, z_n|\mathbf{x}_n, \boldsymbol{\theta}). \quad (6)$$

The vector field $\mathbf{f}$ will be directly obtained from the optimal solution $\boldsymbol{\theta}^*$, and the latent variables $\{z_n : n \in \mathbb{N}_N\}$ determine the inliers. In the next section, we show how to solve the estimation problem using an EM approach.

C. The EM Algorithm

There are several ways to estimate the parameters of the mixture model, such as EM algorithm, gradient descent, and variational inference. The EM algorithm [17] is a general technique dealing with the existence of latent variables. It alternates with two steps: an expectation step (E-step) and a maximization step (M-step).

We follow standard notations [6] and omit some terms that are independent of $\boldsymbol{\theta}$. Considering the negative log posterior function, i.e. Eq. (6), the complete-data log posterior is

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\mathrm{old}}) = & -\frac{1}{2\sigma^2} \sum_{n=1}^{N} P(z_n = 1|\mathbf{x}_n, \mathbf{y}_n, \boldsymbol{\theta}^{\mathrm{old}}) \|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2 \\ & - \frac{D}{2} \ln \sigma^2 \sum_{n=1}^{N} P(z_n = 1|\mathbf{x}_n, \mathbf{y}_n, \boldsymbol{\theta}^{\mathrm{old}}) \\ & + \ln(1-\gamma) \sum_{n=1}^{N} P(z_n = 0|\mathbf{x}_n, \mathbf{y}_n, \boldsymbol{\theta}^{\mathrm{old}}) \\ & + \ln \gamma \sum_{n=1}^{N} P(z_n = 1|\mathbf{x}_n, \mathbf{y}_n, \boldsymbol{\theta}^{\mathrm{old}}) - \frac{\lambda}{2}\phi(\mathbf{f}). \quad (7) \end{aligned}$$

E-step: We use the current parameter values $\boldsymbol{\theta}^{\mathrm{old}}$ to find the posterior distribution of the latent variables. Denote $\mathbf{P} = \mathrm{diag}(p_1, \ldots, p_N)$ a diagonal matrix, where $p_n = P(z_n = 1|\mathbf{x}_n, \mathbf{y}_n, \boldsymbol{\theta}^{\mathrm{old}})$ can be computed by applying Bayes rule:

$$p_n = \frac{\gamma\, e^{-\frac{\|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2}{2\sigma^2}}}{\gamma\, e^{-\frac{\|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2}{2\sigma^2}} + (1-\gamma)\frac{(2\pi\sigma^2)^{D/2}}{a}}, \quad (8)$$

The posterior probability p_n is a soft decision, which indicates to what degree the n-th sample agrees with the current estimated vector field $\mathbf{f}$.

M-step: We determine the revised parameter estimate $\boldsymbol{\theta}^{\mathrm{new}}$ as follows: $\boldsymbol{\theta}^{\mathrm{new}} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\mathrm{old}})$. Considering P is a diagonal matrix and taking derivative of $\mathcal{Q}(\boldsymbol{\theta})$ with respect to σ^2 and γ, and setting them to zero, we obtain

$$\sigma^2 = \frac{(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{V}})^{\mathrm{T}} \widetilde{\mathbf{P}} (\widetilde{\mathbf{Y}} - \widetilde{\mathbf{V}})}{D \cdot \mathrm{tr}(\mathbf{P})}, \quad (9)$$

$$\gamma = \mathrm{tr}(\mathbf{P})/N, \quad (10)$$

where $\widetilde{\mathbf{V}} = (\mathbf{f}(\mathbf{x}_1)^{\mathrm{T}}, \ldots, \mathbf{f}(\mathbf{x}_N)^{\mathrm{T}})^{\mathrm{T}}$, $\widetilde{\mathbf{P}} = \mathbf{P} \otimes \mathbf{I}_{D\times D}$ with $\otimes$ denoting the Kronecker product, and $\mathrm{tr}(\cdot)$ is the trace.

Next we consider the terms of $\mathcal{Q}(\boldsymbol{\theta})$ that are related to $\mathbf{f}$. We obtain a regularized risk functional as [37]:

$$\mathcal{E}(\mathbf{f}) = \frac{1}{2\sigma^2} \sum_{n=1}^{N} p_n \|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2 + \frac{\lambda}{2}\phi(\mathbf{f}). \quad (11)$$

It is a special form of Tikhonov regularization, i.e. equation (1), and the first term could be seen as a weighted empirical error. Thus the maximization of $\mathcal{Q}$ with respect to $\mathbf{f}$ is equivalent to minimizing the regularized risk functional (11).

We model $\mathbf{f}$ by requiring it to lie within an RKHS $\mathcal{H}$ defined by a matrix-valued kernel $\Gamma : \mathbb{R}^P \times \mathbb{R}^P \to \mathbb{R}^{D\times D}$. For the smoothness functional $\phi(\mathbf{f})$, we use the square norm, i.e., $\phi(\mathbf{f}) = \|\mathbf{f}\|_{\mathcal{H}}^2$. Therefore, we have the following representer theorem [37].

Theorem 1: The optimal solution of the regularized risk functional (11) is given by equation (2) with the coefficient set $\{\mathbf{c}_n : n \in \mathbb{N}_N\}$ determined by a linear system

$$(\widetilde{\Gamma} + \lambda\sigma^2 \widetilde{\mathbf{P}}^{-1}) \widetilde{\mathbf{C}} = \widetilde{\mathbf{Y}}. \quad (12)$$

The proof is given in Appendix B. Once the EM algorithm converges, we then obtain a vector field $\mathbf{f}$. Besides, we have

Algorithm 1 The Vector Field Consensus Algorithm

Input: Sample set $S = \{(\mathbf{x}_n, \mathbf{y}_n) : n \in \mathbb{N}_N\}$, kernel $\mathbf{\Gamma}$, regularization constant λ, inlier threshold τ
Output: Vector field $\mathbf{f}$, consensus set (inliers) $\mathcal{I}$

```
Initialize γ, Ṽ = 0_{DN×1}, P = I_{N×N};
Set a to the volume of the output space;
Initialize σ^2 by Eq. (9);
Construct the Gram matrix Γ̃ using the definition of Γ;
repeat
    E-step:
      Update P = diag(p_1, ..., p_N) by Eq. (8);
    M-step:
      Update C̃ by solving linear system (12);
      Update Ṽ by using Eq. (2);
      Update σ^2 and γ by Eqs. (9) and (10);
until Q converges;
The vector field f is determined by Eq. (2);
The consensus set I is determined by Eq. (13).
```

the estimation of the inliers as well. Here we present two particular scenarios:

(i) with a predefined threshold τ, we obtain an inlier set $\mathcal{I}$

$$\mathcal{I} = \{n : p_n > \tau, n \in \mathbb{N}_N\}; \tag{13}$$

(ii) since we have recovered the vector field, we are then able to determine the inliers by checking whether they are consistent with $\mathbf{f}$.

In this paper, we use the first scenario. Moreover, we observe that in practice the posterior probabilities of the samples are mostly (over 99%) either smaller than 0.01 or larger than 0.99, after the EM iteration converges. Therefore, our method is not sensitive to the choice of τ. When such a hard decision is made, the set $\mathcal{I}$ is the so-called consensus set in RANSAC [18]. This is the reason we name our method *vector field consensus* (VFC). We summarize the VFC method in Algorithm 1.

Analysis of convergence. Note that the energy function (6) is not convex so it unlikely that any algorithm can find its global minimum. Our strategy is to initialize the variance σ^2 by a large initial value and then use the EM algorithm. At large sigma, the objective function will be convex in a large region surrounding the global minimum. Hence we are likely to find the global minimum for large variance. As sigma decreases, the position of the global minimum will tend to change smoothly. The objective function will be convex in a small region around its minimum, which makes it likely that using the old global minimum as initial value could converge to the new global minimum. Therefore, as the iteration proceeds, we have a good chance of reaching the global minimum. This is conceptually similar to deterministic annealing [63], which uses the solution of an easy (e.g., smoothed) problem to recursively give initial conditions to increasingly harder problems.

D. Matrix-Valued Kernels

Kernels play a central role in regularization theory as they provide a flexible and computationally feasible way to choose an RKHS. Next, we briefly review some potentially useful kernels which will be adopted and tested in the experiments.

Decomposable kernels. Baldassarre *et al.* [3] discussed a decomposable kernel for interpolating a vector field which has the form

$$\Gamma(\mathbf{x}, \mathbf{x}') = \kappa(\mathbf{x}, \mathbf{x}')\mathbf{A}, \tag{14}$$

where the scalar kernel κ (e.g., Gaussian kernel) encodes the similarity between the inputs, and the positive semidefinite $D \times D$ matrix $\mathbf{A}$ encodes the relationships between the outputs. The matrix-valued kernel can exploit the relationships among the components of the vector field.

Divergence-free and curl-free kernels. Important examples of divergence-free and curl-free fields are incompressible fluid flows and magnetic fields respectively. These kernels have been used in [36] to interpolate divergence-free or curl-free vector fields. The divergence-free kernel is

$$\Gamma_{df}(\mathbf{x}, \mathbf{x}') = \frac{1}{\widetilde{\sigma}^2} e^{-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\widetilde{\sigma}^2}} \left[\left(\frac{\mathbf{x}-\mathbf{x}'}{\widetilde{\sigma}}\right)\left(\frac{\mathbf{x}-\mathbf{x}'}{\widetilde{\sigma}}\right)^{\mathrm{T}} + \left((D-1) - \frac{\|\mathbf{x}-\mathbf{x}'\|^2}{\widetilde{\sigma}^2}\right) \cdot \mathbf{I} \right], \tag{15}$$

and the curl-free kernel is

$$\Gamma_{cf}(\mathbf{x}, \mathbf{x}') = \frac{1}{\widetilde{\sigma}^2} e^{-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\widetilde{\sigma}^2}} \left[\mathbf{I} - \left(\frac{\mathbf{x}-\mathbf{x}'}{\widetilde{\sigma}}\right)\left(\frac{\mathbf{x}-\mathbf{x}'}{\widetilde{\sigma}}\right)^{\mathrm{T}} \right], \tag{16}$$

where $\widetilde{\sigma}$ is the width of the Gaussian part of the kernels. Note that in these two kernels the dimensions of the input and output are the same, i.e. $P = D$. Observe that non-negative linear combinations of matrix-valued kernels still obey the kernel properties. Thus, we can interpolate a vector field and reconstruct its divergence-free and curl-free parts by taking a convex combination of these two matrix-valued kernels, controlled by a parameter $\widetilde{\alpha}$:

$$\Gamma_{\widetilde{\alpha}}(\mathbf{x}, \mathbf{x}') = (1 - \widetilde{\alpha})\Gamma_{df}(\mathbf{x}, \mathbf{x}') + \widetilde{\alpha}\Gamma_{cf}(\mathbf{x}, \mathbf{x}'). \tag{17}$$

E. Fast Implementation

Solving the vector field $\mathbf{f}$ merely requires to solve the linear system (12). However, for large values of N, it may pose a serious problem due to heavy computational (e.g. scales as $O(N^3)$) or memory (e.g. scales as $O(N^2)$) requirements, and, even when it is implementable, one may prefer a suboptimal but simpler method. In this section, we provide a fast implementation based on a similar kind of idea as the subset of regressors method [40].

Rather than searching for the optimal solution in $\mathcal{H}_N$, we use a sparse approximation and search a suboptimal solution in a space with much less basis functions defined as $\mathcal{H}_M = \left\{\sum_{m=1}^{M} \Gamma(\cdot, \widetilde{x}_m)\mathbf{c}_m : \mathbf{c}_m \in \mathcal{Y}\right\}$, and then minimize the regularized risk functional over all the sample data. Here $M \ll N$ and we choose the point set $\{\widetilde{\mathbf{x}}_m : m \in \mathbb{N}_M\}$ as a random subset of $\{\mathbf{x}_n : n \in \mathbb{N}_N\}$ according to [44] and [33]. There, it was found that simply selecting an arbitrary subset of the training inputs performs no worse than more sophisticated

methods. According to the sparse approximation, we search a solution with the form

$$\mathbf{f}(x) = \sum_{m=1}^{M} \Gamma(\mathbf{x}, \widetilde{\mathbf{x}}_m)\mathbf{c}_m, \quad \mathbf{c}_m \in \mathcal{Y}, \tag{18}$$

with the coefficients $\{\mathbf{c}_m : m \in \mathbb{N}_M\}$ determined by a linear system

$$(\widetilde{\mathbf{U}}^{\mathrm{T}}\widetilde{\mathbf{P}}\widetilde{\mathbf{U}} + \lambda\sigma^2\widetilde{\Gamma}_s)\widetilde{\mathbf{C}}_s = \widetilde{\mathbf{U}}^{\mathrm{T}}\widetilde{\mathbf{P}}\widetilde{\mathbf{Y}}, \tag{19}$$

where $\widetilde{\mathbf{C}}_s = (\mathbf{c}_1^{\mathrm{T}}, \ldots, \mathbf{c}_M^{\mathrm{T}})^{\mathrm{T}}$ is the coefficient vector, $\widetilde{\Gamma}_s$ is an $M \times M$ block Gram matrix with the (i, j)-th block $\Gamma(\widetilde{\mathbf{x}}_i, \widetilde{\mathbf{x}}_j)$, $\widetilde{\mathbf{U}}$ is an $N \times M$ block matrix with the (i, j)-th block $\Gamma(\mathbf{x}_i, \widetilde{\mathbf{x}}_j)$.

In contrast to the optimal solution given by the representer theorem, which is a linear combination of the basis functions $\{\Gamma(\cdot, \mathbf{x}_n) : n \in \mathbb{N}_N\}$, the suboptimal solution is formed by a linear combination of arbitrary M-tuples of the basis functions. Generally, this sparse approximation will yield a vast increase in speed and decrease in memory requirements with negligible decrease in accuracy. We call this implementation *SparseVFC*. Compared with the VFC algorithm shown in Algorithm 1, SparseVFC solves a different linear system (19) in Line 9.

F. Computational Complexity

For the VFC algorithm, the corresponding Gram matrix is of size $DN \times DN$. Because of the representer theorem, it needs to solve a linear system (12) to estimate the vector field $\mathbf{f}$. The time complexity is $O(D^3N^3)$ and it is the most time-consuming step in the algorithm. As a result, the total time complexity of the VFC algorithm is $O(mD^3N^3)$, where m is the number of EM iterations. In our current implementation, we just use the MATLAB "\" operator, which implicitly uses Cholesky decomposition to invert a matrix. The space complexity of VFC scales like $O(D^2N^2)$ due to the memory requirements for storing the Gram matrix $\widetilde{\Gamma}$.

For SparseVFC, the corresponding Gram matrix is of size $DM \times DM$, where M is the number of basis functions used for sparse representation. Then the time complexity is reduced to $O(mD^3M^2N + mD^3M^3)$, and the space complexity is reduced to $O(D^2MN + D^2M^2)$. Due to $M \ll N$, the time and space complexities can be written as $O(mD^3M^2N)$ and $O(D^2MN)$. Our experiments demonstrate that SparseVFC is much faster than VFC with negligible performance degradation.

IV. Establishing Point Matching Using VFC

This section describes how we can apply the vector field consensus algorithm to the problem of establishing matches between two sets of points. Our strategy is to construct a putative set of matches by considering all possible matches (between points in the two sets) and rejecting matches between points whose feature descriptor vectors (e.g., SIFT or shape context) are sufficiently different. This putative set typically contains many false matches (outliers), in addition to a small number of true matches (inliers), and hence it is important that the VFC algorithm is highly robust to outliers.

We also address the issue of using parametric and non-parametric geometrical constraints. The non-parametric constraints are more general and allow slow-and-smooth motion fields, while the parametric constraints impose stronger constraints based on rigidity of motion (e.g., the epipolar line constraint). We discuss why there is a relationship between slow-and-smooth and rigid motion, which justifies applying the slow-and-smooth model (described in the last section) to cases where the motion is rigid. In addition, we formulate a variant of the VFC algorithm which uses parametric models.

A. Vector Field Introduced by Point Correspondences

We first establish a set of putative correspondences by considering all matches between the two point sets and then removing matches between points whose feature descriptors are above threshold. Each member of the putative set is comprised of a pair $(\mathbf{u}_n, \mathbf{v}_n)$, where $\mathbf{u}_n$ and $\mathbf{v}_n$ are positions of the two points. The performance of point matching algorithms depends, typically, on the coordinate system in which points are expressed. We use data normalization to control for this. More specifically, we make a linear re-scaling of the correspondences so that the positions in the first and second point sets have zero mean and unit variance. Suppose the normalized correspondence is $(\hat{\mathbf{u}}_n, \hat{\mathbf{v}}_n)$; we convert it into a motion field sample by a transformation $(\hat{\mathbf{u}}_n, \hat{\mathbf{v}}_n) \rightarrow (\mathbf{x}_n, \mathbf{y}_n)$, where $\mathbf{x}_n = \hat{\mathbf{u}}_n$, $\mathbf{y}_n = \hat{\mathbf{v}}_n - \hat{\mathbf{u}}_n$.

B. Kernel Choice

By choosing different kernels, the norm in the corresponding RKHS encodes different notions of smoothness. Usually, for the correspondence problem, the structure of the generated motion field is relatively simple. A decomposable kernel with the form (14) is effective. For the scalar kernel κ, we choose a Gaussian kernel as $\kappa(\mathbf{x}_i, \mathbf{x}_j) = e^{-\beta\|\mathbf{x}_i - \mathbf{x}_j\|^2}$. For the relationship matrix $\mathbf{A}$, we find that empirically using an identity matrix works well. In this case we can solve a more efficient linear system instead of Eq. (12) as

$$(\mathbf{K} + \lambda\sigma^2\mathbf{P}^{-1})\mathbf{C} = \mathbf{Y}, \tag{20}$$

where the Gram matrix $\mathbf{K} \in \mathbb{R}^{N \times N}$ with $\mathbf{K}_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$, and $\mathbf{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_N)^{\mathrm{T}}$ is a matrix of size $N \times D$.

C. Applicability of The Method: Rigid and Non-Rigid Motion

Our basic approach assumes that the motion flow between the two datasets can be modelled non-parametrically which, in practice, requires imposing some type of smoothness, or slow-and-smoothness, constraint. This is a plausible assumption if the transformation between the transformation between the images is a homography or a non-rigid transformation. But it is less clear that this is a good assumption if the underlying transformation is rigid in three-dimensional space (e.g., the epipolar geometry constraint). In this situation, the motion flow may not be smooth if, for example, there are large depth discontinuities. But as we argue below, and our experimental results support, rigid motion in space often corresponds to slow-and-slow motion in the images.

A close relationship between rigid transformation in three-dimensions and slow-and-smooth motion in two dimensions

was shown by Ullman [53] and by [54]. They assumed plausible probability distributions about the rigid motion in three-dimensional (i.e. for the translation and rotation) and showed that the resulting projected motion in the image plane was typically slow-and-smooth. These analyses were performed to address the apparent paradox that humans appear to use slowness and smoothness to resolve ambiguities in matching points between images, but use rigidity assumption to estimate the structure of the moving objects.

Further evidence for a relation between slow-and-smooth and rigidity comes from the empirical studies of Roth and Black [45]. They analyzed the motion flow in images obtained by a camera moving in a fixed environment. They showed that the motion statistics were consistent with a variant of the slow-and-smooth model.

In summary, the analysis in this section shows that it is reasonable to apply our VFC algorithm even if the underlying motion is rigid. This is confirmed by our experimental results which also show that VFC gives better estimates of the motion flow (and for detecting the outliers in the putative set) compared to other approaches which exploit this rigid structure (e.g., RANSAC and the algorithm described in the next section).

D. Extension to Parametric Models

As mentioned in the last section, if the two datasets are related by an underlying rigid transformation then the VFC algorithm may not get the correct result if the image pair contains a large depth discontinuity. Nevertheless, our formulation can be easily extended to a parametric model, e.g., the fundamental matrix or homography. For the sake of clarity, in the following we present a parametric variant of our model to estimate the fundamental matrix alone.

Suppose we are given a set of putative correspondences $S = \{(\mathbf{u}_n, \mathbf{v}_n)\}_{n=1}^{N}$, where $\mathbf{u}$ and $\mathbf{v}$ are homogeneous image coordinates, i.e. $\mathbf{u} = (\mathbf{u}^x, \mathbf{u}^y, 1)^{\mathrm{T}}$. The epipolar line constraint is then represented by a fundamental matrix $\mathbf{F}$: $\mathbf{v}^{\mathrm{T}}\mathbf{F}\mathbf{u} = 0$. We assume Gaussian noise on inliers, i.e., $\mathbf{v}_n^{\mathrm{T}}\mathbf{F}\mathbf{u}_n \sim N(0, \sigma^2)$, and uniform distribution on outliers. By using the same notation as in our non-parametric formulation, we obtain a likelihood as:

$$p(S|\boldsymbol{\theta}) = \prod_{n=1}^{N}\left(\frac{\gamma}{\sqrt{2\pi\sigma^2}}e^{-\frac{(\mathbf{v}_n^{\mathrm{T}}\mathbf{F}\mathbf{u}_n)^2}{2\sigma^2}} + \frac{1-\gamma}{a}\right), \tag{21}$$

where $\boldsymbol{\theta} = \{\mathbf{F}, \sigma^2, \gamma\}$ is the set of unknown parameters. Similar to our non-parametric formulation, a maximum likelihood estimation of $\boldsymbol{\theta}$ can be derived based on the EM algorithm. We omit the details of the derivation, and only present the estimation of $\mathbf{F}$.

The fundamental matrix $\mathbf{F}$ can be estimated by minimizing a weighted error function

$$\mathcal{Q}(\mathbf{F}) = \sum_{n=1}^{N} p_n(\mathbf{v}_n^{\mathrm{T}}\mathbf{F}\mathbf{u}_n)^2. \tag{22}$$

We aim to seek a non-zero solution $\mathbf{F}$. To this end, a condition on the norm such as $\|\mathbf{F}\|_F = 1$ is used. Denote by $\mathbf{f}$ the nine-element vector made up of the entries of $\mathbf{F}$ in row-major order, $\mathbf{P} = \mathrm{diag}(p_1, \ldots, p_n)$, and $\mathbf{A}$ has the following form

$$\begin{bmatrix} \mathbf{v}_1^x\mathbf{u}_1^x & \mathbf{v}_1^x\mathbf{u}_1^y & \mathbf{v}_1^x & \mathbf{v}_1^y\mathbf{u}_1^x & \mathbf{v}_1^y\mathbf{u}_1^y & \mathbf{v}_1^y & \mathbf{u}_1^x & \mathbf{u}_1^y & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{v}_N^x\mathbf{u}_N^x & \mathbf{v}_N^x\mathbf{u}_N^y & \mathbf{v}_N^x & \mathbf{v}_N^y\mathbf{u}_N^x & \mathbf{v}_N^y\mathbf{u}_N^y & \mathbf{v}_N^y & \mathbf{u}_N^x & \mathbf{u}_N^y & 1 \end{bmatrix}.$$

Then the problem may be stated as finding the $\mathbf{f}$ that minimizes $\|\mathbf{P}^{1/2}\mathbf{A}\mathbf{f}\|$ subject to $\|\mathbf{f}\| = 1$. The solution is the unit singular vector corresponding to the smallest singular value of $\mathbf{P}^{1/2}\mathbf{A}$. Specifically, if $\mathbf{P}^{1/2}\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}$ with $\mathbf{D}$ diagonal with positive diagonal entries, arranged in descending order down the diagonal, then $\mathbf{f}$ is the last column of $\mathbf{V}$. Moreover, an important property of $\mathbf{F}$ is that it is singular, in fact of rank 2. To enforce this constraint, we replace $\mathbf{F}$ in each EM iteration by the closest singular matrix to it under a Frobenius norm [22].

For the case of homography, we have a parametric model: $\mathbf{v}_n - \mathbf{H}\mathbf{u}_n = 0$. The derivation of $\mathbf{H}$ is similar to the derivation of $\mathbf{F}$, and we omit the details for clarity.

E. Implementation Details

In the VFC algorithm, if the linear system (12) is solved directly, the matrix inversion operation then causes some problem when the matrix P is singular. For the numerical stability, we cope with this problem by defining a lower bound ε. Diagonal elements of P that are below ε is set to ε. In this paper, we set ε as 10^{-5}. Similarly, we constrain $\gamma \in [0.05, 0.95]$.

In the SparseVFC algorithm, there is a problem to which we need pay attention. We must ensure that the point set $\{\widetilde{\mathbf{x}}_m : m \in \mathbb{N}_M\}$ used to construct the basis functions does not contain two identical points since in this case the coefficient matrix in linear system (19), i.e. $(\widetilde{\mathbf{U}}^{\mathrm{T}}\widetilde{\mathbf{P}}\widetilde{\mathbf{U}} + \lambda\sigma^2\Gamma_s)$, will be singular. Obviously, this may appear in the point correspondence problem, since in the putative correspondence set there may exist one point in the first point set matched to several points in the second point set.

Parameter settings. There are four parameters in the VFC algorithm: β, λ, τ and γ. Parameters β and λ both reflect the amount of the smoothness constraint. Parameter β determines how wide the range of interaction between samples. Parameter λ controls the trade-off between the closeness to the data and the smoothness of the solution. Parameter τ is a threshold, which is used for deciding the correctness of a correspondence. Parameter γ reflects our initial assumption on the amount of inliers in the correspondence sets. In general, we find our method to be very robust to parameter changes. We set $\beta = 0.1$, $\lambda = 3$, $\tau = 0.75$ and $\gamma = 0.9$ throughout this paper.

V. Experimental Results

To evaluate our algorithm, we first design a set of experiments on vector field interpolation to demonstrate the efficiency of our technique in dealing with severe outliers, and then focus on the correspondence problem for building reliable point correspondences for 2D and 3D images. The experiments are performed on a laptop with 2.0 GHz Intel Pentium CPU, 8 GB memory and Matlab Code.

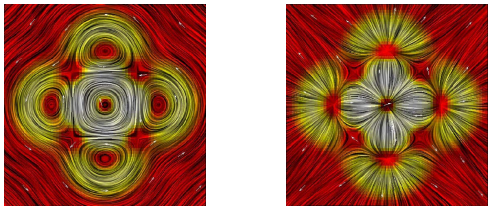

Fig. 2. Visualization of the synthetic 2D vector field. Left: its divergence-free part, corresponding to $\alpha = 0$; right: its curl-free part, corresponding to $\alpha = 1$.

A. Learning With Outliers

We focus on interpolating a synthetic 2D vector field from sparse samples as in [3]. The field is constructed from a function defined by a mixture of five Gaussians, which have the same covariance $0.25\mathbf{I}$ and centered at (0, 0), (1, 0), (0, 1), $(-1, 0)$ and $(0, -1)$ respectively. Its gradient and perpendicular gradient are shown in Fig. 2, which indicate a divergence-free and a curl-free field respectively. The synthetic data is then constructed by taking a convex combination of these two vector fields, controlled by a parameter α which is set to 0 and 0.5. The field is computed on a 70×70 grid over the square $[-2, 2] \times [-2, 2]$. The sparse inlier sets are obtained by uniformly sampled points from the grid. The outliers are generated as follow: the input $\mathbf{x}$ is chosen randomly from the grid; the output $\mathbf{y}$ is generated randomly from a uniform distribution on the square $[-2, 2] \times [-2, 2]$.

The kernel is chosen to be a convex combination of the divergence-free and curl-free kernels, i.e. Eqs. (15) and (16), with width $\widetilde{\sigma} = 0.8$, controlled by a parameter $\widetilde{\alpha}$ which is selected via cross-validation. After interpolating the vector field, we use it to predict the outputs on the whole grid and compare them to the ground truth. The experimental results are evaluated by means of interpolation errors, and the interpolation error is measured by an angular measure of error between the interpolated vector and the ground truth [3]. If $\mathbf{v}_g = (v_g^1, v_g^2)$ and $\mathbf{v}_e = (v_e^1, v_e^2)$ are the ground truth and estimated fields, we consider the transformation $\mathbf{v} \rightarrow \tilde{\mathbf{v}} = \frac{1}{\|(v^1, v^2, 1)\|}(v^1, v^2, 1)$. The interpolation error is defined as $err = \arccos(\tilde{\mathbf{v}}_e, \tilde{\mathbf{v}}_g)$.

The Tikhonov regularization on sample sets without outliers is used for comparison. For each set with a fixed number of inliers, we add outliers for VFC so that the inlier percentage varies from 0.9 to 0.1. Generally speaking, the performance of the Tikhonov regularization on a sample set without outliers can be considered as an upper bound performance of VFC on the sample set with outliers.

The results are reported in Fig. 3, in which we consider both the noiseless and noise cases. For the noise case, we add a Gaussian noise with zero mean and uniform standard deviation 0.1 to the inliers. As shown, the performance consistently improves with the increase of the cardinality of the sample set. When the sample set is small, the performance of the Tikhonov regularization without outliers is better than VFC, and the performance of VFC becomes worse as the outlier percentage increases. However, the difference in performance between them becomes small when the sample set is large.

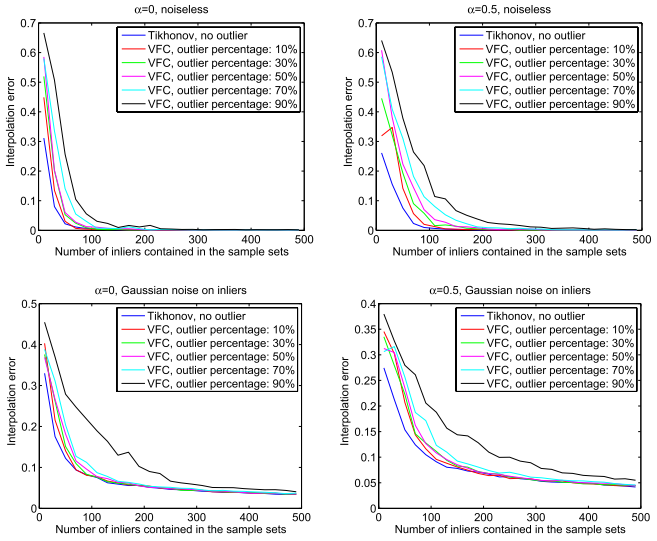

Fig. 3. Detailed results on synthetic 2D vector fields. Top: combination coefficient $\alpha = 0$ and 0.5 respectively, noiseless inliers; bottom: combination coefficient $\alpha = 0$ and 0.5 respectively, Gaussian noise with zero mean and uniform standard deviation 0.1 on inliers.

In conclusion, the performance of VFC is influenced both by the outlier percentage and sample set size, and it can reach the upper bound performance when the size of the given samples grows to an appropriate number, regardless of the percentage of the outliers.

B. Feature Correspondence on 2D Image Datasets

In this section, we focus on establishing feature correspondences for 2D real images. The open source VLFEAT toolbox [56] is used to determine the putative correspondences of SIFT [31]. All parameters are set as the default values except for the distance ratio threshold t. In the VLFEAT toolbox, it is defined as the ratio of the Euclidean distance of the second-closest neighbor and the closest neighbor. Usually, the greater is the value of t, the smaller amount of correspondences with higher inlier percentage will be. The default value of t is 1.5 and the smallest possible value is 1.0, which corresponds to the nearest neighbor strategy.

The experimental results are evaluated by precision and recall, where the precision is defined as the ratio of the preserved inlier number and the preserved correspondence number, and the recall is defined as the ratio of the preserved inlier number and the inlier number contained in the putative correspondences. We compare our VFC algorithm with other four methods which remove outliers from given putative point correspondences, such as ICF [26], GS [29], [30], RANSAC [18] and MLESAC [49]. We implement ICF and tune all parameters accordingly to find optimal settings. For GS and MLESAC, we implement them based on the publicly available codes. Throughout all the experiments, five algorithms' parameters are all fixed.

1) Results on Image Pairs of Homography: We test our method on the dataset of Mikolajczyk *et al* [38], which contains image pairs either of planar scenes or taken by camera in a fixed position during acquisition. The images, therefore,

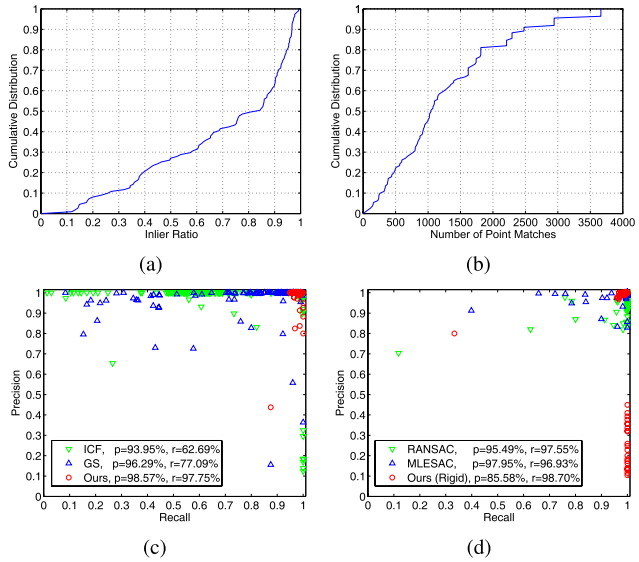

Fig. 4. Experimental results on the dataset of Mikolajczyk *et al* [38]. (**a**) Cumulative distribution function of initial inlier ratio. (**b**) Cumulative distribution function of number of point matches in the image pairs. (**c**) Precision-recall statistics for ICF, GS and VFC. (**d**) Precision-recall statistics for RANSAC, MLESAC and our method with parametric model. Our VFC (red circles, upper right corner) has the best precision and recall overall.

always obey homography. The ground truth homographies are supplied by the dataset. We use all the 40 image pairs, and for each pair, we set the SIFT distance ratio threshold t as 1.5, 1.3, 1.0 respectively. To determine the match correctness on this dataset, we similarly use an overlap error ϵ_S as in [38]: we reduce the scale of feature points to be $1/3$ of the original scale, and a correspondence is regarded as inlier if $\epsilon_S > 0$. The cumulative distribution function of original inlier percentage is shown in Fig. 4(a). The average precision of all image pairs is 69.57%, and about 30 percent of the correspondence sets have inlier percentage below 50%. Fig. 4(b) presents the cumulative distribution of the number of point matches contained in the experimental image pairs. We see that most of the image pairs have large scale of point matches (i.e. in the order of 1000's).

The results of five methods are summarized in Fig. 4(c) and (d), in which each scattered dot represents a precision-recall pair on an image pair. On the left is a comparison of our method to non-parametric model based methods ICF and GS, while on the right RANSAC and MLESAC based on geometric constraint (homography) are used for comparison. The average precision-recall pairs are (93.95%, 62.69%), (96.29%, 77.09%), (95.49%, 97.55%), (97.95%, 96.93%) and (98.57%, 97.75%) for ICF, GS, RANSAC, MLESAC and VFC respectively. As shown, ICF usually has high precision or recall, but not simultaneously. It lacks robustness when the outlier percentage is high or the viewpoint change is large. GS has high precision and low recall. This is probably because GS cannot estimate the factor for affinity matrix automatically and it is not affine-invariant. MLESAC performs a little better than RANSAC, and they both achieve quite satisfactory performance. This can be explained by the lack of complex constraints between the elements of the homography matrix. Our proposed method VFC has the best precision-recall trade-off. We also observe that the outlier removal capability of VFC is not affected by the large view angle, image rotation and affine transformation since these cases are all contained in the dataset. In fact, VFC performs well except when the initial number (not the percentage) of inliers is very small.

TABLE I
AVERAGE PRECISION-RECALL AND RUN-TIME COMPARISON OF RANSAC, VFC AND SPARSEVFC ON THE DATASET OF MIKOLAJCZYK

	RANSAC [18]	VFC	SparseVFC
(p, r)	(95.49, 97.55)	(98.57, 97.75)	(98.57, 97.78)
t (ms)	3784	6085	21

Since the scenes in the test image pairs are all rigid, we test the performance of our parametric variant (i.e., homography) as presented in Section IV-D. The results are shown in Fig. 4(d), marked by red circles, where we get an average precision-recall pair (85.58%, 98.70%). We see that our algorithm works quite well on most pairs, and fails on a small part of them (about 20%). In fact, we find that the inlier percentages in the failure image pairs are all below 50%. For the image pairs with inlier percentages over 50%, the average precision-recall pair is about (99.84%, 99.21%), compared to (98.15%, 99.87%), (99.38%, 99.79%) and (99.71%, 97.60%) in RANSAC, MLESAC and VFC respectively. That is to say, for handling a rigid scene, our algorithm with parametric model is somewhat sensitive to noise, however, if the outlier percentage is not so high, e.g., less than 50%, it still works quite well and yields comparable results.

The SparseVFC algorithm is also tested on this dataset, where the number M of basis functions used for sparse approximation is fixed to 15. The precision-recall pairs are summarized in Table I. We see that SparseVFC approximates VFC quite well, especially SparseVFC. The average run-times of VFC and SparseVFC are also presented in Table I. For comparison, we provide the run-time of RANSAC as well. The average run-times of VFC and RANSAC have the same order of magnitude. However, SparseVFC achieves a significant speedup with respect to VFC and RANSAC, more specifically, of two orders of magnitude, without any performance degradation.

2) Results on Image Pairs of Non-Rigid Object: The traditional methods such as RANSAC and similar techniques depend on a parametric geometrical model, for example, the fundamental matrix. If there exist some deformable objects with different shapes in the image pairs (this often happens in the area of *image retrieval* or *image-based non-rigid registration*), then these parametric model-based methods can no longer work, since the parametric model between the image pairs is not known apriori. However, our proposed VFC is a general method and it does not depend on any particular parametric model. Instead, it just uses a smoothness constraint so long as the deformation does not destroy the smoothness of the field.

To validate this idea, we consider two image pairs with deformable objects as shown in Fig. 5. In the first image

TABLE II

PERFORMANCE COMPARISON ON THE IMAGE PAIRS OF *DogCat* AND *T-shirt*. THE PAIRS IN THE TABLE ARE PRECISION-RECALL PAIRS (%) (THE SAME REPRESENTATION IS USED IN THE FOLLOWING EXPERIMENTS)

	ICF	GS	RANSAC	VFC
DogCat	(92.19, 63.44)	(97.70, 91.40)	(97.89, 100.0)	(100.0, 100.0)
T-shirt	(77.94, 72.60)	(94.83, 75.34)	(80.00, 87.67)	(90.67, 93.15)

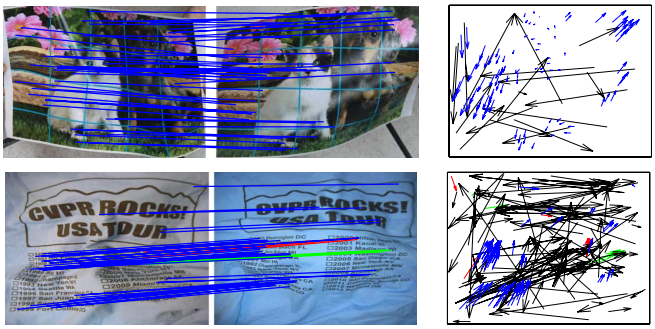

Fig. 5. Experimental results on two image pairs of non-rigid objects: *DogCat* and *T-shirt*. Top: results on *DogCat*, the initial inlier percentage is 82.30%, and the precision-recall pair is (100.0%, 100.0%); bottom: results on *T-shirt*, the initial inlier percentage is 36.50%, and the precision-recall pair is (90.67%, 93.15%). The lines and arrows indicate inlier detection results (blue = true positive, black = true negative, green = false negative, red = false positive). For visibility, in the image pairs, only 50 randomly selected correspondences are presented, and the true negatives are not shown. Best viewed in color (the same representation is used in the following experiments).

pair, we first add a regular grid on it, and then warp it and take two views with different deformations. The second image pair consists of scenes of two different deformations with illumination changes of a T-shirt. The match correctness is determined by manually refining the results of our VFC algorithm. The results are shown in Fig. 5 as well. On the *DogCat* pair, our VFC method correctly removes all the outliers and keeps all the inliers. On the *T-shirt* pair, there are still a few false positives and false negatives in the result since we could not precisely estimate the true warp function between the image pair under this framework. The average run-time of VFC on these two image pairs is about 55 milliseconds.

Recent work [51] justifies a simple RANSAC-driven deformable registration technique with an affine model that is at least as accurate as other methods based on the optimization of fully deformable models. Therefore, besides ICF and GS, we compare VFC to RANSAC as well on these two image pairs. The result is shown in Table II. We see that RANSAC performs well in case of small or moderate distortion, e.g., in the *DogCat* pair. However, when the deformation is relatively large, e.g., in the *T-shirt* pair, it cannot obtain satisfactory results, since just an affine model is not capable to handle a large complex deformation. Our method VFC has the best precision-recall scores. In general, VFC is effective for establishing feature correspondence on image pairs related by smooth motion fields, including both rigid and non-rigid cases.

3) Results on Wide Baseline Images: The motion fields introduced by the image pairs in the previous sections are

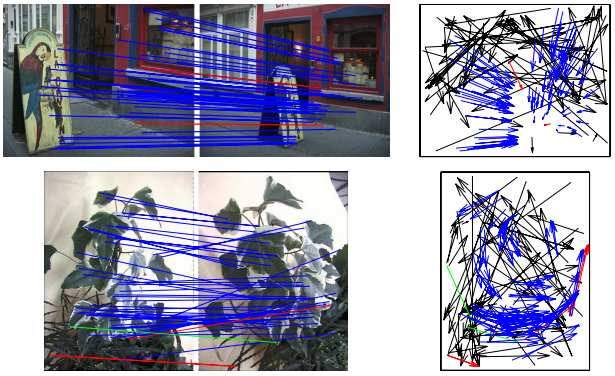

Fig. 6. Experimental results on two wide baseline image pairs: *Mex* and *Tree*. Top: results on *Mex*, the initial inlier percentage is 51.90%, and the precision-recall pair is (96.47%, 100.0%); bottom: results on *Tree*, the initial inlier percentage is 56.29%, and the precision-recall pair is (94.85%, 97.87%).

TABLE III

PERFORMANCE COMPARISON ON THE IMAGE PAIRS OF *Mex* AND *Tree*

	ICF	GS	RANSAC	MLESAC	VFC
Mex	(96.15, 60.98)	(93.83, 92.68)	(91.76, 95.12)	(92.05, 98.78)	(96.47, 100.0)
Tree	(92.75, 68.09)	(97.62, 87.23)	(94.68, 94.68)	(98.82, 89.36)	(94.85, 97.87)

usually smooth, and we obtain good performance. Now we test the VFC method on wide baseline image pairs, in which the motion fields are in general not continuous. The test images are from the dataset of Tuytelaars *et al.* [52], and the match correctness is determined by manually refining the results of RANSAC.

We first consider two wide baseline image pairs, *Mex* and *Tree*, as shown in Fig. 6. The *Mex* pair is a structured scene and the *Tree* pair is an unstructured scene. On the *Mex* pair, as shown on the top row of Fig. 6, there are 158 putative correspondences with 76 outliers; the inlier percentage is 51.90%; after using the VFC to remove the outliers, 85 correspondences are preserved, including all the 82 inliers. The precision-recall pair is about (96.47%, 100.0%). A similar result on the *Tree* pair is presented on the bottom row of Fig. 6. The average run-time of VFC on these two image pairs is about 17 milliseconds.

The performance of VFC compared to other four approaches is shown in Table III. The geometry model used in RANSAC and MLESAC is epipolar geometry. We see that MLESAC is slightly better than GS and RANSAC. The recall of ICF is quite low, although it has a satisfactory precision score. However, VFC can successfully distinguish inliers from outliers, and it has the best trade-off between precision and recall. These results suggest that even though the motion field is discontinuous and not consistent with the smoothness constraint, the VFC method is still effective for establishing feature correspondences.

Since the smoothness constraint is imposed as a prior in our approach, it may be problematic in some particular cases. We now consider an image pair shown in Fig. 7. In this image pair, there exists a large depth discontinuity; the point

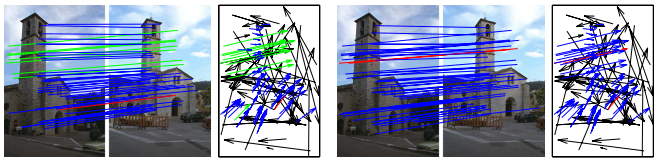

Fig. 7. Results on the image pair of *Valbonne*. Left: result of VFC, the correspondences on the sky are falsely removed; right: first use point correspondences preserved by VFC to estimate fundamental matrix, and then use it for outlier removal; in this time all the inliers are preserved.

TABLE IV
THE INLIER PERCENTAGE IS CHANGED BY ADDING OUTLIERS

inlier pct.	54.76%	37.65%	9.56%	3.69%	2.28%
ICF	(91.67, 63.77)	(100.0, 21.51)	(13.29, 100.0)	(3.69, 100.0)	(2.28, 100.0)
GS	(91.78, 97.10)	(92.31, 90.32)	(84.44, 63.33)	(86.05, 30.83)	(84.00, 17.50)
RANSAC	(94.52, 100.0)	(83.50, 92.47)	(54.31, 89.17)	-	-
MLESAC	(94.44, 98.55)	(95.00, 81.72)	(83.64, 76.67)	-	-
VFC	(98.33, 85.51)	(94.25, 88.17)	(90.76, 90.00)	(86.96, 83.33)	(85.47, 83.33)

correspondences on the sky violate the smoothness prior, which will be removed by our VFC method, as shown on the left of Fig. 7. It should be noted that this does not influence the recovery of epipolar geometry, since the point correspondences preserved by VFC in general do not lie in a dominant plane, and thereby are not geometrically degenerate with respect to the fundamental matrix [15].

To validate this idea, we again take the *Valbonne* pair for example, and apply our parametric variant to estimate the epipolar geometry, e.g., the fundamental matrix, based on the correspondences preserved by our non-parametric model VFC. After we estimate the fundamental matrix, we use it to determine match correctness of the whole set of putative correspondences. The result is shown on the right of Fig. 7. We see that all the inliers are preserved, including the point correspondences on the sky. This suggests that the epipolar geometry has been correctly estimated.

4) Robustness Test: We next test the robustness of VFC and compare it to ICF, GS, RANSAC and MLESAC on the image pair shown in Fig. 7. In our evaluation we consider the following two scenarios.

On the one hand, reducing the distance ratio threshold would generate more inliers. For instance, changing t from 1.5 to 1, the number of inliers will increase from 69 to 120 in the *Valbonne* pair. This is sometimes important for the possible subsequent analysis such as fundamental matrix estimation. Here we design a series of experiments from this perspective. For each image pair, we generate five correspondence sets by the following procedure: the distance ratio thresholds are first set to 1.5, 1.3 and 1.0 respectively; then we fix threshold to 1.0 and add $2,000$ and $4,000$ random outliers respectively. The result is presented in Table IV. We observe that the performance of VFC is satisfactory, and it can tolerate even 90% outliers. As the inlier percentage decreases, the precision and recall of VFC decrease gradually. Still, the results are acceptable compared to other four alternative methods.

On the other hand, images taken at close range may result in plenty of outliers while having only a few inliers.

TABLE V
THE INLIER PERCENTAGE IS CHANGED BY REDUCING INLIERS

inlier pct.	46.73%	41.24%	34.48%	25.97%	14.93%
ICF	(76.36, 84.00)	(59.68, 92.50)	(44.44, 93.33)	(32.76, 95.00)	(18.00, 90.00)
GS	(90.91, 100.0)	(85.11, 100.0)	(71.43, 100.0)	(48.78, 100.0)	(23.26, 100.0)
RANSAC	(88.00, 88.00)	(86.96, 100.0)	(81.08, 100.0)	(73.08, 95.00)	(27.27, 60.00)
MLESAC	(90.00, 90.00)	(86.49, 80.00)	(79.41, 90.00)	(55.17, 80.00)	(32.00, 80.00)
VFC	(97.73, 86.00)	(96.88, 77.50)	(92.59, 83.33)	(90.48, 95.00)	(75.00, 90.00)

One standard example is in the endoscopic images, since they often involve low texture, abundant specularity, blurs and extreme illumination changes [58]. Here we test the robustness of the VFC by reducing the percentage of inliers. For each image pair, we generate five correspondence sets by the following procedure: we first fix the distance ratio threshold to be 1.5 and then randomly remove inliers so that the numbers of inliers become 50, 40, 30, 20 and 10 respectively. The initial number of correspondence and inlier are 126 and 69 respectively. The result is presented in Table V. We see that VFC becomes ineffective when both the inlier number and the inlier percentage in the sample set are very small. However, in other cases, the performance of VFC is still satisfactory compared with other four competing methods.

C. Feature Correspondence on 3D Surfaces

In this section, we establish feature correspondences for 3D surfaces. We adopt the datasets used in [66], which contain two types of 3D data: rigid and non-rigid objects. In the rigid case, the test datasets are *Dino* and *Temple* datasets; each surface pair is from the same rigid object which can be aligned using a rotation, translation and scale. In the non-rigid case, the dataset is the INRIA *Dance-1* sequence, each surface pair is from the same moving person.

We determine the putative correspondences by using the method of Zaharescu *et al*. [66] which detects correspondences between nontrivial feature points on the 2D manifolds, such as the photometric and local curvature data. The feature point detector is called *MeshDOG*, and the feature descriptor is called *MeshHOG*.

The match correctness is determined as follows. For the rigid objects such as the *Dino* and *Temple* datasets, the correspondence between the two surfaces can be formulated as $\mathbf{y} = s\mathbf{R}\mathbf{x}+\mathbf{t}$, where $\mathbf{R}_{3\times3}$ is a rotation matrix, s is a scaling parameter, and $\mathbf{t}_{3\times1}$ is a translation vector. We can use some robust rigid point registration methods such as the Coherent Point Drift (CPD) [39] to solve for these three parameters, and then the match correctness can be accordingly determined. On the INRIA *Dance-1* sequence, which contains non-rigid objects, the match correctness is difficult to determine; we just visualize the results in image pairs.

1) Results on Rigid Objects: We test the VFC method on two surface pairs of rigid objects, the *Dino* and *Temple* datasets, which satisfy similarity transformations. For comparisons, we choose RANSAC combined with similarity transformation. The correspondence between two surfaces can be formulated as $\mathbf{y} = s\mathbf{R}\mathbf{x} + \mathbf{t}$. This model has seven degrees of freedom: three for rotation matrix $\mathbf{R}$, three for translation

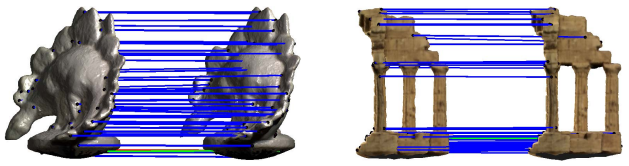

Fig. 8. Experimental results on rigid object datasets of *Dino* and *Temple*. Left: results on *Dino*, the initial inlier percentage is about 81.23%, and the precision-recall pair is (98.87%, 99.62%); right: results on *Temple*, the initial inlier percentage is about 89.96%, and the precision-recall pair is (99.07%, 99.53%).

TABLE VI

PERFORMANCE COMPARISON ON THE *Dino* DATASET

inlier pct.	32.00%	19.92%	11.35%	6.10%	3.17%
RANSAC	(99.62, 99.62)	(99.62, 99.62)	(99.63, 100.0)	(100.0, 100.0)	(98.64, 99.32)
VFC	(98.86, 98.12)	(99.23, 97.37)	(98.48, 95.57)	(99.25, 93.97)	(98.59, 94.58)

vector $\mathbf{t}$ and one for scaling factor s. Therefore, three point correspondences are sufficient to recover the similarity transformation. The only restriction is that the three points must be in "general position", which means that they should not be collinear. To obtain the closed form solution for these three parameters, we use the method of Umeyama [55].

The results of the VFC are shown in Fig. 8. For the *Dino* dataset, there are 325 putative correspondences with 61 outliers; after using the VFC to remove outliers, 266 correspondences are preserved, in which 263 are inliers. That is to say, 58 of 61 outliers are eliminated while discarding only 1 inlier. A similar result on the *Temple* dataset is presented on the right of Fig. 8. The average run-time of VFC on these two surface pairs is about 48 milliseconds. Experiments on these datasets with rotation and scale transformations are also performed. We observe similar performances.

However, using RANSAC obtains even better results: the precision-recall pairs of RANSAC on these two datasets are (100.0%, 100.0%) and (99.07%, 100.0%). Actually, in this case, RANSAC is only influenced by noise on inliers, but not the random outliers. On the one hand, despite how large the proportion of outliers is, the sampling rule ensures with high probability, p (usually is chosen at 0.99), at least one of the random samples of points is free from outliers; that is to say that we can work out the correct model generally. On the other hand, for a point on the first object, there is one, and only one, point on the other object corresponds to it by the similarity transformation; if we obtain the parameters of the similarity transformation, all outliers could be removed. This is different from the case of RANSAC with fundamental matrix on the 2D image pairs. For a point in one image, there is one line in the other image corresponding to it and all points on this line satisfy the epipolar line constraint.

We then test the robustness of the VFC on a surface pair, the *Dino* dataset, and compare it with the RANSAC algorithm. We generate five correspondence sets by adding different numbers of additional outliers: 500, 1, 000, 2, 000, 4, 000 and 8, 000 respectively. Here each outlier is generated by randomly choosing one vertex from each of the surfaces. The results are shown in Table VI. As we expect, RANSAC is not influenced by outliers. Similar to the 2D case, the performance of VFC is also quite satisfactory, and it can tolerate 90% outliers. As the inlier percentage decreases, the recall of VFC decreases gradually while having slight changes on the precision.

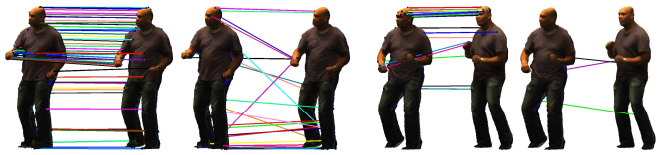

Fig. 9. Experimental results on non-rigid object real datasets of the INRIA *Dance-1* sequence. Left two: results on frames 525 and 527; right two: results on frames 530 and 550. For each group, the left pair denotes the identified suspect inliers, and the right pair denotes the removed suspect outliers. For visibility, at most 50 randomly selected correspondences are presented.

From these results we observe that the VFC is not influenced by the dimension of the input data; the performance is as good as that in the 2D case, although the RANSAC is more effective in this rigid case.

2) Results on Non-Rigid Objects: We further conduct experiments on non-rigid object case, and we used the INRIA *Dance-1* sequence.

We first test two nearby frames. In this case, the object often observes small deformations. The results are presented in the left two figures of Fig. 9. There are 191 putative correspondences. After the VFC is used to remove the outliers, 164 correspondences are preserved.

We then consider two frames that are far apart. In this case, the object usually has a large deformation, leading to less putative correspondences. The results are shown in the right two figures of Fig. 9. There are 23 putative correspondences, and 20 of them are preserved after using the VFC for outlier removal. Note that the preserved correspondences contain two on the fist which seem not fit the spatially smooth field introduced by the other identified suspect inliers. This could be due to the sparsity of the sample set, which increases the uncertainty of the vector field. Unlike the outliers in the bottom right figure of Fig. 9, the two correspondences on the fist just slightly violate the spatial smoothness. Thus, it is possible to find a smooth field which agrees with those preserved correspondences.

In conclusion, for the feature correspondence problem, VFC demonstrates its capability of handling 3D data which contains both rigid and non-rigid objects.

D. Non-Rigid Point Set Registration

Point set registration aims to align two point sets $\{\mathbf{x}_n\}_{n=1}^N$ (the model point set) and $\{\mathbf{y}_l\}_{l=1}^L$ (the target point set). Typically, in the non-rigid case, it requires estimating a non-rigid transformation $\mathbf{f}$ which warps the model point set to the target point set. Recall that our VFC method is able to generate a smoothly interpolated vector field with adherence to a set of observed input-output pairs. Therefore, it could be used to recover the transformation between two point sets with a set of putative correspondences.

We determine the putative correspondences by using the shape context descriptor [4], using the Hungarian method for matching with the χ^2 test statistic as the cost measure. The two steps of estimating correspondences and transformations

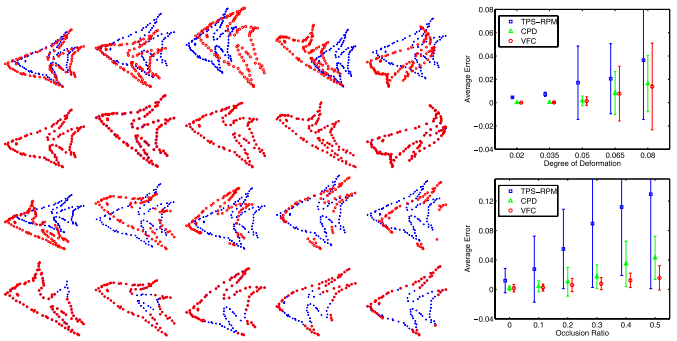

Fig. 10. Point set registration results of our VFC method on a *fish* shape with deformation (top) and occlusion (bottom) presented in every two rows. The goal is to align the model point set (blue pluses) onto the target point set (red circles). For each group of experiments, the upper figure is the model and target point sets, and the lower figure is the registration result. From left to right, increasing degree of degradation. The rightmost figures are comparisons of the registration performance of our method with TPS-RPM [12] and CPD [39] on the corresponding distortions. The error bars indicate the registration error means and standard deviations over 100 trials.

are iterated to obtain a reliable result. We use a fixed number of iterations, typically 10 but more refined schemes are possible.

We tested our method on a synthesized *fish* shape as in [12]. We test the robustness of our VFC on different degrees of deformations and occlusions, and for each deformation level 100 samples are generated. Fig. 10 shows the registration results. We see that for both deformation and occlusion with moderate degradation, our method is able to produce an almost perfect alignment. The matching performance degrades gradually and gracefully as the degree of distortion in the data increases. To provide a quantitative comparison, we report the results of other two state-of-the-art algorithms such as TPS-RPM [12] and CPD [39] which are implemented using publicly available codes. The registration error on a pair of shape is quantified as the average Euclidean distance between a point in the warped model and the corresponding point in the target. Then the registration performance of each algorithm is compared by the mean and standard deviation of the registration error of all the 100 samples in each distortion level. The statistical results, error means, and standard deviations for each setting are summarized in the last column of Fig. 10. As shown, our VFC method achieves similar matching performance compared to CPD on the deformation test, and both algorithms perform better than TPS-RPM. However, in the occlusion test, VFC consistently outperforms the other two algorithms in all degrees of rotations.

VI. Conclusion

In this paper, we proposed and studied a new vector field interpolation algorithm called *vector field consensus* (VFC) that is robust and fast. It simultaneously generates a smoothly interpolated vector field and estimates the consensus set by an iterative EM algorithm. We apply it to point correspondence problems in computer vision, in which the feature correspondences between image pairs are determined based on the coherence of the underlying motion fields rather than the geometric constraints. Experiments on 2D and 3D real image datasets demonstrate the capability of VFC being able to tolerate 90% outliers. Quantitative results demonstrate that VFC outperforms state-of-the-art methods such as RANSAC. In addition, we describe a variant of VFC which uses a parametric model (e.g., exploiting rigidity) which we show in more effective than RANSAC, but less effective that VFC if there are many outliers. We also provide an efficient implementation of VFC called SparseVFC, which significantly reduces the computational complexity without much performance degradation.

Appendix A
Vector-Valued RKHS

We review the basic theory of vector-valued reproducing kernel Hilbert space, and for further details and references we refer to [37] and [10].

Let $\mathcal{Y}$ be a real Hilbert space with inner product (norm) $\langle\cdot,\cdot\rangle$, $(\|\cdot\|)$, for example, $\mathcal{Y}\subseteq\mathbb{R}^D$, $\mathcal{X}$ a set, for example, $\mathcal{X}\subseteq\mathbb{R}^P$, and $\mathcal{H}$ a Hilbert space with inner product (norm) $\langle\cdot,\cdot\rangle_{\mathcal{H}}$, $(\|\cdot\|_{\mathcal{H}})$. Note that a norm can be induced by an inner product, for example, $\forall\mathbf{f}\in\mathcal{H}$, $\|\mathbf{f}\|_{\mathcal{H}}=\sqrt{\langle\mathbf{f},\mathbf{f}\rangle_{\mathcal{H}}}$.

Definition 1: A Hilbert space $\mathcal{H}$ is an RKHS if the evaluation maps $ev_{\mathbf{x}}:\mathcal{H}\to\mathcal{Y}$ are bounded, i.e. if $\forall\mathbf{x}\in\mathcal{X}$ there exists a positive constant $C_{\mathbf{x}}$ such that

$$\|ev_{\mathbf{x}}(\mathbf{f})\|=\|\mathbf{f}(\mathbf{x})\|\leq C_{\mathbf{x}}\|\mathbf{f}\|_{\mathcal{H}},\quad\forall\mathbf{f}\in\mathcal{H}. \tag{23}$$

A reproducing kernel $\Gamma:\mathcal{X}\times\mathcal{X}\to\mathcal{B}(\mathcal{Y})$ is then defined as: $\Gamma(\mathbf{x},\mathbf{x}'):=ev_{\mathbf{x}}ev_{\mathbf{x}'}^*$, where $\mathcal{B}(\mathcal{Y})$ is the space of bounded operators on $\mathcal{Y}$, for example, $\mathcal{B}(\mathcal{Y})\subseteq\mathbb{R}^{D\times D}$, and $ev_{\mathbf{x}}^*$ is the adjoint of $ev_{\mathbf{x}}$.

Remark 1: The kernel Γ reproduces the value of a function $\mathbf{f}\in\mathcal{H}$ at a point $\mathbf{x}\in\mathcal{X}$. Indeed, $\forall\mathbf{x}\in\mathcal{X}$ and $\mathbf{y}\in\mathcal{Y}$, we have $ev_{\mathbf{x}}^*\mathbf{y}=\Gamma(\cdot,\mathbf{x})\mathbf{y}$, so that $\langle\mathbf{f}(\mathbf{x}),\mathbf{y}\rangle=\langle\mathbf{f},\Gamma(\cdot,\mathbf{x})\mathbf{y}\rangle_{\mathcal{H}}$.

Remark 2: An RKHS defines a corresponding reproducing kernel. Conversely, a reproducing kernel defines a unique RKHS.

More specifically, for any $N\in\mathbb{N}$, $\{\mathbf{x}_n:n\in\mathbb{N}_N\}\subseteq\mathcal{X}$, and a reproducing kernel Γ, a unique RKHS can be defined by considering the completion of the space

$$\mathcal{H}_N=\left\{\sum_{n=1}^{N}\Gamma(\cdot,\mathbf{x}_n)\mathbf{c}_n:\mathbf{c}_n\in\mathcal{Y}\right\}, \tag{24}$$

with respect to the norm induced by the inner product

$$\langle\mathbf{f},\mathbf{g}\rangle_{\mathcal{H}}=\sum_{i,j=1}^{N}\langle\Gamma(\mathbf{x}_j,\mathbf{x}_i)\mathbf{c}_i,\mathbf{d}_j\rangle\quad\forall\mathbf{f},\mathbf{g}\in\mathcal{H}_N, \tag{25}$$

where $\mathbf{f}=\sum_{i=1}^{N}\Gamma(\cdot,\mathbf{x}_i)\mathbf{c}_i$ and $\mathbf{g}=\sum_{j=1}^{N}\Gamma(\cdot,\mathbf{x}_j)\mathbf{d}_j$.

Appendix B
Proof of Theorem 1

For any given reproducing kernel Γ, we can define a unique RKHS $\mathcal{H}_N$ as in Eq. (24). Let $\mathcal{H}_N^{\perp}$ be a subspace of $\mathcal{H}$,

$$\mathcal{H}_N^{\perp}=\{\mathbf{f}\in\mathcal{H}:\mathbf{f}(\mathbf{x}_n)=0,n\in\mathbb{N}_N\}. \tag{26}$$

Form the reproducing property, i.e. Remark 1, $\forall\mathbf{f}\in\mathcal{H}_N^{\perp}$

$$\langle\mathbf{f},\sum_{n=1}^{N}\Gamma(\cdot,\mathbf{x}_n)\mathbf{c}_n\rangle_{\mathcal{H}}=\sum_{n=1}^{N}\langle\mathbf{f}(\mathbf{x}_i),\mathbf{c}_n\rangle=0. \tag{27}$$

Thus $\mathcal{H}_N^\perp$ is the orthogonal complement of $\mathcal{H}_N$; then every $\mathbf{f} \in \mathcal{H}$ can be uniquely decomposed in components along and perpendicular to $\mathcal{H}_N$: $\mathbf{f} = \mathbf{f}_N + \mathbf{f}_N^\perp$, where $\mathbf{f}_N \in \mathcal{H}_N$ and $\mathbf{f}_N^\perp \in \mathcal{H}_N^\perp$. Since by orthogonality $\|\mathbf{f}_N + \mathbf{f}_N^\perp\|_{\mathcal{H}}^2 = \|\mathbf{f}_N\|_{\mathcal{H}}^2 + \|\mathbf{f}_N^\perp\|_{\mathcal{H}}^2$ and by the reproducing property $\mathbf{f}(\mathbf{x}_n) = \mathbf{f}_N(\mathbf{x}_n)$, the regularized risk functional then satisfies

$$\begin{aligned}\mathcal{E}(\mathbf{f}) &= \frac{1}{2\sigma^2}\sum_{n=1}^{N} p_n \|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2 + \frac{\lambda}{2}\|\mathbf{f}_N + \mathbf{f}_N^\perp\|_{\mathcal{H}}^2 \\ &\geq \frac{1}{2\sigma^2}\sum_{n=1}^{N} p_n \|\mathbf{y}_n - \mathbf{f}_N(\mathbf{x}_n)\|^2 + \frac{\lambda}{2}\|\mathbf{f}_N\|_{\mathcal{H}}^2. \end{aligned} \tag{28}$$

Therefore, the optimal solution of the regularized risk functional (11) comes from the space $\mathcal{H}_N$, and hence has the form (2). To solve for the coefficients, we consider the definition of the smoothness functional $\phi(\mathbf{f})$ and the inner product (25), the regularized risk functional then can be conveniently expressed in the following matrix form:

$$\mathcal{E}(\mathbf{f}) = \frac{1}{2\sigma^2}\|\widetilde{\mathbf{P}}^{1/2}(\widetilde{\mathbf{Y}} - \widetilde{\Gamma}\widetilde{\mathbf{C}})\|^2 + \frac{\lambda}{2}\widetilde{\mathbf{C}}^{\mathrm{T}}\widetilde{\Gamma}\widetilde{\mathbf{C}}. \tag{29}$$

where $\widetilde{\Gamma}$ is an $N \times N$ block matrix with the (i, j)-th block $\Gamma(\mathbf{x}_i, \mathbf{x}_j)$, and $\widetilde{\mathbf{C}} = (\mathbf{c}_1^{\mathrm{T}}, \ldots, \mathbf{c}_N^{\mathrm{T}})^{\mathrm{T}}$ is the coefficient vector. Taking the derivative of the last Eq. with respect to $\widetilde{\mathbf{C}}$ and setting it to zero, we obtain the linear system in Eq. (12). Thus the coefficient set $\{\mathbf{c}_n : n \in \mathbb{N}_N\}$ of the optimal solution $\mathbf{f}$ is determined by the linear system (12).

REFERENCES

[1] M. A. Álvarez and N. D. Lawrence, "Computationally efficient convolved multiple output Gaussian processes," *J. Mach. Learn. Res.*, vol. 12, no. 1, pp. 1425–1466, 2011.

[2] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, no. 3, pp. 337–404, 1950.

[3] L. Baldassarre, L. Rosasco, A. Barla, and A. Verri, "Multi-output learning via spectral filtering," *Mach. Learn.*, vol. 87, no. 3, pp. 259–301, 2012.

[4] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 24, pp. 509–522, Apr. 2002.

[5] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.

[6] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.

[7] M. J. Black and A. Rangarajan, "On the unification of line processes, outlier rejection and robust statistics with applications in early vision," *Int. J. Comput. Vis.*, vol. 19, no. 1, pp. 57–91, 1996.

[8] P. Boyle and M. Frean, "Dependent Gaussian processes," in *Advances in Neural Information Processing Systems*, Cambridge, MA, USA: MIT Press, 2005, pp. 217–224.

[9] B. Cabral and L. C. Leedom, "Imaging vector fields using line integral convolution," in *Proc. 20th Annu. Conf. Comput. Graph. Interactive Tech.*, vol. 27. 1993, pp. 263–270.

[10] C. Carmeli, E. De Vito, and A. Toigo, "Vector valued reproducing kernel Hilbert spaces of integrable functions and mercer theorem," *Anal. Appl.*, vol. 4, no. 4, pp. 377–408, 2006.

[11] M. Cho and K. M. Lee, "Progressive graph matching: Making a move of graphs via probabilistic voting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 398–405.

[12] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Understand.*, vol. 89, nos. 2–3, pp. 114–141, 2003.

[13] O. Chum and J. Matas, "Matching with PROSAC—Progressive sample consensus," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 220–226.

[14] O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," in *Proc. Pattern Recognit. Symp. German Assoc. Pattern Recognit. (DAGM)*, 2003, pp. 236–243.

[15] O. Chum, T. Werner, and J. Matas, "Two-view geometry estimation unaffected by a dominant plane," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 772–779.

[16] T. Corpetti, E. Mémin, and P. Pérez, "Dense estimation of fluid flows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 365–380, Mar. 2002.

[17] A. Dempster, N. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc. Series B*, vol. 39, no. 1, pp. 1–38, 1977.

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

[19] D. Geiger and A. L. Yuille, "A common framework for image segmentation," *Int. J. Comput. Vis.*, vol. 6, no. 3, pp. 227–243, 1991.

[20] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[21] S. Gold, A. Rangarajan, C. P. Lu, S. Pappu, and E. Mjolsness, "New algorithms for 2-D and 3-D point matching: Pose estimation and correspondence," *Pattern Recognit.*, vol. 31, no. 8, pp. 1019–1031, 1998.

[22] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[23] P. J. Huber, *Robust Statistics*. New York, NY, USA: Wiley, 1981.

[24] A. E. Johnson and M. Hebert, "Using spin-images for efficient object recognition in cluttered 3-D scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433–449, May 1999.

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

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

[27] B. Lin, S. Yang, C. Zhang, J. Ye, and X. He, "Multi-task vector field learning," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2012, pp. 296–304.

[28] C. Liu, J. Yuen, and A. Torralba, "SIFT flow: Dense correspondence across different scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 978–994, May 2011.

[29] H. Liu and S. Yan, "Common visual pattern discovery via spatially coherent correspondence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1609–1616.

[30] H. Liu and S. Yan, "Robust graph mode seeking by graph shift," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 671–678.

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

[32] H. Lu and A. L. Yuille, "Ideal observers for detecting motion: Correspondence noise," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2006.

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

[34] J. Ma, J. Zhao, J. Tian, Z. Tu, and A. Yuille, "Robust estimation of nonrigid transformation for point set registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2147–2154.

[35] J. Ma, J. Zhao, Y. Zhou, and J. Tian, "Mismatch removal via coherent spatial mapping," in *Proc. Int. Conf. Image Process.*, 2012, pp. 1–4.

[36] I. Macêdo and R. Castro, "Learning divergence-free and curl-free vector fields with matrix-valued kernels," Instituto Nacional de Matematica Pura e Aplicada, Brasil, Tech. Rep. , 2008.

[37] C. A. Micchelli and M. Pontil, "On learning vector-valued functions," *Neural Comput.*, vol. 17, no. 1, pp. 177–204, 2005.

[38] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, *et al.*, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, no. 1, pp. 43–72, 2005.

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

[40] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, vol. 78, no. 9, pp. 1481–1497, Sep. 1990.

[41] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, no. 6035, pp. 314–319, 1985.

[42] R. Raguram, J. M. Frahm, and M. Pollefeys, "A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 500–513.
[43] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
[44] R. Rifkin, G. Yeo, and T. Poggio, "Regularized least-squares classification," in *Advances in Learning Theory: Methods, Model and Applications*. Cambridge, MA, USA: MIT Press, 2003.
[45] S. Roth and M. J. Black, "On the spatial statistics of optical flow," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 33–50, 2007.
[46] P. J. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection*. New York, NY, USA: Wiley, 1987.
[47] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 2nd ed. Pacific Grove, CA, USA: Brooks/Cole Company, 1999.
[48] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-posed Problems*. Washington, DC, USA: Winston, 1977.
[49] P. H. S. 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.
[50] L. Torresani, V. Kolmogorov, and C. Rother, "Feature correspondence via graph matching: Models and global optimization," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 596–609.
[51] Q.-H. Tran, T.-J. Chin, G. Carneiro, M. S. Brown, and D. Suter, "In defence of RANSAC for outlier rejection in deformable registration," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 274–287.
[52] T. Tuytelaars and L. van Gool, "Matching widely separated views based on affine invariant regions," *Int. J. Comput. Vis.*, vol. 59, no. 1, pp. 61–85, 2004.
[53] S. Ullman, *The Interpretation of Visual Motion*, vol. 28, Cambridge, MA, USA: MIT Press, 1979.
[54] S. Ullman and A. L. Yuille, *Rigidity and Smoothness of Motion*. Cambridge, MA, USA: MIT Press, 1987.
[55] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 4, pp. 376–380, Apr. 1991.
[56] A. Vedaldi and B. Fulkerson, "VLFeat—An open and portable library of computer vision algorithms," in *Proc. 18th Annu. ACM Int. Conf. Multimedia*, 2010, pp. 1469–1472.
[57] G. Wahba, *Spline Models for Observational Data*. Philadelphia, PA, USA: SIAM, 1990.
[58] H. Wang, D. Mirota, M. Ishii, and G. D. Hager, "Robust motion estimation and structure recovery from endoscopic image sequences with an adaptive scale kernel consensus estimator," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–7.
[59] Y. Weiss and E. H. Adelson, "Slow and smooth: A Bayesian theory for the combination of local motion signals in human vision," Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 1624, 1998.
[60] S. Wu, H. Lu, and A. L. Yuille, "Model selection and parameter estimation in motion perception," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2008.
[61] L. Xu and A. L. Yuille, "Robust principal component analysis by self-organizing rules based on statistical physics approach," *IEEE Trans. Neural Netw.*, vol. 6, no. 1, pp. 131–144, Jan. 1995.
[62] S. Yu, V. Tresp, and K. Yu, "Robust multi-task learning with t-processes," in *Proc. Int. Conf. Mach. Learn.*, 2007, pp. 1103–1110.
[63] A. L. Yuille, "Generalized deformable models, statistical physics, and matching problems," *Neural Comput.*, vol. 2, no. 1, pp. 1–24, 1990.
[64] A. L. Yuille and N. M. Grzywacz, "A computational theory for the perception of coherent visual motion," *Nature*, vol. 333, no. 6168, pp. 71–74, 1988.
[65] A. L. Yuille and N. M. Grzywacz, "A mathematical analysis of the motion coherence theory," *Int. J. Comput. Vis.*, vol. 3, no. 2, pp. 155–175, 1989.
[66] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud, "Surface feature detection and description with applications to mesh matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 373–380.
[67] J. Zhao, J. Ma, J. Tian, J. Ma, and D. Zhang, "A robust method for vector field learning with application to mismatch removing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 2977–2984.
[68] S. Zhu, K. Yu, and Y. Gong, "Predictive matrix-variate t models," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2008, pp. 1721–1728.

Jiayi Ma received the B.S. degree from the Department of Mathematics, Huazhong University of Science and Technology (HUST), Wuhan, China, in 2008. He is currently pursuing the Ph.D. degree with the School of Automation, HUST. From 2012 to 2013, he was with the Department of Statistics, University of California at Los Angeles. His current research interests include in the areas of computer vision and machine learning.

Ji Zhao received the B.S. degree in automation from the Nanjing University of Posts and Telecommunication and the Ph.D. degree in control science and engineering from Huazhong University of Science and Technology in 2005 and 2012, respectively. Since 2012, he has been a Post-Doctoral Research Associate with the Robotics Institute, Carnegie Mellon University. His current research interests include image classification, image segmentation, and kernel-based learning.

Jinwen Tian received the Ph.D. degree in pattern recognition and intelligent systems from Huazhong University of Science and Technology (HUST), China, in 1998. He is a Professor and Ph.D. Supervisor of pattern recognition and artificial intelligence with HUST. His current research interests include remote sensing image analysis, wavelet analysis, image compression, computer vision, and fractal geometry.

Alan L. Yuille received the B.A. degree in mathematics and the Ph.D. degree in theoretical physics studying under Stephen Hawking from the University of Cambridge, in 1976 and 1980, respectively. He held a post-doctoral position with the Physics Department, University of Texas at Austin, and the Institute for Theoretical Physics, Santa Barbara. He then joined the Artificial Intelligence Laboratory, MIT, from 1982 to 1986, and followed this with a faculty position with the Division of Applied Sciences, Harvard, from 1986 to 1995, rising to the position of an Associate Professor. From 1995 to 2002, he was a Senior Scientist with the Smith-Kettlewell Eye Research Institute, San Francisco. In 2002, he accepted a position as a Full Professor with the Department of Statistics, University of California, Los Angeles. He has over two hundred peer-reviewed publications in vision, neural networks, and physics, and has co-authored two books: *Data Fusion for Sensory Information Processing Systems* (with J. J. Clark) and Two- and Three-Dimensional Patterns of the Face (with P. W. Hallinan, G. G. Gordon, P. J. Giblin, and D. B. Mumford). He co-edited the book *Active Vision* (with A. Blake). He received several academic prizes.

Zhuowen Tu is an Assistant Professor with the Department of Cognitive Science, and the Department of Computer Science and Engineering, University of California, San Diego (UCSD). Before joining UCSD, he was an Assistant Professor with the University of California, Los Angeles. From 2011 to 2013, he took a leave to work with Microsoft Research Asia. He received the Ph.D. degre from the Ohio State University and the M.E. degree from Tsinghua University. He received the NSF CAREER Award in 2009 and the David Marr prize in 2003.