



A Mixture Model for Robust Point Matching under Multi-Layer Motion

Jiayi Ma¹, Jun Chen², Delie Ming^{1*}, Jinwen Tian¹

1 National Key Laboratory of Science & Technology on Multi-Spectral Information Processing, School of Automation, Huazhong University of Science and Technology, Wuhan, Hubei, China, **2** Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, Hubei, China

Abstract

This paper proposes an efficient mixture model for establishing robust point correspondences between two sets of points under multi-layer motion. Our algorithm starts by creating a set of putative correspondences which can contain a number of false correspondences, or outliers, in addition to the true correspondences (inliers). Next we solve for correspondence by interpolating a set of spatial transformations on the putative correspondence set based on a mixture model, which involves estimating a consensus of inlier points whose matching follows a non-parametric geometrical constraint. We formulate this as 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 non-parametric geometrical constraints on the correspondence, as a prior distribution, in a reproducing kernel Hilbert space (RKHS). 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 further provide a fast implementation based on sparse approximation which can achieve a significant speed-up without much performance degradation. We illustrate the proposed method on 2D and 3D real images for sparse feature correspondence, as well as a public available dataset for shape matching. The quantitative results demonstrate that our method is robust to non-rigid deformation and multi-layer/large discontinuous motion.

Citation: Ma J, Chen J, Ming D, Tian J (2014) A Mixture Model for Robust Point Matching under Multi-Layer Motion. PLoS ONE 9(3): e92282. doi:10.1371/journal.pone.0092282

Editor: Magnus Rattray, University of Manchester, United Kingdom

Received: July 12, 2013; **Accepted:** February 20, 2014; **Published:** March 21, 2014

Copyright: © 2014 Ma et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work is supported by the National Natural Science Foundation of China (Nos. 61273279 and 61273241, http://www.nsf.gov.cn/e_nsf/desktop/zn/0101.htm) and China Scholarship Council (No. 201206160008, <http://en.csc.edu.cn/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mingdelie@hust.edu.cn

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, camera self-calibration, tracking, image retrieval, and object recognition [1]. 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 [2]). 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 [3–5]. This strategy is commonly used for situations where the geometrical constraints are parametric, such as requiring that corresponding points lie on epipolar lines [1]. Examples of this strategy include the RANSAC algorithm [3] and analogous robust hypothesize-and-verify methods [4,6,7]. 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 [5].

Recently, some new non-parametric model-based methods have also been developed to deal with the non-rigidity, such as identifying point correspondences by correspondence function (ICF) [5], vector field consensus (VFC) [8,9], mismatch removal via coherent spatial mapping [10], as well as L_2 -Minimizing Estimate-based method (RPM- L_2E) [11]. These methods work well when the scene contains some deformable objects. However, since they fit a smooth transformation for the scene motion, when the motion contains large discontinuities or multi-layer, for

example, large depth discontinuities or motion inconsistencies, the smoothness prior will be violated and hence these methods will be badly degraded. Some graph matching based methods such as dual decomposition method [12] and graph shift (GS) [13] have also been proposed to capture different layers of motions, and hence robust to the motion with large discontinuities. However, these methods could not handle the 3D case.

In this paper, we generalize the former non-parametric model-based methods to deal with large discontinuities. Rather than these methods which interpolates a global transformation, our approach uses a mixture model and fits the correspondences with a set of transformations. This enables us to capture multi-layer motion, and hence robust to large discontinuous motion. To illustrate the main idea of our approach, 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 the left of Fig. 1, which contains a number of outliers. Due to the motions of the fox and the ground are quite different in the scene, it is hard to fit a global transformation smooth enough on all the inlier correspondences. Therefore, the non-parametric model-based methods such as VFC [8] will only preserve the majority of the inliers which locate on the ground, as shown in the middle of Fig. 1. However, our approach fits multiple transformations, which can capture multi-layer motion, and hence both the inliers on the fox and the ground will be preserved, as shown in the right of Fig. 1.

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 methods like RANSAC which use robust criteria for performing correspondence assuming parametric geometric constraints. Next we discuss approaches for solving matching problems which solve for a correspondence matrix between point sets.

There has been considerable study of robust estimation in the statistics literature [14,15]. 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 [14]. 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).

The RANSAC algorithm matches two point sets by first computing a putative set and then using robust methods to impose parametric geometric constraints [3]. RANSAC uses a hypothe-

size-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 [4], LO-RANSAC [16] and PROSAC [6]. 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 [7].

In the recent past, some new non-parametric model-based methods have also been developed, such as ICF [5], VFC [8,17], RPM- L_2E [11]. The ICF rejects outliers by learning a correspondence function pair which maps points in one image to their corresponding points in another. While the VFC converts the outlier rejection problem into a robust vector field interpolation problem which interpolates a non-parametric smooth motion field to fit the potential inliers. Similar to VFC, the RPM- L_2E also fits a non-parametric spatial transformation, and the difference is that it uses a robust estimator to deal with outliers rather than explicitly modeling the outlier distribution in VFC.

Another strategy for point correspondences is to formulate this problem in terms of a correspondence matrix between points (in the two datasets) together with a parametric, or non-parametric, geometric constraint [18–22]. These approaches relate closely to earlier work on mathematical models of human perception of long-range motion. This includes Ullman’s minimal mapping theory [23] and Yuille and Grzywacz’s motion coherence theory [24] which formulate correspondence in terms of vector field interpolation and use Gaussian kernels. The iterated closest point (ICP) algorithm [18] is one of the best known point correspondence/registration approaches. It uses nearest-neighbor relationships to assign a binary correspondence, and then uses estimated correspondence to refine the transformation. Efficient versions of ICP use sampling processes, either deterministic or based on heuristics [25]. The nearest point strategy of ICP can be replaced by soft assignments within a continuous optimization framework, e.g., the TPS-RPM [20,26]. In the recent past, the point registration is typically solved by probabilistic methods [21,22,27,28]. The kernel correlation based method [27] models each one of the two point sets by two probability distributions and measures the dissimilarity between the two distributions. It was later improved in [22]. In [21] as well as in [22] and [28], the Gaussian mixture model is used to recast the point-to-point assignment problem into that of estimating the parameters of a



Figure 1. Schematic illustration of our approach for point matching. Left: a set of putative correspondences computed by SIFT matching. Middle: point matching results of a non-parametric model-based method, e.g., VFC [8]. Right: point matching results of our approach; different color indicates different component of the mixture model in our approach. doi:10.1371/journal.pone.0092282.g001

mixture. This is done within the framework of maximum likelihood and the expectation-maximization (EM) algorithm [29].

Point correspondence has also been formulated as a graph matching problem, such as the dual decomposition (DD) [12], Spectral Matching (SM) [30], and graph shift (GS) [13,31]. 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. The SIFT-flow algorithm [32] 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 [24].

Method

Problem Formulation

Suppose now we are given a set of putative image point correspondences $S = \{(\mathbf{x}_n, \mathbf{y}_n) : n \in \mathbb{I}N_K\}$, which may be perturbed by noise and outliers. The non-parametric model-based methods [8,11] aim to fit a non-parametric transformation \mathbf{f} to the underlying inliers, i.e., $\mathbf{y}_n = \mathbf{f}(\mathbf{x}_n)$ for any inlier $(\mathbf{x}_n, \mathbf{y}_n)$, and hence remove outliers. The transformation \mathbf{f} here is continuous and smooth, which is not able to handle multi-layer motion. In this case, a more robust method is desirable to provide stable results. To this end, we consider a mixture model and fit the correspondences with a set of transformations $\{\mathbf{f}_k : k \in \mathbb{I}N_K\}$ instead of just one global transformation.

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 (D is the dimension of the data, typically $D = 2$ or 3), and the distribution is assumed to be uniform $\frac{1}{a}$ with a being a constant. Let us introduce a set of latent variables $\{\mathbf{z}_n : n \in \mathbb{I}N_K\}$, where \mathbf{z}_n has an 1-of- $K+1$ representation in which a particular element z_{nk} is equal to 1 and all other elements are equal to 0. More specifically, the notation $z_{nk} = 1$ ($k \in \mathbb{I}N_K$) means that the n -th correspondence is an inlier and it corresponds to the transformation \mathbf{f}_k , while $z_{n,K+1} = 1$ means that the n -th correspondence is an outlier. The marginal distribution over \mathbf{z}_n is specified in terms of the mixing coefficients π_k , such that

$$p(z_{nk} = 1) = \pi_k, \tag{1}$$

where the parameters $\{\pi_k : k \in \mathbb{I}N_{K+1}\}$ satisfy $0 \leq \pi_k \leq 1$ together with

$$\sum_{k=1}^{K+1} \pi_k = 1 \tag{2}$$

in order to be valid probabilities. The likelihood of a correspondence $(\mathbf{x}_n, \mathbf{y}_n)$ corresponding to the transformation \mathbf{f}_k is drawn from a Gaussian distribution with mean 0 and covariance $\sigma^2 \mathbf{I}$:

$$p(\mathbf{x}_n, \mathbf{y}_n | z_{nk} = 1, \theta) = \mathcal{N}(\mathbf{y}_n - \mathbf{f}_k(\mathbf{x}_n) | \mathbf{0}, \sigma^2 \mathbf{I}), \quad \forall k, 1 \leq k \leq K, \tag{3}$$

where $\theta = \{\mathbf{f}, \sigma^2, \pi_k\}$ includes a set of unknown parameters. We denote the set of all observed data by \mathbf{X} and \mathbf{Y} , in which the n -th rows represent \mathbf{x}_n^T and \mathbf{y}_n^T respectively, and similarly we denote the set of all latent variables by \mathbf{Z} . By making the i.i.d. data assumption, the likelihood is a mixture model given by

$$\begin{aligned} p(\mathbf{X}, \mathbf{Y} | \theta) &= \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Y}, \mathbf{Z} | \theta) = \prod_{n=1}^N \sum_{\mathbf{z}_n} p(\mathbf{x}_n, \mathbf{y}_n, \mathbf{z}_n | \theta) \\ &= \prod_{n=1}^N \left(\sum_{k=1}^K \frac{\pi_k}{(2\pi\sigma^2)^{D/2}} e^{-\frac{\|\mathbf{y}_n - \mathbf{f}_k(\mathbf{x}_n)\|^2}{2\sigma^2}} + \frac{\pi_{K+1}}{a} \right). \end{aligned} \tag{4}$$

distribution function is nonzero only in a bounded region (here we omit the indicator function for clarity).

We want to recover the transformations $\{\mathbf{f}_k, k \in \mathbb{I}N_K\}$ from the data S . Taking a probabilistic approach, we assume \mathbf{f}_k to be a realization of a random field with a known prior probability distribution $p(\mathbf{f}_k)$. The prior is used to impose constraints on \mathbf{f}_k , assigning significant probability only to those functions that satisfy those constraints. We consider the slow-and-smooth model [33] which has been shown to account for a range of motion phenomena, the prior of \mathbf{f}_k then has the form:

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

where $\phi(\mathbf{f}_k)$ is a smoothness functional and λ a positive real number (we will discuss the details of \mathbf{f}_k later). By applying Bayes rule, we seek a MAP solution of θ , i.e.,

$$\theta^* = \arg \max_{\theta} p(\theta | \mathbf{X}, \mathbf{Y}) = \arg \max_{\theta} p(\mathbf{X}, \mathbf{Y} | \theta) \prod_{k=1}^K p(\mathbf{f}_k). \tag{6}$$

This is equivalent to seeking the minimal energy

$$E(\theta) = - \sum_{n=1}^N \ln \sum_{\mathbf{z}_n} p(\mathbf{x}_n, \mathbf{y}_n, \mathbf{z}_n | \theta) - \sum_{k=1}^K \ln p(\mathbf{f}_k). \tag{7}$$

The set of transformations $\{\mathbf{f}_k, k \in \mathbb{I}N_K\}$ will be directly obtained from the optimal solution θ^* , and the latent variables $\{\mathbf{z}_n : n \in \mathbb{I}N_N\}$ determine the inliers. In the next section, we show how to solve the estimation problem using an EM approach.

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 [29] 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 [34] and omit some terms that are independent of θ . Denote $p(z_{nk} = 1 | \mathbf{x}_n, \mathbf{y}_n, \theta)$ by $\gamma(z_{nk})$, considering the negative log posterior function, i.e. equation (7), the complete-data log posterior is then given by

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{old}) = & -\frac{1}{2\sigma^2} \sum_{n=1}^N \sum_{k=1}^K \gamma(z_{nk}) \|\mathbf{y}_n - \mathbf{f}_k(\mathbf{x}_n)\|^2 \\ & - \frac{D}{2} \ln \sigma^2 \sum_{n=1}^N \sum_{k=1}^K \gamma(z_{nk}) \\ & + \sum_{n=1}^N \sum_{k=1}^{K+1} \gamma(z_{nk}) \ln \pi_k - \frac{\lambda}{2} \sum_{k=1}^K \phi(\mathbf{f}_k). \end{aligned} \tag{8}$$

E-step: We use the current parameter values $\boldsymbol{\theta}^{old}$ to find the posterior distribution of the latent variables, which can be found by applying Bayes' rules

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{y}_n - \mathbf{f}_k(\mathbf{x}_n) | \mathbf{0}, \sigma^2 \mathbf{I})}{\sum_{j=1}^k \pi_j \mathcal{N}(\mathbf{y}_n - \mathbf{f}_j(\mathbf{x}_n) | \mathbf{0}, \sigma^2 \mathbf{I}) + \pi_{K+1}/a}, \tag{9}$$

where $k \in \mathbb{I}_K$ and $n \in \mathbb{I}_N$. The posterior probability $\gamma(z_{nk})$ indicates to what degree the n -th sample agrees with the current estimated transformation \mathbf{f}_k .

M-step: We determine the revised parameter estimate $\boldsymbol{\theta}^{new}$ as follows: $\boldsymbol{\theta}^{new} = \arg \max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{old})$. Taking derivative of $\mathcal{Q}(\boldsymbol{\theta})$ with respect to π_k and σ^2 , and setting them to zero, together with equation (2), we obtain

$$\pi_k = N_k/N, \tag{10}$$

$$\sigma^2 = \frac{\sum_{n=1}^N \sum_{k=1}^K \gamma(z_{nk}) \|\mathbf{y}_n - \mathbf{f}_k(\mathbf{x}_n)\|^2}{D \sum_{k=1}^K N_k}, \tag{11}$$

where $N_k = \sum_{n=1}^N \gamma(z_{nk})$.

Next we consider the terms of $\mathcal{Q}(\boldsymbol{\theta})$ that are related to \mathbf{f}_k . We obtain a regularized risk functional as [8]:

$$\mathcal{E}(\mathbf{f}_k) = \frac{1}{2\sigma^2} \sum_{n=1}^N \gamma(z_{nk}) \|\mathbf{y}_n - \mathbf{f}_k(\mathbf{x}_n)\|^2 + \frac{\lambda}{2} \phi(\mathbf{f}_k). \tag{12}$$

We model \mathbf{f}_k by requiring it to lie within a specific functional space \mathcal{H} , namely a vector-valued reproducing kernel Hilbert space (RKHS) [35]. We define the RKHS \mathcal{H} by a diagonal Gaussian kernel [8]: $\Gamma(\mathbf{x}_i, \mathbf{x}_j) = e^{-\beta \|\mathbf{x}_i - \mathbf{x}_j\|^2} \mathbf{I}$. 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 [36].

Theorem 1. *The optimal solution of the regularized risk functional (12) is given by*

$$\mathbf{f}_k(\mathbf{x}) = \sum_{n=1}^N \Gamma(\mathbf{x}, \mathbf{x}_n) \mathbf{c}_{nk}, \quad k \in \mathbb{I}_K \tag{13}$$

with the coefficient set $\{\mathbf{c}_{nk} : n \in \mathbb{I}_N, k \in \mathbb{I}_K\}$ determined by a linear system

$$(\mathbf{\Gamma} + \lambda \sigma^2 \mathbf{P}_k^{-1}) \mathbf{C}_k = \mathbf{Y}, \tag{14}$$

where $\mathbf{\Gamma} \in \mathbb{R}^{N \times N}$ is the kernel matrix with the ij -th entry

$\Gamma_{ij} = e^{-\beta \|\mathbf{x}_i - \mathbf{x}_j\|^2}$, $\mathbf{P}_k = \text{diag}(\gamma(z_{1k}), \dots, \gamma(z_{Nk}))$ is a diagonal matrix, and $\mathbf{C}_k = (\mathbf{c}_{1k}, \dots, \mathbf{c}_{Nk})^T \in \mathbb{R}^{N \times D}$.

Proof. The kernel $\mathbf{\Gamma}$ has the following reproducing property, for all $\mathbf{c} \in \mathcal{Y}$ and $\mathbf{x} \in \mathcal{X}$

$$\langle \mathbf{f}_k(\mathbf{x}), \mathbf{c} \rangle = \langle \mathbf{f}_k, \mathbf{\Gamma}_x \mathbf{c} \rangle_{\mathcal{H}}, \tag{15}$$

where $\mathbf{\Gamma}_x \mathbf{c} = \mathbf{\Gamma}(\cdot, \mathbf{x}) \mathbf{c}$. For any given reproducing kernel $\mathbf{\Gamma}$, a unique RKHS can be defined as

$$\mathcal{H}_N = \left\{ \sum_{n=1}^N \Gamma_{\mathbf{x}_n} \mathbf{c}_n \mid \mathbf{c}_n \in \mathcal{Y}, n=1, \dots, N \right\}. \tag{16}$$

For further details about the reproducing property we refer the readers to [36–38].

Let \mathcal{H}_N^\perp be a subspace of \mathcal{H} ,

$$\mathcal{H}_N^\perp = \{\mathbf{f}_k \in \mathcal{H} : \mathbf{f}_k(\mathbf{x}_n) = 0, n \in \mathbb{I}_N\}. \tag{17}$$

From the reproducing property, i.e. equation (15), $\forall \mathbf{f}_k \in \mathcal{H}_N^\perp$

$$\langle \mathbf{f}_k, \sum_{n=1}^N \Gamma(\cdot, \mathbf{x}_n) \mathbf{c}_{nk} \rangle_{\mathcal{H}} = \sum_{n=1}^N \langle \mathbf{f}_k(\mathbf{x}_n), \mathbf{c}_{nk} \rangle = 0. \tag{18}$$

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

$$\begin{aligned} \mathcal{E}(\mathbf{f}_k) &= \frac{1}{2\sigma^2} \sum_{n=1}^N \gamma(z_{nk}) \|\mathbf{y}_n - \mathbf{f}_k(\mathbf{x}_n)\|^2 + \frac{\lambda}{2} \|\mathbf{f}_{kN} + \mathbf{f}_{kN}^\perp\|_{\mathcal{H}}^2 \\ &\geq \frac{1}{2\sigma^2} \sum_{n=1}^N \gamma(z_{nk}) \|\mathbf{y}_n - \mathbf{f}_{kN}(\mathbf{x}_n)\|^2 + \frac{\lambda}{2} \|\mathbf{f}_{kN}\|_{\mathcal{H}}^2. \end{aligned} \tag{19}$$

Therefore, the optimal solution of the regularized risk functional (12) comes from the space \mathcal{H}_N , and hence has the form (13). To solve for the coefficients, we rewrite the regularized risk functional in the following matrix form:

$$\mathcal{E}(\mathbf{f}_k) = \frac{1}{2\sigma^2} \|\mathbf{P}^{1/2}(\mathbf{Y} - \mathbf{\Gamma} \mathbf{C}_k)\|_F^2 + \frac{\lambda}{2} \text{tr}(\mathbf{C}_k^T \mathbf{\Gamma} \mathbf{C}_k). \tag{20}$$

where $\mathbf{\Gamma} \in \mathbb{R}^{N \times N}$ is the kernel matrix with the ij -th entry $\Gamma_{ij} = e^{-\beta \|\mathbf{x}_i - \mathbf{x}_j\|^2}$, $\mathbf{P}_k = \text{diag}(\gamma(z_{1k}), \dots, \gamma(z_{Nk}))$ is a diagonal matrix, $\mathbf{C}_k = (\mathbf{c}_{1k}, \dots, \mathbf{c}_{Nk})^T \in \mathbb{R}^{N \times D}$ is the coefficient matrix, $\|\cdot\|$ is the Frobenius norm, and $\text{tr}(\cdot)$ denotes the trace. Taking the derivative of the last equation with respect to \mathbf{C}_k and setting it to zero, we obtain the linear system in equation (14). Thus the coefficient set $\{\mathbf{c}_{nk} : n \in \mathbb{I}_N\}$ of the optimal solution \mathbf{f}_k is determined by the linear system (14).

Once the EM algorithm converges, we then obtain the set of transformations $\{\mathbf{f}_k : k \in \mathbb{I}_K\}$. Besides, with a predefined thresh-

old τ we obtain the inliers \mathcal{I} as well:

$$\mathcal{I} = \{n : \sum_{k=1}^K \Gamma(z_{nk}) > \tau, n \in \mathbb{I}_N\} \quad (21)$$

We summarize our method in Table 1. Since our robust point matching method is based on a mixture model, we named our method *RPM-MM*.

Fast Implementation

Solving the transformation \mathbf{f}_k merely requires to solve the linear system (14). 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. To address this problem, in this section we provide a fast implementation based on a similar kind of idea as the subset of regressors method [17,39].

Rather than searching for the optimal solution in \mathcal{H}_N i.e., equation (16), we use a sparse approximation and search a suboptimal solution in a space \mathcal{H}_M with much less basis functions defined as

$$\mathcal{H}_M = \left\{ \sum_{m=1}^M \Gamma(\cdot, \tilde{\mathbf{x}}_m) \mathbf{c}_m : \mathbf{c}_m \in \mathcal{Y} \right\}, \quad (22)$$

and then minimize the regularized risk functional over all the sample data. Here $M \ll N$ and we choose the point set $\{\tilde{\mathbf{x}}_m : m \in \mathbb{I}_M\}$ as a random subset of $\{\mathbf{x}_n : n \in \mathbb{I}_N\}$ according to [17]. 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

$$\mathbf{f}_k(x) = \sum_{m=1}^M \Gamma(\mathbf{x}, \tilde{\mathbf{x}}_m) \mathbf{c}_{mk}, \quad k \in \mathbb{I}_K$$

Table 1. Algorithm 1: The RPM-MM Algorithm.

Input: Correspondences $S = \{(\mathbf{x}_n, \mathbf{y}_n) : n \in \mathbb{I}_N\}$, parameters λ, β, τ
Output: Transformations $\{\mathbf{f}_k : k \in \mathbb{I}_K\}$, inliers \mathcal{I}
1. Initialize $\gamma(z_{nk}), \mathbf{C}_k = \mathbf{0}_{N \times D}$;
2. Compute π_k, σ^2 by equations (10) and (11);
3. Set the constant a and compute the kernel matrix Γ ;
4. repeat
5. <i>E-step:</i>
6. Update $\gamma(z_{nk})$ by equation (9);
7. <i>M-step:</i>
8. Update \mathbf{C}_k by solving linear system (14);
9. Compute $\{\mathbf{f}_k(\mathbf{x}_n)\}$ by equation (13);
10. Update π_k and σ^2 by equations (10) and (11);
11. until \mathcal{Q} converges;
12. $\{\mathbf{f}_k : k \in \mathbb{I}_K\}$ and \mathcal{I} are determined by equations (13) and (21).

doi:10.1371/journal.pone.0092282.t001

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

$$(\mathbf{U}^T \mathbf{P}_k \mathbf{U} + \lambda \sigma^2 \mathbf{\Gamma}^s) \mathbf{C}_k^s = \mathbf{U}^T \mathbf{P}_k \mathbf{Y}, \quad (24)$$

where $\mathbf{C}_k^s = (\mathbf{c}_{1k}, \dots, \mathbf{c}_{Mk})^T$ is the coefficient matrix, $\mathbf{\Gamma}^s \in \mathbb{R}^{M \times M}$ with the (i,j) -th element $e^{-\beta \|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|^2}$, $\mathbf{U} \in \mathbb{R}^{N \times M}$ with the (i,j) -th element $e^{-\beta \|\mathbf{x}_i - \tilde{\mathbf{x}}_j\|^2}$.

In contrast to the optimal solution given by the representer theorem, which is a linear combination of the basis functions $\{\Gamma(\cdot, \mathbf{x}_n) : n \in \mathbb{I}_N\}$, 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. Compared with the original algorithm shown in Algorithm 1, the fast version solves a different linear system (24) in Line 8.

Computational Complexity

For the RPM-MM algorithm, the corresponding kernel matrix $\mathbf{\Gamma}$ is of size $N \times N$. According to Theorem 1, we need to solve a linear system (14) for each transformation \mathbf{f}_k . The time complexity is $O(DN^3)$, which is the most time-consuming step in the algorithm. As a result, the total time complexity of our algorithm is $O(mKDN^3)$, 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 RPM-MM scales like $O(N^2)$ due to the memory requirements for storing the kernel matrix $\mathbf{\Gamma}$.

For the fast implementation, the corresponding kernel matrix is of size $M \times M$, where M is the number of basis functions used for sparse representation. Then the time complexity is reduced to $O(mKDM^2N + mKDM^3)$, and the space complexity is reduced to $O(MN + M^2)$. Typically, in point matching problems, the number of the point matches N is in the order of 10^3 , and the required number of basis function M is in the order of 10^1 . Therefore, both the time and space complexities can be simply written as $O(N)$. This is significant for large datasets. Our experiments demonstrate that the fast version is much faster than the original RPM-MM algorithm with negligible performance degradation.

Table 2. Algorithm 2: Non-Rigid Point Set Registration via RPM-MM.

Input: Two point sets $\{\mathbf{x}_n\}_{n=1}^N, \{\mathbf{y}_l\}_{l=1}^L$
Output: Aligned model point set $\{\tilde{\mathbf{x}}_n\}_{n=1}^N$
1. Compute feature descriptors for the target point set $\{\mathbf{y}_l\}_{l=1}^L$;
2. repeat
3. Compute feature descriptors for the model point set $\{\mathbf{x}_n\}_{n=1}^N$;
4. Estimate the initial correspondences based on feature descriptors;
5. Solve the transformations $\{\mathbf{f}_k\}_{k=1}^K$ by using RPM-MM;
6. Warp the model point set according to the transformations $\{\mathbf{f}_k\}_{k=1}^K$;
7. Update the model point set by using the warped model point set;
8. until reach the maximum iteration number;
9. $\{\tilde{\mathbf{x}}_n\}_{n=1}^N$ is given by the warped model point set in the last iteration.

doi:10.1371/journal.pone.0092282.t002

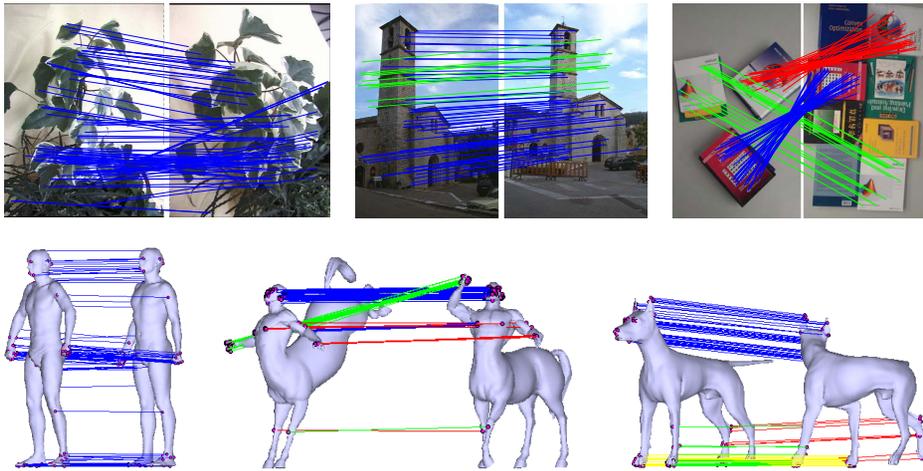


Figure 2. Matching results on 2D image pairs (*Tree, Church and Books*) and 3D surface pairs (*Person, Centaur and Dog* [42]). Different color denotes different component of the mixture model. For visibility, in the image, we only show at most 100 randomly selected elements of the preserved correspondences by our method.
doi:10.1371/journal.pone.0092282.g002

Extension to 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 $\{\mathbf{y}_i\}_{i=1}^L$ (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. Moreover, for point sets with multi-layer motion, it may need multiple transformations to achieve satisfying results. Recall that our RPM-MM method is able to generate a set of non-rigid transformations with adherence to a set of point correspondences. Therefore, it could be used to recover the transformation(s) between two point sets with a set of putative correspondences. Next we discuss how to establish initial correspondences.

Typically, for a pure point matching problem, the appearance information is not available. In general, if the two point sets have similar shapes, the corresponding points have similar neighborhood structures which could be incorporated into a feature descriptor. Thus finding correspondences between two point sets is equivalent to finding for each point in one point set (e.g., the model) the point in the other point set (e.g., the target) that has the most similar feature descriptor. Fortunately, the initial correspondences need not be very accurate, since our method is robust to noise and outliers. Inspired by these facts, we use shape context [40] as the feature descriptor, using the Hungarian method for matching with the χ^2 test statistic as the cost measure.

Table 3. Performance comparison on 2D image pairs in Fig. 2: *Tree, Church, and Books*.

Inlier pct.	56.29%	54.76%	75.74%
RANSAC [3]	(94.68, 94.68)	(94.52, 100.00)	-
ICF [5]	(92.75, 68.09)	(91.67, 63.77)	(91.24, 40.53)
GS [13]	(97.62, 87.23)	(91.78, 97.10)	(100.00, 82.48)
VFC [8]	(94.85, 97.87)	(98.33, 85.51)	(97.79, 70.44)
RPM-MM	(94.85, 97.87)	(97.14, 98.57)	(99.82, 98.05)
RPM-MM (fast)	(94.85, 97.87)	(95.77, 98.55)	(99.82, 98.23)

The pair denotes the precision-recall pair (%).
doi:10.1371/journal.pone.0092282.t003

After we get the rough correspondences between two point sets based on their shape features, we fit a set of transformations and use them to warp the model points. To this end, we need to determine the attribution of each model point (i.e., which component of the mixture model a model point belongs to). First, according to the matching results based on the rough correspondences, we can determine the attributions of the inliers; for the rest of the model points, we attribute them to their nearest neighbors which already have attributions. The two steps of estimating correspondences and transformations are iterated to obtain a reliable result. In this paper, we use a fixed number of iterations, typically 10 but more when there are large degradations on the data. We summarize our non-rigid point set registration method in Table 2.

Implementation Details

The performance of point matching algorithms typically depends on the coordinate system in which points are expressed; here we use data normalization to control for this. More specifically, we perform a linear re-scaling of the correspondences so that the points in the two sets both have zero mean and unit variance. Furthermore, we define the transformation \mathbf{f}_k as the initial position plus a displacement function v_k : $\mathbf{f}_k(\mathbf{x}) = \mathbf{x} + \mathbf{v}_k(\mathbf{x})$ [11], and solve for \mathbf{v}_k instead of \mathbf{f}_k . This can be achieved simply by setting the output \mathbf{y}_n to be $\mathbf{y}_n - \mathbf{x}_n$. The use of displacement function achieves more robustness.

The EM algorithm is well known to converge to a local maximum. To initialize the EM iteration, we first use the K-means

Table 4. Performance comparison on 3D surface pairs in Fig. 2: *Person, Centaur, and Dog*.

Inlier pct.	56.40%	78.23%	86.18%
VFC [8]	(99.22, 98.46)	(99.53, 78.85)	(96.58, 82.65)
RPM-MM	(99.22, 98.46)	(97.11, 93.51)	(95.71, 96.27)
RPM-MM (fast)	(99.22, 98.46)	(97.04, 93.09)	(95.74, 96.40)

The pair denotes the precision-recall pair (%).
doi:10.1371/journal.pone.0092282.t004

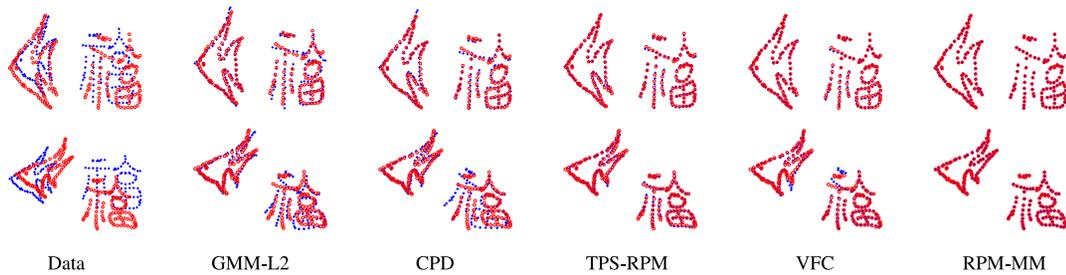


Figure 3. Point set registration results. The data contains two independent shape models with different movements and distortions. The goal is to align the model point set (blue pluses) onto the target point set (red circles). From left to right: model and target point sets, registration results of GMM- L_2 [22], CPD [21], TPS-RPM [20], VFC [8] and RPM-MM. doi:10.1371/journal.pone.0092282.g003

algorithm to cluster the correspondences into K_0 clusters. To this end, we convert the correspondences into a set of motion field samples $\{\mathbf{y}_n - \mathbf{x}_n\}_{n=1}^N$ and do clustering on them. The first K largest clusters are considered as inliers of the K -component mixture model in our approach, and then are used to initialize $\{\gamma(z_{nk}) \in \{0, 1\} : n \in \mathbb{N}_N, k \in \mathbb{N}_K\}$. In our evaluation, we set $K_0 = 10$, and K can be set adaptively according to the cardinalities of the clusters. For instance, if the ratio of the cardinalities between a certain cluster and the largest cluster is bigger than a threshold, e.g., 0.2, then it will be preserved as initial inliers, typically $K = 2$ or 3.

There are mainly three parameters in our algorithm: β , λ and τ . Parameters β and λ both reflect the amount of smoothness regularization. Parameter β determines how wide the range of interaction between correspondences. 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. In general, we find our method to be very robust to parameter changes. We set $\beta = 0.1$, $\lambda = 1$, and $\tau = 1/K$ throughout the experiments. The constant a in the uniform distribution is set to the volume of the output space (e.g., the bounding box of the points) after data normalization, typically $a = 10$. Besides, for the fast implementation, the number M of basis functions used for sparse approximation is fixed to 15.

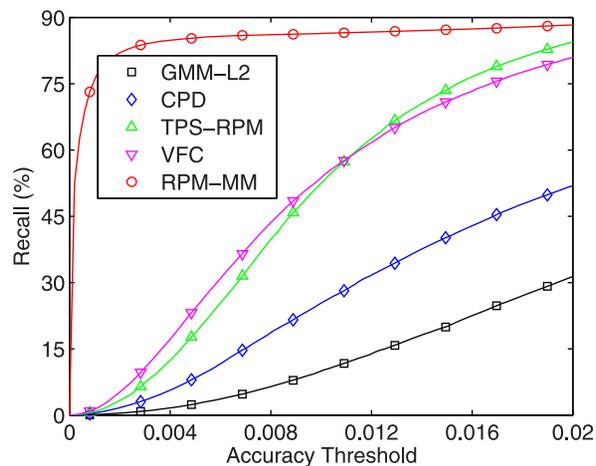


Figure 4. Performances of non-rigid point set registration algorithms over 100 samples. doi:10.1371/journal.pone.0092282.g004

Experimental Results

In order to evaluate the performance of our algorithm, we conducted two types of experiments: i) sparse image feature correspondence on 2D images and 3D surfaces; ii) non-rigid shape matching on synthetic 2D datasets.

Results on Sparse Feature Correspondence

We present a few representative matching results on both 2D image pairs and 3D surface pairs, as shown in Fig. 2. For the 2D images, the first two pairs (i.e., *Tree* and *Church*) are wide baseline images which are public available [41]. The third pair (i.e., *Books*) contains some shared content which is taken by ourselves. It often occurs in image or object retrieval. The 3D surface pairs (i.e., *Person*, *Centaur* and *Dog*) are objects with different amounts of non-rigid deformation, which come from a surface correspondence benchmark [42]. In our evaluation, we use the SIFT [2] and MeshDOG/MeshHOG [43] to establish putative correspondences for 2D and 3D cases, respectively.

The match correctness is determined as follows. For the 2D images, a method combining subjectivity and objectivity is considered. We first fit the epipolar geometry (e.g., fundamental matrix) by RANSAC and use it to determine the match correctness. We further confirm them artificially. Although the judgment of correct match and mismatch seems arbitrary, we make the benchmark before performing experiments to ensure objectivity. For the 3D surfaces, the ground truth correspondences are supplied by the dataset.

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 RPM-MM algorithm with other four methods which remove outliers from given putative point correspondences, such as RANSAC [3], ICF [5], GS [13], and VFC [8]. We implement ICF and tune all parameters accordingly to find optimal settings. For RANSAC, GS and VFC, we implement them based on the publicly available codes. Throughout all the experiments, five algorithms' parameters are all fixed.

The results of our RPM-MM are presented in Fig. 2, we see that for an image pair with relatively simple structures, such as *Tree* or *Person* which involves a small amount of rotation, viewpoint change, or non-rigid deformation, the mixture model in our algorithm then degenerates and contains only one component. In this case Our RPM-MM algorithm is equal to a normal non-parametric model-based method, e.g., VFC [8]. For images with large view point change or non-rigid deformation, the mixture

model will contain multiple components to capture multi-layer motion, such as *Church*, *Books*, *Centaur* and *Dog*. Note that in the image pair of *Church*, our RPM-MM preserves the correspondences on the sky; this is very useful since removing inliers outstanding in depth tend to make the recovery of epipolar geometry unstable and ill conditioned [44]. Moreover, the matching result of *Books* shows our method's capability in image retrieval.

We further quantitatively compare our RPM-MM to four state-of-the-art point matching algorithms: RANSAC, ICF, GS and VFC. Tables 3 and 4 report the 2D and 3D results respectively. As shown in Table 3, ICF and VFC have low recalls when the scene contains large discontinuities. In fact, they tend to preserve typically one major component (i.e., the matches marked by blue lines in Fig. 2) or two components of the correspondences. RANSAC has satisfactory performance when the relationship of correspondence is rigid, e.g., epipolar geometry. But it can not work in the non-rigid case, e.g., *Books*. The graph matching based method GS generally can obtain better performance than ICF and VFC in case of large discontinuity, e.g., *Books*. But its recalls are still relatively low compared to our RPM-MM. In Table 4 we only use VFC for comparison since the other three methods are not applicable for either the 3D case or non-rigid deformation. We again observe that VFC fails to keep most of the inliers under large non-rigid deformations. Our RPM-MM in general has the best precision-recall trade-off, and it is not affected by large non-rigid deformation or multi-layer motion. In addition, we also test the fast version of our method on these six image pairs, as shown in the last rows of Tables 3 and 4. The average number of putative correspondences on these six image pairs is about 630, and the average run times of our RPM-MM and its fast implementation are about 6 s and 0.2 s on an Intel Core 2.0 GHz PC with Matlab code. We see that the use of sparse approximation leads to an essential speedup without much performance degradation.

Results on Shape Matching

We next evaluate our RPM-MM on the non-rigid point set matching problem, more specifically, shape matching. For the dataset, we choose the same synthesized data as in [20]. The data consists of two different shape models, where the first model consists of 96 points representing a fish shape, and the second model is a more complex pattern consisting of 108 points representing a Chinese character. We combine the two shapes to generate a model point set, i.e., the blue pluses as shown in Fig. 3. To get a target point set, we apply two randomly generated non-rigid transformations on the model point set and warp the two shape models respectively, i.e., the red circles as shown in Fig. 3. The model is warped through progressively larger degrees of non-rigid warpings, and we generate 100 samples in total.

Fig. 3 presents some registration results of our RPM-MM and four other state-of-the-art registration algorithms: GMM- L_2 [22], CPD [21], TPS-RPM [20] and VFC [8], which are implemented using publicly available codes. In the first row, the data contains a relatively slight deformation; in this case, it is possible to search a

single transformation which approximates the two-layer motion well. Therefore, all the five algorithms are able to generate satisfying results. However, the matching performance degrades gradually as the degree of deformation increases. In the second row, the data contains a relatively large deformation; in this case, just one transformation cannot capture the two-layer motion, and hence the matching performance degrades. By contrast, our RPM-MM uses a mixture model which interpolates a set of transformations, and hence is robust to the two-layer motion. Therefore, it still can produce an almost perfect alignment, as shown in the last column of Fig. 3.

To provide a quantitative comparison, we report the registration results of the five algorithms on all the 100 samples. We compute the recall as the metric used in [22]. Here the recall, or true positive rate, is defined as the proportion of true positive correspondences to the ground truth correspondences and a true positive correspondence is counted when the pair falls within a given accuracy threshold in terms of pairwise distance, e.g., the Euclidean distance between a point in the warped model and the corresponding point in the target. Fig. 4 plots the recall curves of the five algorithms on all the 100 samples. We see that VFC and TPS-RPM perform much better than CPD and GMM- L_2 , and our RPM-MM is far better than the other four methods. This is not surprise since our RPM-MM interpolates multiple transformations to capture multi-layer motion while the other four algorithms interpolate just a global transformation. That is to say, for point sets with multi-layer motion, it is necessary to interpolate multiple local transformations (as in our RPM-MM) rather than only one global transformation to achieve accurate results. We also tested the fast version of our algorithm on this dataset and obtained almost the same recall curve as the original version, here we omit it in the figure for clarity.

Conclusion

Within this paper, we have proposed and studied a new robust point matching algorithm based on a mixture model (RPM-MM). It interpolates a set of transformations to fit different layers of the motion correspondence by an iterative EM algorithm, and hence establish reliable correspondence between two images. Quantitative comparisons on both sparse feature correspondence and shape matching demonstrate that our algorithm outperforms state-of-the-art point matching methods, especially when the motion of the scene contains large discontinuities.

Acknowledgments

We are very grateful to Ji Zhao for helpful suggestions.

Author Contributions

Conceived and designed the experiments: JM. Performed the experiments: JM JC DM JT. Analyzed the data: JM JC DM JT. Contributed reagents/materials/analysis tools: JM JC DM JT. Wrote the paper: JM.

References

- Hartley R, Zisserman A (2003) Multiple View Geometry in Computer Vision (2nd ed.). Cambridge: Cambridge University Press.
- Lowe D (2004) Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60: 91–110.
- Fischler MA, Bolles RC (1981) Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography. Communications of the ACM 24: 381–395.
- Torr PHS, Zisserman A (2000) MLESAC: A new robust estimator with application to estimating image geometry. Computer Vision and Image Understanding 78: 138–156.
- Li X, Hu Z (2010) Rejecting mismatches by correspondence function. International Journal of Computer Vision 89: 1–17.
- Chum O, Matas J (2005) Matching with PROSAC - progressive sample consensus. In: Proceedings of IEEE conference on Computer Vision and Pattern Recognition. pp. 220–226.
- Raguram R, Frahm JM, Pollefeys M (2008) A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus. In: Proceedings of European Conference on Computer Vision. pp. 500–513.

8. Zhao J, Ma J, Tian J, Ma J, Zhang D (2011) A robust method for vector field learning with application to mismatch removing. In: Proceedings of IEEE conference on Computer Vision and Pattern Recognition. pp. 2977–2984.
9. Ma J, Zhao J, Tian J, Yuille AL, Tu Z (2014) Robust point matching via vector field consensus. *IEEE Transactions on Image Processing*.
10. Ma J, Zhao J, Zhou Y, Tian J (2012) Mismatch removal via coherent spatial mapping. In: Proceedings of International Conference on Image Processing. pp. 1–4.
11. Ma J, Zhao J, Tian J, Tu Z, Yuille A (2013) Robust estimation of nonrigid transformation for point set registration. In: Proceedings of IEEE conference on Computer Vision and Pattern Recognition. pp. 2147–2154.
12. Torresani L, Kolmogorov V, Rother C (2008) Feature correspondence via graph matching: Models and global optimization. In: Proceedings of European Conference on Computer Vision. pp. 596–609.
13. Liu H, Yan S (2010) Common visual pattern discovery via spatially coherent correspondence. In: Proceedings of IEEE conference on Computer Vision and Pattern Recognition. pp. 1609–1616.
14. Huber PJ (1981) *Robust Statistics*. New York: John Wiley & Sons.
15. Rousseeuw PJ, Leroy A (1987) *Robust Regression and Outlier Detection*. New York: John Wiley & Sons.
16. Chum O, Matas J, Kittler J (2003) Locally optimized RANSAC. In: Proceedings of Pattern Recognition Symposium of the German Association for Pattern Recognition (DAGM). pp. 236–243.
17. Ma J, Zhao J, Tian J, Bai X, Tu Z (2013) Regularized vector field learning with sparse approximation for mismatch removal. *Pattern Recognition* 46: 3519–3532.
18. Besl PJ, McKay ND (1992) A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14: 239–256.
19. Gold S, Rangarajan A, Lu CP, Pappu S, Mjolsness E (1998) New algorithms for 2-d and 3-d point matching: Pose estimation and correspondence. *Pattern Recognition* 31: 1019–1031.
20. Chui H, Rangarajan A (2003) A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding* 89: 114–141.
21. Myronenko A, Song X (2010) Point set registration: Coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32: 2262–2275.
22. Jian B, Vemuri BC (2011) Robust point set registration using gaussian mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33: 1633–1645.
23. Ullman S (1979) *The Interpretation of Visual Motion*, volume 28. MIT press Cambridge, MA.
24. Yuille AL, Grzywacz NM (1989) A mathematical analysis of the motion coherence theory. *International Journal of Computer Vision* 3: 155–175.
25. Rusinkiewicz S, Levoy M (2001) Efficient variants of the icp algorithm. In: *3-D Digital Imaging and Modeling*. pp. 145–152.
26. Rangarajan A, Chui H, Bookstein F (1997) The softassign procrustes matching algorithm. In: *Information Processing in Medical Imaging*. pp. 29–42.
27. Tsin Y, Kanade T (2004) A correlation-based approach to robust point set registration. In: *Proceedings of European Conference on Computer Vision*. pp. 558–569.
28. Horaud R, Forbes F, Yguel M, Dewaele G, Zhang J (2011) Rigid and articulated point registration with expectation conditional maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33: 587–602.
29. Dempster A, Laird N, Rubin DB (1977) Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society Series B* 39: 1–38.
30. Leordeanu M, Hebert M (2005) A spectral technique for correspondence problems using pairwise constraints. In: *Proceedings of International Conference on Computer Vision*.
31. Liu H, Yan S (2010) Robust graph mode seeking by graph shift. In: *Proceedings of International Conference on Machine Learning*. pp. 671–678.
32. Liu C, Yuen J, Torralba A (2011) Sift ow: Dense correspondence across different scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.
33. Yuille AL, Grzywacz NM (1988) A computational theory for the perception of coherent visual motion. *Nature* 333: 71–74.
34. Bishop CM (2006) *Pattern Recognition and Machine Learning*. Springer.
35. Aronszajn N (1950) Theory of reproducing kernels. *Transactions of the American Mathematical Society* 68: 337–404.
36. Micchelli CA, Pontil M (2005) On learning vector-valued functions. *Neural Computation* 17: 177–204.
37. Carmeli C, De Vito E, Toigo A (2006) Vector valued reproducing kernel hilbert spaces of integrable functions and mercer theorem. *Analysis and Applications* 4: 377–408.
38. Baldassarre L, Rosasco L, Barla A, Verri A (2011) Multi-output learning via spectral filtering. Technical report, MIT Computer Science and Artificial Intelligence Laboratory, Cambridge.
39. Poggio T, Girosi F (1990) Networks for approximation and learning. *Proceedings of the IEEE* 78: 1481–1497.
40. Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24: 509–522.
41. Tuytelaars T, van Gool L (2004) Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision* 59: 61–85.
42. Kim VG, Lipman Y, Funkhouser T (2011) Blended intrinsic maps. *ACM Transactions on Graphics* 30: 79.
43. Zaharescu A, Boyer E, Varanasi K, Horaud R (2009) Surface feature detection and description with applications to mesh matching. In: *Proceedings of IEEE conference on Computer Vision and Pattern Recognition*. pp. 373–380.
44. Chum O, Werner T, Matas J (2005) Two-view geometry estimation unaffected by a dominant plane. In: *Proceedings of IEEE conference on Computer Vision and Pattern Recognition*. pp. 772–779.