Model-free Consensus Maximization
for Non-Rigid Shapes

Thomas Probst[1], Ajad Chhatkuli[1], Danda Pani Paudel[1], and Luc Van Gool[1,2]

[1] Computer Vision Lab, ETH Zürich, Switzerland
[2] VISICS, ESAT/PSI, KU Leuven, Belgium
{probstt,ajad.chhatkuli,paudel,vangool}@vision.ee.ethz.ch

Abstract. Many computer vision methods use consensus maximization to relate measurements containing outliers with the correct transformation model. In the context of rigid shapes, this is typically done using Random Sampling and Consensus (RANSAC) by estimating an analytical model that agrees with the largest number of measurements (inliers). However, small parameter models may not be always available. In this paper, we formulate the model-free consensus maximization as an Integer Program in a graph using 'rules' on measurements. We then provide a method to solve it optimally using the Branch and Bound (BnB) paradigm. We focus its application on non-rigid shapes, where we apply the method to remove outlier 3D correspondences and achieve performance superior to the state of the art. Our method works with outlier ratio as high as 80%. We further derive a similar formulation for 3D template to image matching, achieving similar or better performance compared to the state of the art.

1 Introduction

Consensus maximization is a powerful tool in computer vision that has enabled practical applications of highly complex algorithms such as Structure-from-Motion (SfM) [1–3] to work despite incorrect measurements and noise. Apart from heuristic strategies such as Random Sampling and Consensus (RANSAC) [4], globally optimal consensus maximizers [5–11] have been widely studied for rigid shapes, where there exists a simple analytical transformation between two sets of measurements. In contrast, such tools have not been explored in earnest for the model-free scenario, where simple analytical transformation models cannot explain the measurements. An important field where model-free approaches are needed is in non-rigid shape registration. Consensus maximization in non-rigid shapes have applications in augmented reality, object animations and shape analysis, among others.

While a large number of works have tackled non-rigid registration problem between images or shapes [12–16], little attention has been given to identifying outliers in matching correspondences. A few methods solve the problem in the images of non-rigid shapes [17, 18] and between a template shape and an image [19] through locally optimal approaches. The difficulty of assigning a suitable minimal parameter model to non-rigid transformations makes it highly challenging to devise a consensus maximizer.

In this paper, we propose a common framework of seeking consensus in a model-free correspondence set. Our key idea is that despite lacking a model which can explain

each instance in a matching set individually, one can consider the agreement between two or more instances using certain rules to formulate constraints. In non-rigid shapes, a rule widely applied for reconstruction and registration is the isometric deformation prior. Isometry implies that the geodesic distances are preserved despite deformations. Using these theoretical understandings, we provide our contributions in three different aspects. First we show how a model-free consensus maximization problem can be posed as a graph problem and solved as an Integer Program if we have inlier/outlier rules on the matching sets. Such an Integer Program can be solved optimally using a BnB approach. Second, we apply this formulation for removing outliers in non-rigid shape correspondences under the isometry prior. We show that our method can handle as much as 80% outlier correspondences on isometric surfaces. We provide extensive experiments on several isometric and partial shapes, as well as 'loosely' isometric partial inter-subject human shapes, where we obtain results that improve over the state-of-the-art methods. To show the generic nature of the introduced consensus maximizer, we also formulate a 3D template-to-image outlier removal problem using the piecewise rigidity and smoothness prior. We conduct extensive experiments in order to analyze the behavior of the proposed algorithms and to compare with the state-of-the-art methods.

2 Related Work

We briefly highlight the related works that are relevant to non-rigid registration problems. The first problem is that of maximizing consensus between matched 3D surface points in non-rigid 3D shapes using the isometry prior. Isometry is a widely used prior in registration [14, 15, 20–22] as well as 3D reconstruction [23, 24]. Most non-rigid shape registration methods [15, 20–22] start with a 3D descriptor such as the SHOT descriptor [25] or heat kernels [26] and establish correspondences between shapes through energy minimization. Others compute the registration by blending conformal maps [14, 27]. Any such matching method results in good and bad matches. In the following sections, we study how the outlier matches from various methods can be removed in practical cases, including complete, partial and inter-subject scenarios.

  3D template-to-image matching is yet another important problem in non-rigid shapes that can be used to localize cameras [28] or for template-based 3D reconstruction [19, 24, 29, 30]. Eliminating outlier matches in such cases is addressed in [19] by using a local iterative approach. Most other methods which solve image registration [12, 18] do not use a 3D geometric prior explicitly. We address the problem of consensus maximization in this setting with piece-wise rigidity and smoothness prior. A recent method [16] solves the combinatorial matching problem with similar constraints but does not focus on the problem of identifying outlier matches.

3 Background and Theory

Notations. We represent sets and graphs as special Latin characters, e.g., $\mathcal{V}$. We use lowercase Latin letters i, j, k or l to represent indices or sets of indices. For example, $\mathcal{V}_i$ is an element of the set $\mathcal{V}$. We write known or unknown scalars also in lowercase letters, such as z. We use uppercase bold Latin letters to represent matrices (e.g., M)

and lowercase bold Latin letters to represent vectors (e.g., v). We use lowercase Greek letter ϵ to represent thresholds. We use uppercase Greek letters to represent mappings or functions (e.g., Φ). We use $\|.\|$ to denote the ℓ_2 – norm and $|\,.\,|$ to denote the ℓ_1 – norm of a vector or the cardinality of a set. Unless stated otherwise, we write primed letters to represent quantities related to the transformed set.

3.1 Outliers

Let $\Phi : \Omega \to \Omega'$ be a transformation function between two spatial domains. Φ is related to the matching sets $\mathcal{P} = \{\mathcal{P}_i : \mathcal{P}_i \in \Omega, \ i = 1, \ldots, p\}$ and $\mathcal{P}' = \{\mathcal{P}'_i : \mathcal{P}'_i \in \Omega', \ i = 1, \ldots, p\}$. In practice, Φ may be a rigid or non-rigid transformation function or such transformations followed by camera projection. Each member $\mathcal{P}_i$ corresponds to the member $\mathcal{P}'_i$ in the second set. This defines a set of matches $\mathcal{C} \subset \mathcal{P} \times \mathcal{P}'$ that may contain outliers. The outlier set $\mathcal{O}$ is defined with a distance function Δ as:

$$\forall i \in \{1 \ldots p\}, \quad \Delta\left(\Phi(\mathcal{P}_i), \mathcal{P}'_i\right) \geq \epsilon \implies i \in \mathcal{O}. \tag{1}$$

A correspondence pair $(\mathcal{P}_i, \mathcal{P}'_i)$, also simply denoted as i, is an outlier if the distance between the mapping of $\mathcal{P}_i$ and its correspondence $\mathcal{P}'_i$, is greater than the threshold ϵ.

3.2 Consensus Maximization

Using the definition of outliers in (1), the problem of consensus maximization is defined as the minimization of the cardinality of the set $\mathcal{O}$ for the unknown Φ:

$$\begin{aligned}
\underset{\Phi}{\text{minimize}} \quad & |\mathcal{O}| \\
\text{subject to} \quad & \Delta\left(\Phi(\mathcal{P}_i), \mathcal{P}'_i\right) \geq \epsilon \implies i \in \mathcal{O}.
\end{aligned} \tag{2}$$

Problem (2) implies that we wish to find the mapping Φ which results in the least number of disagreements given by the cardinality of $\mathcal{O}$, in the given matching set $\mathcal{C}$. In rigid SfM related problems, Φ can be often expressed using a linear or non-linear function with a small fixed number of parameters. This means that equation (1) can be evaluated point-wise[1] and also estimated using a very small size of point correspondence set, known as the minimal set. There is no doubt that such problems can be efficiently solved using RANSAC and other globally optimal methods highlighted in section 1. It should be noted that even if Φ can be parameterized, very recently problem (2) was shown to be NP-hard with W[1]-complexity [31, 32], meaning that solving it optimally is very expensive. We call such a problem, when Φ can be parameterized (with a reasonably small number of parameters), as model-based consensus maximization. In the sections below we focus on the model-free case. Note that most formulations on consensus maximization are written as maximization of the inlier set cardinality rather than the minimization of the outlier set cardinality. However, these definitions are equivalent and we choose the latter for convenience.

---

[1]Although in some cases such as that of the Fundamental Matrix, Φ cannot be determined point-wise, it can be estimated for a minimal set. Thus, a RANSAC problem can be formulated.

3.3   Generic Rules-based Consensus Maximization

In contrast to model-based problems, for many applications such as those related to non-rigid shapes, Φ cannot be represented with a small size of parameters and therefore it cannot be estimated using a minimal point set. As a consequence, Δ cannot be evaluated point-wise. For example, consider the case when Φ represents the mapping between the two instances of a non-rigid surface. Such a map may be represented by Free-Form Deformation (FFD) [18, 33] or specialized latent space models such as SMPL [34] for human body, both requiring a large number of points to fit the latent parameters. In such cases problem (2) is impractical to solve in its original form.

Therefore, we offer an alternative consensus maximization formulation which is easier to solve for a special class of problems. A problem belongs to this special class if the sets $\mathcal{P}$ and $\mathcal{P}'$ have a common underlying structure which can be measured using subsets of the match set $\mathcal{C}$, without explicitly computing the transformation function Φ. To obtain a tractable formulation, we define a set of binary variables $\{z\}, z_i \in \{0, 1\}$ and $i \in \{1, \ldots, p\}$ such that $z_i = 1 \iff i \in \mathcal{O}$. Let a binary valued function $\Theta :$ $(\mathcal{S}_a, \mathcal{S}_b) \to \{0, 1\}$ measure the agreement between two small subsets $\mathcal{S}_a, \mathcal{S}_b \subset \mathcal{C}$. Θ evaluates to 1 if the subsets $\mathcal{S}_a$ and $\mathcal{S}_b$ agree up to some threshold ϵ and 0 otherwise. Then the following is an alternative of the original problem (2):

$$\begin{aligned} &\underset{\{z\}}{\text{minimize}} \quad \sum_i z_i \\ &\text{subject to} \\ &\Theta(\mathcal{S}_a, \mathcal{S}_b) = 0 \implies \exists (\mathcal{P}_i, \mathcal{P}'_i) \in \mathcal{S}_a \cup \mathcal{S}_b : z_i = 1, \\ &\forall \, (\mathcal{S}_a, \mathcal{S}_b) : \mathcal{S}_a \neq \mathcal{S}_b, \end{aligned} \quad (3)$$

The function Θ can be thought of as a rule which uses priors on the sets $\mathcal{P}$ and $\mathcal{P}'$ to measure the agreement on the matched subsets. The subsets $\mathcal{S}_a$ and $\mathcal{S}_b$ sampled from the match set $\mathcal{C}$, are the minimal sets such that Θ can be evaluated. Problem (3) simply means, in case two subsets chosen on the basis of a prior do not agree with each other, *at least one member from the union of those subsets must be an outlier*. This is the key idea of our work. Although solving problem (3) optimally does not guarantee an optimal solution for problem (2), the latter is a close relaxation of the former. Therefore solving problem (3) amounts to solving the model-free consensus maximization. Problem (3) is still a combinatorial problem and is NP-hard. In the next section we give more insights into the problem with a graph structure and provide a globally optimal method for solving it with integer programming.

4   Consensus Maximization with a Graph

We represent the union of all samples $\mathcal{S}_a$ and $\mathcal{S}_b$ as the nodes and the connection between them as the edges of a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$. The node set $\mathcal{V}$ consists of all unique sampled subsets $\mathcal{S}_a$ and $\mathcal{S}_b$. An edge $(\mathcal{S}_a, \mathcal{S}_b) \in \mathcal{E}$ connects the nodes $\mathcal{S}_a$ and $\mathcal{S}_b$ and induces the agreement function $\Theta(\mathcal{S}_a, \mathcal{S}_b)$. We use the index $k \in \{1 \ldots v\}$ to denote the nodes $\mathcal{V}$ and the index $l \in \{1 \ldots e\}$ to denote the edges $\mathcal{E}$. Figure 1 illustrates this representation of the problem.
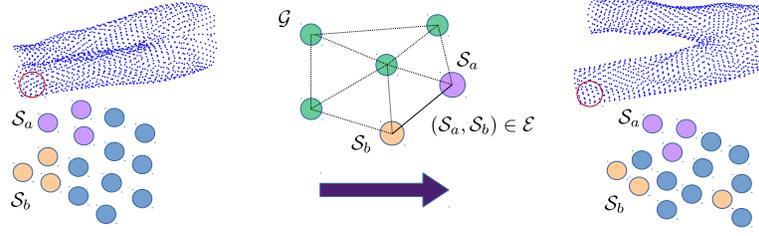
Fig. 1: **Graph formulation for consensus maximization.** The selected point sets (nodes) are drawn as orange and purple circles in the graph, connected by edges representing the compatiblity between the sets. The point clouds are taken from [35].

4.1 Graph Formulation

Given the graph $\mathcal{G}$, we would still like to compute the original binary variable set $\{z\}$. With a slight abuse of notations, we define the binary variable set of a node as $z_k \triangleq \{z_i\} : (\mathcal{P}_i, \mathcal{P}'_i) \in \mathcal{V}_k$. Similarly we define the binary variable set of an edge as $z_l \triangleq \{z_i\} : (\mathcal{P}_i, \mathcal{P}'_i) \in \mathcal{V}_{k_a} \cup \mathcal{V}_{k_b}$ for $\mathcal{E}_l = (\mathcal{V}_{k_a}, \mathcal{V}_{k_b})$. The constraint on the binary variables can then be compactly expressed as:

$$\Sigma z_l + \Theta(\mathcal{E}_l) \geq 1, \tag{4}$$

where Σz_l represents the sum of all the elements in the set z_l. Problem (3) with constraint (4) is an example of graph optimization where we need to compute the node properties z_k for each node k using the edge measurements $\Theta(\mathcal{E}_l)$.

4.2 Integer Programming

Using the constraint of (4) in a graph, we propose an efficient way to solve the consensus maximization problem, under the framework of Integer Programming, as:

$$\begin{aligned} \underset{\{z\}}{\text{minimize}} \quad & \sum z_i \\ \text{subject to} \quad & \sum z_l \geq 1, \quad \forall l \in \{1 \dots e\}, \quad \text{if } \Theta(\mathcal{E}_l) = 0. \end{aligned} \tag{5}$$

Problem (5) can be optimally solved using any off-the-shelf solver for Integer Programming. This is done using the popular BnB method. Often such problems in consensus maximization are solved using the so-called big M method [36]. Such a method is needed when a binary decision function Θ cannot be defined for a given edge $\mathcal{E}_l$. In that case, the integer inequality in problem (5) is written as $M \sum z_l + \epsilon \geq \Lambda(\mathcal{E}_l)$ using the scalar-valued function Λ and a scalar threshold ϵ. Here, M is a chosen large scalar number that makes the problem feasible when Λ is large. However, in this work we consider only those problems that can be expressed with a binary rule Θ.

Relaxed alternatives and BnB. Integer programming problems are generally non-convex in nature. They can be simplified by further relaxing the binary or integer constraint with real bounds. In contrast, we opt for the BnB approach keeping the integer constraint in order to obtain a globally optimal solution even in case of high outlier ratio. Such an approach computes the lower and upper bound of the cost iteratively and terminates with a certificate of ϵ sub-optimality if they are equal. We compare the relaxed and the globally optimal methods in section 6. In the next section, we describe two different problems in non-rigid shapes which can be expressed in the form of problem (5).

5   Non-Rigid Shapes

Non-rigid objects have deformations that cannot be parameterized with a small fixed set of parameters. Nevertheless, they do obey some shape priors. We provide our methods for two problems in non-rigid shapes below, based on such deformation priors.

5.1   Shape Matching with Isometry

We consider two different shapes $\mathcal{P}$ and $\mathcal{P}'$ related by an unknown deformation Φ. We want to establish the set of outlier points $\mathcal{O}$ on the matching set $\mathcal{C}$. Such problems may arise, for example, when registering 3D non-rigid surfaces using image matches [28] or when registering different shapes with a 3D feature point descriptor [25, 26]. In order to solve the problem, we consider the isometric deformation prior which assumes that the surface distances are preserved under deformations. The prior allows us to use the following graph attributes:

$$
\begin{aligned}
\mathcal{V}_k &= (\mathcal{P}_i, \mathcal{P}'_i) \\
\Theta(\mathcal{E}_l) &= \begin{cases} 1 & \text{if} \quad \|\Psi(\mathcal{P}_{i_1}, \mathcal{P}_{i_2}) - \Psi(\mathcal{P}'_{i_1}, \mathcal{P}'_{i_2})\| \le \epsilon \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
\tag{6}
$$

where Ψ is the function that measures the geodesic distance between two points on a surface. Each graph node consists of a single matching pair in $\mathcal{C}$. Therefore each constraint in (6) obtained for an edge consists of only two binary variables, making the problem highly sparse. Although, we only show the problem formulation using isometry, other deformation priors such as conformality may be used in problem (6).

Practical considerations. While the method works perfectly for isometric surfaces, objects which are undergoing topological changes such as a tearing piece of paper or loosely isometric surfaces such as a human body pose additional difficulty, as isometry is not always satisfied in such cases. We therefore provide a more practical approach to solve the problem in algorithm 1. In algorithm 1, separately applying the method for different clusters also addresses the non-linear time complexity of the integer programming problem. This allows us to use the method in dense point surfaces as the time complexity with the number of clusters is always linear. To estimate the geodesics, we compute a mesh by Delaunay triangulation when a mesh is not provided.

Algorithm 1 : $\{z\}$ = shapeRegistration $(\mathcal{P}, \mathcal{P}', \mathcal{C})$

---

  1. Cluster initial matches $\mathcal{C}$ into m disjoint clusters using k-means.
  2. For each cluster $c \in \{1 \ldots m\}$,
   (a) Compute nearest neighbors and establish edges $\mathcal{E}_l$.
   (b) For each edge compute the agreement function $\Theta(\mathcal{E}_l)$.
   (c) Formulate constraints (6) with Θ.
  3. Aggregate all the results from each cluster c.

5.2 Template to Image Matching

Template-based reconstruction is a well-studied problem [19, 23, 24, 30] which uses matches between the template shape $\mathcal{P}$ and the deformed shape's image $\mathcal{I}$ to reconstruct the deformed surface. Again, the matches established may consist of outliers, in which case the reconstruction obtained can be of poor quality. Here, we propose the use of piece-wise rigidity and surface smoothness as the priors to define the agreement function Θ. Despite non-rigidity, surface smoothness has been successfully used in the state-of-the-art template-based reconstruction methods [19, 30]. We use a similar approach by considering that the relative camera to object pose changes smoothly over the surface. Using these priors we define the graph attributes as follows:

$$
\begin{aligned}
\mathcal{V}_{k_1} &= \{(\mathcal{P}_i, \mathcal{I}_i)\}, \quad i = \{i_1, i_2, i_3\}, \quad i_1, i_2, i_3 \in \{1 \ldots p\} \\
\mathcal{V}_{k_2} &\in \mathcal{N}(\mathcal{V}_{k_1}) \\
\Theta(\mathcal{E}_l) &= \begin{cases} 1 & \text{if} \quad \Delta\left(\mathsf{R}_{k_1}^{\top}, \mathsf{R}_{k_2}\right) \leq \epsilon_1 \text{ and } |\mathsf{t}_{k_1} - \mathsf{t}_{k_2}| \leq \epsilon_2 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
\tag{7}
$$

where R_{k_1} and R_{k_2} represent the rotations of the absolute pose estimated using the nodes $\mathcal{V}_{k_1}$ and $\mathcal{V}_{k_2}$ respectively, for the image $\mathcal{I}$. We define $\mathcal{N}(.)$ to be the set valued function giving neighboring nodes in the graph. Similarly t_{k_1} and t_{k_2} represent the camera translations. The rule Θ measures how well the poses agree for the two nodes. To that end, Δ is the function used to measure the distance between two rotations. We use two hyperparameters ϵ_1 and ϵ_2 to threshold the change in rotation and translation respectively. Local rigidity and surface smoothness imply that the poses should also change smoothly. The absolute pose problem can be solved using any of the so-called PnP methods [37–39]. We consider only the minimal problem that uses three non-collinear matched points and is also known as the P3P method [37]. The solutions obtained with P3P have a 4-fold ambiguity. This can be disambiguated either by using an additional matching point pair or by simply choosing the solution that minimizes Δ. The nodes are sampled such that each edge requires only four unique point matches and therefore each inequality constraint will consist of four binary variables.

Practical considerations. Piecewise rigidity is a stronger prior compared to isometry. For non-rigid shapes, this holds true only for close neighbors. In contrast to the shape matching problem of 5.1, each edge here requires four point matches. For that reason, it requires the matching set to be dense enough so that the points obey rigidity at least

---

Algorithm 2 : $\{z\}$ = templateImageRegistration $(\mathcal{P}, \mathcal{I}, \mathcal{C})$

---
1. Cluster initial matches $\mathcal{C}$ into m disjoint clusters using k-means.
2. For each cluster $c \in \{1 \ldots m\}$,
 (a) Compute various triangulations of the point clusters and establish edges with two triangles.
 (b) For each such pair of triangles with shared edge, evaluate $\Theta(\mathcal{E}_l)$.
 (c) Formulate constraints (7) with Θ.
3. Aggregate all the results from m clusters.

---

locally. Algorithm 2 describes the implementation of the method. A very naive simplification of algorithm 2 can be made by considering all points that produce a high number of 1's in the agreement function Θ to be inliers. We term such a voting method as local filtering which can find obvious inliers in the template-image matching problem.

5.3   Complexity Analysis

The combinatorial complexity of problem (5) depends on four main aspects: the number of points p, the neighborhood size q, the cluster size r and the cardinality of the minimal set required to represent a vertex set $\mathcal{S}$, say s in the graph (see Fig. 1). The complexity for a single Integer Program as reported in Table 1 can be directly obtained from the combinatorics in graph. Although the template-to-image matching complexity ($s = 3$) is high, the problem demands only local agreement, which allows us to use a small local neighborhood ($q = 15$) for creating the vertices (triangles in this case) with on average 30 edges per point. This is not the case in the shape matching and we use a fully-connected graph ($q = p/r - 1$) on any cluster as the geodesic measurements are valid irrespective of the points' proximity.

Table 1: **Complexity Analysis.** Solving for n points and minimal set size s with full connectivity (cluster size $r = 1$) and q-connectivity (cluster size r).

Full-connectivity		q-connectivity	
Vertices	Edges	Vertices	Edges
$\binom{p}{s}$	$\binom{\binom{p}{s}}{2}$	$\frac{p}{sr}\binom{q}{s-1}$	$\binom{\frac{p}{sr}\binom{q}{s-1}}{2}$

6   Experimental Results

We present the results and analysis of our proposed methods in this section on several standard datasets. We refer to the integer program based methods as *exact* or the proposed method. We also compare with the simplified method where the binary constraints in problem (5) have been relaxed to real, which we refer to as the *relaxed* method. We compare and use several matching or outlier removal methods. We write the spline-warp based image outlier removal method [18] as *featds*. We write the graph matching method [12] as *maxpoolm*. We test the template-image outlier removal method based on mesh Laplacian [19] as *laplacian*. Apart from these image-based methods we also use shape matching methods. We write the recent deformable shape kernel matching method [15] as KM. We write the deep functional map [22] as DFM

and the blended intrinsic maps [14] as BIM. We implement our methods in MATLAB with YALMIP [40] and MOSEK [41] for integer programming. Below we describe in detail the experiments for each of the discussed non-rigid registration problems.

Clustering and threshold parameters. For some experiments, we apply clustering to handle the high number of point matches. For template-to-image matching and the Hand dataset, the number of point matches is low ($n < 200$) and therefore the number of clusters is 1. For the human shapes and the newspaper dataset we choose the number of clusters as 5 based on neighborhood (k-means clustering). Note that the result aggregation is straight forward, since the clusters are disjoint. For Fig. 2, to vary the number of points, we randomly sub-sampled the points to a fixed number. Regarding thresholds, we fix $\epsilon = 20\%$ distance error relative to the template for shape matching (Sec. 6.1) unless stated otherwise. In the template-to-image (Sec. 6.2) matching case, we use $\epsilon_1 = 10^o$ and $\epsilon_2 = 40\%$ for all datasets.

6.1   Non-rigid Shape Matching

We begin by analyzing the behavior of the proposed methods on synthetic data where the ground truth correspondences are available for the shape matching problem. We also compare the proposed methods with the state-of-the-art methods on several real datasets. All these are outlined below.

Mocap data. We test with two cloth-capture data [35]. The datasets consist of a cloth falling (toss) and a moving pair of trousers (stepping trousers). The datasets are generated with mocap and consist of registered real 3D points. We synthetically generate outliers by randomly re-assigning matches to evaluate our methods.

Figure 2 (a) compares the *relaxed* and *exact* versions of the proposed method. We observe that, for low outlier ratio, it is possible to remove all the outliers using the *relaxed* method. However, it breaks down as the percentage of outliers increase beyond 50%, while the *exact* solution still correctly detects the inliers even in conditions with 80% outliers. Note that the proposed method does not detect any false positive inliers. Figure 2 (b) shows how the *exact* method behaves with the number of iterations. We observe that the method quickly computes the upper bound cost or the pessimistic inlier set while it takes a while to obtain the certificate of optimality. We find this behavior to be consistent to many other experimental setups. Figure 2 (c) shows the number of open nodes at each iteration, describing how BnB evaluates and prunes branches. To investigate time complexity, we also plot the execution time for the *exact* method in figure 2 (d). It can be observed that the execution time increases with increase in the number of points. However, this is not a problem in practice thanks to the clustering framework presented in algorithm 1.

KINECT Newspaper dataset. The RGB-D data obtained from depth-camera sensors such as KINECT make an important field of application for the method. We investigated our method on the Newspaper dataset[2] [42]. It consists of a double sheet of
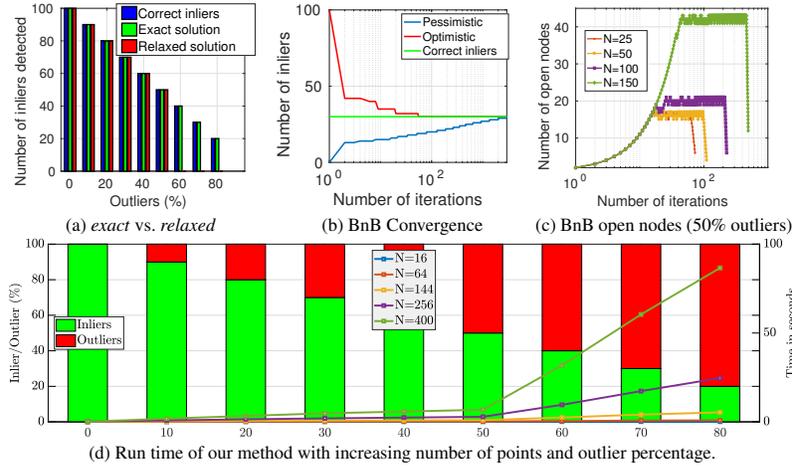
---

[2] Dataset was provided by the authors.

Fig. 2: **Analysis of our method.** Number of inliers detected, convergence of the proposed method, and time taken for the mocap cloth dataset [35] under various setups. Note that the number of iterations in (b) and (c) are in log-scale.

newspaper being torn into two parts. Figure 3 shows the inliers and outliers for a part of the template image with our method. Due to the local neighborhood computed using both point sets, the *exact* method can robustly handle the topological changes. On the other hand, the *relaxed* method does not work well from lack of enough constraints[3].

Hand dataset. The hand dataset [42] consists of two different instances of a hand and their 3D ground truth obtained with SfM. Due to the non-rigid deformation, the detected SIFT correspondences consist of very few matches with a large percentage (more than 70%) of outliers. The shape matching methods [14, 15] completely fail on this dataset and we do not show them here. We show the results of the *exact* method in figure 4 and the next best performing methods in figure 5. These qualitative results clearly show that the compared methods do not perform well in such difficult cases.

Human body shapes. In the next set of experiments, we use our methods on human body scans from the FAUST [43] dataset. To introduce challenging outliers, we consider a partial matching scenario by cutting out one arm and one leg from the mesh, and matching it to the full one. Thanks to the mesh registrations provided by the dataset, we can exactly evaluate inliers and outliers based on geodesic deviations to the ground truth correspondences (deviations greater than 15cm are considered as outliers). We compare our *relaxed* and *exact* methods against matches estimated by DFM [22], KM [15], and BIM [14]. Although BIM [14] produced visually good correspondences, it suffered from mirror-image ambiguity, that could not be resolved. Therefore we compare to BIM only where proper evaluations were possible.

---

[3]The complete set of results are provided in the supplementary document.

(a) *exact*

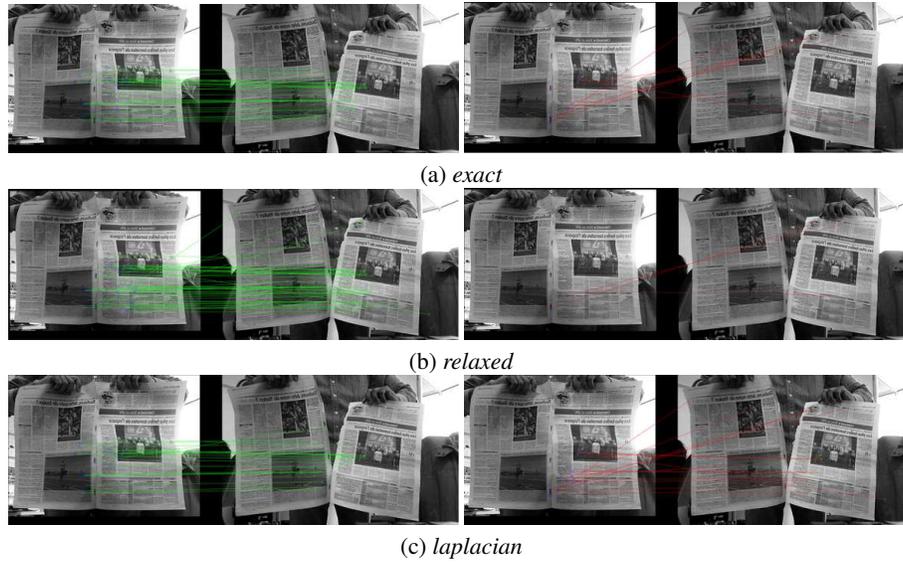(b) *relaxed*

(c) *laplacian*

Fig. 3: **Newspaper dataset.** Visualization of inlier and outlier matches from our *exact* and two next best performing methods for an example pair of the Newspaper dataset. Left column shows the inlier detection and the right column shows the outlier detection.
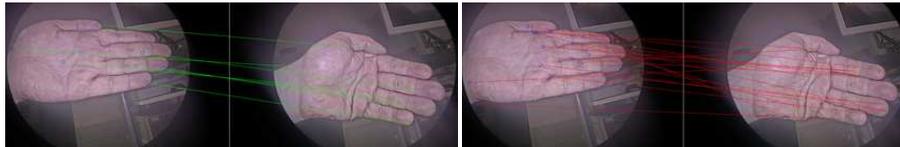

Fig. 4: **SfM Hand dataset.** Inlier detections (left) and outlier detections (right) of our *exact* method.
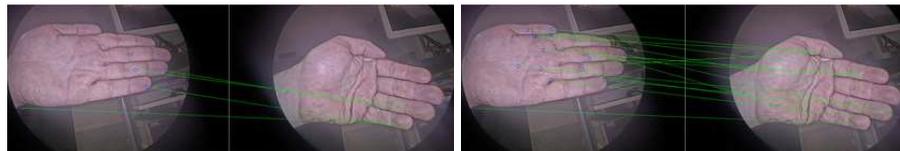

Fig. 5: Inlier detections with *laplacian* (left) and *relaxed* (right).

Since our method is designed for isometric shapes, we conduct the first experiment in the *intra-subject* case (same subject in 9 different poses). We observe that our method can successfully eliminate more than 90% outliers produced by DFM and KM while removing only a few true inliers, as shown in the first column of Table 2.

In inter-subject body shape matching applications however, the isometry assumption holds only to some extent. We use two challenging datasets to test such scenarios. The first one is on *inter-subject* matching on the FAUST data, again in the partial matching setting. Since the body shape varies across subjects, isometry doesn't hold anymore. The results presented in the second column of Table 2 demonstrate that this problem is significantly harder than the isometric matching. We see that the matches resulting from BIM contain outliers that are very hard to detect, and only 15% can be removed without sacrificing many inliers. For DFM and KM, we reliably detect more than 80% and 90% resp., and therefore improve the matching robustness for subsequent tasks.

Our third experiment with human body shapes involves dense correspondence estimation from a depth map to the 3D model. We rendered synthetic depth map mimicking the projection and noise properties of KINECT from an articulated MPII Human Shape model [44] using variations of upright poses and body shapes. To compute the geodesics on this modality, we triangulated the point cloud using 2D Delaunay triangulation. Applying DFM and KM on the raw input does not work well, since SHOT [45] and HKS [46] are not reliable features for depth maps. We therefore take initial matches from a metric regression forest [47] trained on the specific task of dense correspondence estimation. We then compare our methods, KM and ICP on top of these matches in the third column of Table 2. We can conclude that, although provided with inital matches, KM fails to correctly match the two modalities. Our method however shows promising results even though the shapes are non-isometric, and geodesics are computed on the triangulated point cloud. Interestingly, our result is comparable to that of the articulated non-rigid ICP which exploits additional information such as the kinematics and a stronger shape prior. Fig. 6 shows a qualitative example from our test set.

In summary, we showed that our method can be used on top of generic matching methods to robustly detect outliers for isometric deformations, and some classes of non-isometric registration such as inter-subject body shapes. Moreover, we can confirm our results on the synthetic data and conclude that even the *relaxed* method provides good results if the proportion of the outliers is below 50%.

6.2   Template to Image Matching

The template 3D to image matching is an important problem in non-rigid geometry. Most reconstruction methods [19, 30] are sensitive to outlying correspondences and proceed by first removing outliers in matches. We use problem (7) to formulate the template to image outlier removal method using the absolute pose. We test our results on three datasets: KINECT Paper [48], T-Shirt [49] and the MPI Sintel [50] all of which contain the groundtruth 3D data and images. We select a random single pair for each dataset and compute the SIFT matches. We count the number of inliers and outlier matches manually for each of the methods' output. We compare our methods with three other state-of-the-art methods *laplacian*, *featds* and *maxpoolm*. Similarly, as discussed in section 5.2 we also report the results of the *relaxed* method. We further report the

Method	FAUST				MPII HumanShape	
	intra-subject		*inter-subject*		*from rendered depth map*	
	Inliers / Outliers	Time [s]	Inliers / Outliers	Time [s]	Inliers / Outliers	Time [s]
BIM	-	-	3381 / 1602	3	-	-
BIM+Ours (*relaxed*)	-	-	3269 / **1362**	10	-	-
BIM+Ours (*exact*)	-	-	3267 / 1395	32.9	-	-
DFM	4211 / 772	1	3756 / 1227	1	272 / 3728	1
DFM+Ours (*relaxed*)	3918 / **31**	19	3437 / **93**	15	-	-
DFM+Ours (*exact*)	3918 / **31**	24	3437 / **93**	19.4	-	-
KM	4736 / 181	89	4051 / 860	92	572 / 3387	53
KM+Ours (*relaxed*)	4554 / 18	104	3634 / **161**	107	-	-
KM+Ours (*exact*)	4556 / **17**	110	3634 / **161**	115	-	-
RF	-	-	-	-	3220 / 780	<1
RF+KM	-	-	-	-	1162 / 269	3
RF+Ours (*relaxed*)	-	-	-	-	2800 / **137**	14
RF+Ours (*exact*)	-	-	-	-	2800 / **137**	15
RF+ICP	-	-	-	-	3166 / 159	301
	mostly isometric		non-isometric			

Table 2: **Non-rigid 3D shape matching.** Results on FAUST [43] intra- and inter-subject, as well as matching depth maps to the MPII HumanShape [44] model. We report the number of true positive (inliers) and false positive (remaining outliers) matches.

Fig. 6: **Qualitative results.** Non-isometric shape matching from depth map. Left to right: body mesh model [44], RF [47], RF+KM [15], RF+Ours, RF+ICP, input depth map. Correspondences are color-coded, gray indicates removed matches.

Method	Kinect Paper		T-shirt		Sintel	
	Inliers	Time(s)	Inliers	Time(s)	Inliers	Time(s)
Local-filtering	46 / 142	4.22	95 / 351	6.10	17 / 68	2.03
relaxed	99 / 142	5.56	291 / 351	7.52	44 / 68	3.51
exact	114 / 142	7.59	**309** / 351	9.66	**53** / 68	5.01
laplacian	**126** / 142	**1.15**	301 / 351	7.84	44 / 68	0.53
featds	76 / 142	3.93	304 / 351	**1.46**	42 / 68	**0.32**
maxpoolm	3 / 142	159.96	6 / 351	608.55	16 / 68	7.88

Table 3: **3D template to image matching.** Comparison on three different real datasets.

results of the local-filtering method as another baseline where the inliers are decided based on the local neighborhood voting.

We test all the methods with favorable parameters. The reported inliers are manually validated. The results show that our method performs in par with *laplacian* designed exactly for the template-based outlier removal. Note that the *exact* method consistently detects more number of inliers than other methods. Our method performs better than *featds* in multi-body situation as *featds* uses a single spline-based warp and computes the residuals to identify outliers. We visualize the results of outlier removal in figure 7 for the proposed method and two other best performing methods: *featds* and *laplacian*.

Fig. 7: Inliers (left) vs. Outliers (right) for the T-shirt dataset using the *exact* method. The performance of our method is better than that of the two compared methods designed for non-rigid matching. More results are provided in the supplementary material.

7   Conclusions and Future Work

In this paper we brought forward a theory on model-free consensus maximization using integer programming and an optimal method to solve it using Branch and Bound. We formulated two different registration problems using our consensus maximizer: isometric shape outlier removal and template-image outlier removal. We obtained very good results at up to 80% mismatches in non-rigid shape registration and >25% mismatches in template-image registration. We obtained these results by solving a close relaxation of the original problem with guaranteed optimality. We showed with extensive experiments that our methods consistently performs on par or better than the existing methods.

Although the focus of this paper was on non-rigid shapes, many vision problems can be converted to formulation 5 with three or less variables per graph node. A non-exhaustive list includes: *i)* one variable problems: relative pose on robot navigation [51], *ii)* two variable problems: robust triangulation [52] and pure translation estimation [53], and *iii)* three variable problems: image to image affine homography and three-view modulus constraints [54]. For future works, we intend to tackle some of these problems using the formulation we developed in this paper.

Acknowledgements

References

1. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Second edn. Cambridge University Press, ISBN: 0521540518 (2004)
2. Longuet-Higgins, H.: A computer algorithm for reconstructing a scene from two projections. Nature **293** (1981) 133–135
3. Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE Trans. Pattern Anal. Mach. Intell. **26**(6) (2004) 756–777
4. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6) (1981) 381–395
5. Chin, T.J., Kee, Y.H., Eriksson, A., Neumann, F.: Guaranteed outlier removal with mixed integer linear programs. In: CVPR. (2016)
6. Speciale, P., Paudel, D.P., Oswald, M.R., Kroeger, T., Gool, L.V., Pollefeys, M.: Consensus maximization with linear matrix inequality constraints. In: CVPR. (2017)
7. Bazin, J.C., Li, H., Kweon, I.S., Demonceaux, C., Vasseur, P., Ikeuchi, K.: A branch-and-bound approach to correspondence and grouping problems. IEEE transactions on pattern analysis and machine intelligence **35**(7) (2013) 1565–1576
8. Hartley, R.I., Kahl, F.: Global optimization through rotation space search. IJCV **82**(1) (2009) 64–79
9. Bazin, J.C., Seo, Y., Hartley, R.I., Pollefeys, M.: Globally optimal inlier set maximization with unknown rotation and focal length. In: ECCV. (2014)
10. Li, H.: Consensus set maximization with guaranteed global optimality for robust geometry estimation. In: ICCV. (2009)
11. Zheng, Y., Sugimoto, S., Okutomi, M.: Deterministically maximizing feasible subsystem for robust model fitting with unit norm constraint. In: CVPR. (2011)
12. Cho, M., Sun, J., Duchenne, O., Ponce, J.: Finding matches in a haystack: A max-pooling strategy for graph matching in the presence of outliers. In: CVPR. (2013)
13. Collins, T., Mesejo, P., Bartoli, A.: An analysis of errors in graph-based keypoint matching and proposed solutions. In: ECCV. (2014)
14. Kim, V.G., Lipman, Y., Funkhouser, T.: Blended intrinsic maps. In: ACM Transactions on Graphics (TOG). Volume 30. (2011) 79
15. Lähner, Z., Vestner, M., Boyarski, A., Litany, O., Slossberg, R., Remez, T., Rodolà, E., Bronstein, A.M., Bronstein, M.M., Kimmel, R., Cremers, D.: Efficient deformable shape correspondence via kernel matching. In: 3DV. (2017)
16. Bernard, F., Schmidt, F.R., Thunberg, J., Cremers, D.: A combinatorial solution to non-rigid 3d shape-to-image matching. In: CVPR. (2017)
17. Pilet, J., Lepetit, V., Fua, P.: Fast non-rigid surface detection, registration and realistic augmentation. International Journal of Computer Vision **76**(2) (2008) 109–122
18. Pizarro, D., Bartoli, A.: Feature-based deformable surface detection with self-occlusion reasoning. International Journal of Computer Vision **97**(1) (2012) 54–70
19. Ngo, T.D., Östlund, J.O., Fua, P.: Template-based monocular 3D shape recovery using laplacian meshes. IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(1) (2016) 172–187
20. Aflalo, Y., Dubrovina, A., Kimmel, R.: Spectral generalized multi-dimensional scaling. International Journal of Computer Vision **118**(3) (2016) 380–392
21. Vestner, M., Litman, R., Rodol, E., Bronstein, A., Cremers, D.: Product manifold filter: Non-rigid shape correspondence via kernel density estimation in the product space. In: CVPR. (2017)

22. Litany, O., Remez, T., Rodola, E., Bronstein, A.M., Bronstein, M.M.: Deep functional maps: Structured prediction for dense shape correspondence. In: ICCV. (2017)
23. Salzmann, M., Fua, P.: Linear local models for monocular reconstruction of deformable surfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(5) (2011) 931–944
24. Bartoli, A., Gérard, Y., Chadebecq, F., Collins, T., Pizarro, D.: Shape-from-template. IEEE Trans. Pattern Anal. Mach. Intell. **37**(10) (2015) 2099–2118
25. Salti, S., Tombari, F., Di Stefano, L.: Shot: Unique signatures of histograms for surface and texture description. Computer Vision and Image Understanding **125** (2014) 251–264
26. Ovsjanikov, M., Mérigot, Q., Mémoli, F., Guibas, L.: One point isometric matching with the heat kernel. In: Computer Graphics Forum. (2010)
27. Le, H., Chin, T.J., Suter, D.: Conformal surface alignment with optimal mobius search. In: CVPR. (2016)
28. Innmann, M., Zollhöfer, M., Nießner, M., Theobalt, C., Stamminger, M.: Volumedeform: Real-time volumetric non-rigid reconstruction. In: ECCV. (2016)
29. Wandt, B., Ackermann, H., Rosenhahn, B.: 3d reconstruction of human motion from monocular image sequences. IEEE transactions on pattern analysis and machine intelligence **38**(8) (2016) 1505–1516
30. Chhatkuli, A., Pizarro, D., Bartoli, A., Collins, T.: A stable analytical framework for isometric shape-from-template by surface integration. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(5) (2017) 833–850
31. Chin, T.J., Suter, D.: The maximum consensus problem: recent algorithmic advances. Volume 7. Morgan & Claypool Publishers (2017)
32. Chin, T.J., Cai, Z., Neumann, F.: Robust fitting in computer vision: Easy or hard? arXiv preprint arXiv:1802.06464 (2018)
33. Brunet, F., Bartoli, A., Hartley, R.: Monocular template-based 3D surface reconstruction: Convex inextensible and nonconvex isometric methods. Computer Vision and Image Understanding **125** (2014) 138–154
34. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: SMPL: A skinned multi-person linear model. ACM Trans. Graphics (Proc. SIGGRAPH Asia) **34**(6) (2015) 248:1–248:16
35. White, R., Crane, K., Forsyth, D.: Capturing and animating occluded cloth. In: SIGGRAPH. (2007)
36. McCormick, G.P.: Computability of global solutions to factorable nonconvex programs: Part iconvex underestimating problems. Mathematical programming **10**(1) (1976) 147–175
37. Kneip, L., Li, H., Seo, Y.: Upnp: An optimal o(n) solution to the absolute pose problem with universal applicability. In Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., eds.: Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I. (2014)
38. Lepetit, V., Moreno-Noguer, F., Fua, P.: Epnp: An accurate o(n) solution to the pnp problem. Int. J. Comput. Vision **81**(2) (2009) 155–166
39. Urban, S., Leitloff, J., Hinz, S.: Mlpnp - a real-time maximum likelihood solution to the perspective-n-point problem. In: ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences. Volume 3. (2016) 131–138
40. Löfberg, J.: Yalmip : A toolbox for modeling and optimization in MATLAB. In: Proceedings of the CACSD Conference. (2004)
41. ApS, M.: The MOSEK optimization toolbox for MATLAB manual. Version 7.1 (Revision 28). (2015)
42. Chhatkuli, A., Pizarro, D., Collins, T., Bartoli, A.: Inextensible non-rigid shape-from-motion by second-order cone programming. In: CVPR. (2016)

43. Bogo, F., Romero, J., Loper, M., Black, M.J.: FAUST: Dataset and evaluation for 3D mesh registration. In: Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Piscataway, NJ, USA, IEEE (2014)
44. Pishchulin, L., Wuhrer, S., Helten, T., Theobalt, C., Schiele, B.: Building statistical shape spaces for 3d human modeling. Pattern Recognition (2017)
45. Tombari, F., Salti, S., Di Stefano, L.: Unique signatures of histograms for local surface description. In: Proc. ECCV. Volume 6313. (09 2010) 356–369
46. Sun, J., Ovsjanikov, M., Guibas, L.: A concise and provably informative multi-scale signature based on heat diffusion. In: Comput. Graph. Forum. Volume 28. (07 2009) 1383–1392
47. Pons-Moll, G., Taylor, J., Shotton, J., Hertzmann, A., Fitzgibbon, A.: Metric regression forests for correspondence estimation. International Journal of Computer Vision (2015) 1–13
48. Varol, A., Salzmann, M., Fua, P., Urtasun, R.: A constrained latent variable model. In: CVPR. (2012)
49. Chhatkuli, A., Pizarro, D., Bartoli, A.: Non-rigid shape-from-motion for isometric surfaces using infinitesimal planarity. In: BMVC. (2014)
50. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In: ECCV. (2012)
51. Scaramuzza, D.: 1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints. Int. J. Comp. Vision **95**(1) (2011) 74–85
52. Li, H.: A practical algorithm for L_∞ triangulation with outliers. In: CVPR. (2007)
53. Fredriksson, J., Enqvist, O., Kahl, F.: Fast and reliable two-view translation estimation. In: CVPR. (2014)
54. Pollefeys, M., Gool, L.V.: Stratified self-calibration with the modulus constraint. IEEE Trans. Pattern Anal. Mach. Intell. **21**(8) (1999) 707–724