Robust Region Detection via Consensus Segmentation of Deformable Shapes

E. Rodolà[1], S. Rota Bulò[2], and D. Cremers[1]

[1]Technische Universität München, Garching, Germany
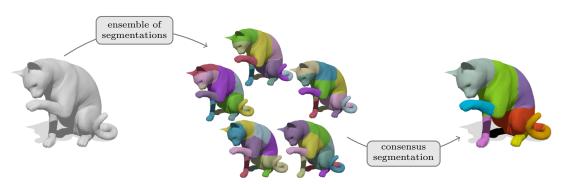[2]Fondazione Bruno Kessler, Trento, Italy

Figure 1: *The proposed framework allows to obtain robust surface segmentations in a principled way. Here, an initial ensemble of 50 segmentations (we show only 5 for visualization purposes) is generated via a clustering process on the Global Point Signature embedding of the shape. Given this ensemble, the corresponding consensus is defined as the unknown segmentation that is as close as possible to all the others. Note that the number of regions in the ensemble and in the final consensus segmentation are not necessarily the same. The detected regions are stable across non-rigid deformations of the shape.*

Abstract

We consider the problem of stable region detection and segmentation of deformable shapes. We pursue this goal by determining a consensus segmentation from a heterogeneous ensemble of putative segmentations, which are generated by a clustering process on an intrinsic embedding of the shape. The intuition is that the consensus segmentation, which relies on aggregate statistics gathered from the segmentations in the ensemble, can reveal components in the shape that are more stable to deformations than the single baseline segmentations. Compared to the existing approaches, our solution exhibits higher robustness and repeatability throughout a wide spectrum of non-rigid transformations. It is computationally efficient, naturally extendible to point clouds, and remains semantically stable even across different object classes. A quantitative evaluation on standard datasets confirms the potentiality of our method as a valid tool for deformable shape analysis.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Shape Analysis

1. Introduction

Detecting and matching features among three-dimensional shapes is at the heart of a broad range of visual tasks in computer vision and graphics, including object recognition, surface reconstruction, and pose transfer to name just a few. Inspired by their increasing success in the 2D domain [MTS*05], in recent years the interest has shifted towards the definition of region (or component) detectors as robust alternatives to the widespread point-based methods.

Due to the increased robustness with respect to their point-wise counterparts, region-based features are being successfully employed for problems of shape matching [OBCS*12, PBB*13] and shape retrieval [LBB12]. In this context, of particular relevance is the robustness or invariance of the detector under non-rigid transformations of the surface being analysed. Previous attempts to solve the problem with the language of diffusion geometry had alternating success due to their limited resilience to strong deformations.

In this work, we present a novel perspective on this problem. Our approach draws inspiration from the family of so-called ensemble methods, originating in the context of data clustering. Before providing a summary of our contribution in Sect. 1.2, we present hereafter a brief review of the most relevant works in the literature.

1.1. Related work

Perhaps the most closely related work to ours is the approach taken by Golovinskiy and Funkhouser in [GF08]. Their main idea is to generate a large set of randomized *cuts* on the mesh structure of the shape, with concavity weights assigned to each edge based on their exterior dihedral angles. The generated cuts are then scored according to a measure of overlap among them, so as to produce a global *partition function* for the mesh. The partition function provides a measure of where natural part boundaries occur in the shape. In practice, strong cuts mainly appear at the boundaries between large, nearly-convex parts of the surface. With this approach, segmentations are only obtained as a by-product of consistent cuts via a hierarchical extension to the main algorithm [GF08]. The authors reported overall average results, in line with other *rigid* segmentation approaches.

More recently, Litman *et al.* [LBB11] explicitly formulated region detection as a problem of seeking *maximally stable* components on deformable 3D models, according to a particular stability criterion based on surface area. Their approach is based on the definition of a *component tree* over the mesh graph of the shape, which is then traversed efficiently by an optimization procedure. Differently from [GF08], the framework of [LBB11] accounts for both vertex- and edge-weighted representations of the mesh graph. Most importantly, the weighting functions employed by Litman *et al.* capture the *intrinsic* properties of the underlying manifold, being constructed from diffusion geometric quantities. This choice makes the approach invariant under non-rigid transformations of the shape, and gives successful results on standard benchmarks. The approach was further extended to deal with shapes represented as volumes [LBB12], comparing favourably with the mesh-based counterpart.

Inspired by the cognitive theory of saliency of visual parts, Sipiran and Bustos [SB13] introduced a region detector based on the idea of shape "protrusions". The method differs from the other approaches in that it provides an *incomplete* decomposition of the shape, *i.e.* the union of the detected regions does not cover the whole surface. Visually, these regions correspond to the rigid parts of the shape (*e.g.* the hands, feet and head of a human shape). Other notable approaches to deformable shape segmentation include the persistence-based technique of Skraba *et al.* [SOCG10]. The method is based on the ideas of topological persistence, according to which the "relevant" (*i.e.* stable) segments of a shape correspond to the basins of attraction of some function f defined over it. First, connected components on the surface are associated with local maxima of f;

a merging procedure then traverses the hierarchy of components to produce a final stable segmentation of the shape. The method is provably stable under isometric transformations, but its performance rapidly decreases in presence of topological noise. Heat walk [BPVR11] is based on similar principles, but employs a diffusion geometric edge weighting akin to [LBB11]. Other methods based on ideas from diffusion geometry [Reu10, GBAL09] successfully tackle the nearly-isometric case, but are generally sensitive to shape deformations that are far from being isometric. Aubry *et al.* [ASC11] formulated the segmentation problem as one of Gaussian mixture clustering in a proper descriptor space. The method shows promising results, but relies on the assumption of having several poses of the same shape available, in order to produce stable regions on previously unseen transformations. To our knowledge, the methods of Litman *et al.* [LBB11, LBB12] provide the best results overall, on a broad range of shape transformations.

1.2. Contributions

In this paper, we develop the idea of detecting robust regions in three-dimensional shapes by looking for a consensus among a set (*a.k.a.* ensemble) of candidate segmentations of the shape itself. Our motivation is that consensus segmentations tend to reveal more stable region structures than the individual putative segmentations forming the ensemble. Indeed, this property of ensemble methods is common knowledge within the data clustering community [SG02, FJ05]. Works like [AG12] successfully exploited these ideas for image segmentation, although with the aim to improve the segmentation quality, rather than detecting robust segments. To our knowledge, this is the first work trying to detect repeatable regions in shapes as the outcome of an ensemble clustering process. First, we create an ensemble of isometry-invariant segmentations by running multiple times, and with different parametrizations, a simple clustering algorithm (*i.e.* k-means) in an appropriate space (Sect. 2). The ensemble is then fed to a consensus clustering algorithm specifically designed for the task (Sect. 3). Compared to existing approaches, our method has the following key advantages:

- The provided framework is *general*, in that it does not fix any specific representation of the shape. The algorithm revolves around the general notion of shape "node", which can then be instantiated to the specific problem at hand (*e.g.* mesh faces when working with meshes).
- The algorithm is *scalable*. We propose an effective way to rephrase the node-based formulation in terms of more abstract objects that we call "base segments". Further, we provide theoretical results relating the two formulations.
- Quantitative evaluations on two standard datasets confirm the improved *stability* of our approach in the presence of strong non-rigid transformations (Sect. 4).
- Finally, we show that the computational cost of finding a repeatable segmentation is low, and that this can be done with limited data (*i.e.* with fairly small ensembles).

2. Isometry-invariant segmentation of a shape

In this section we introduce the key concepts and results that set the stage for consensus segmentation as a robust tool for deformable shape analysis. These definitions will be the starting point for our method in Sect. 3.

2.1. Diffusion geometry

We model shapes as compact connected Riemannian manifolds $\mathcal{M}$ endowed with the standard measure μ induced by the volume form. The space $(\mathcal{M}, \mu)$ features the symmetric Laplace-Beltrami operator (or Laplacian) $\Delta_{\mathcal{M}}$ acting on smooth functions over $\mathcal{M}$. We call function ϕ an *eigenfunction* with eigenvalue λ if $\Delta_{\mathcal{M}} \phi = \lambda \phi$. The set of all eigenvalues $\{\lambda_1 < \lambda_2 < \dots\}$ of $\Delta_{\mathcal{M}}$ is an infinite discrete subset of $\mathbb{R}_+$ with $\lambda_k \to \infty$ as $k \to \infty$ [Ber03]. Assuming the manifold $\mathcal{M}$ has no boundary, we have an extra eigenvalue $\lambda_0 = 0$ with multiplicity 1 (by connectedness), whose corresponding eigenfunction is constant. Consider the diffusion process on $\mathcal{M}$ described by the heat equation

$$\Delta_{\mathcal{M}} u = \frac{\partial}{\partial t} u, \qquad (1)$$

with the initial condition $u(t = 0, p)$ describing the initial heat distribution at $p \in \mathcal{M}$. Any Riemannian manifold $\mathcal{M}$ possesses a unique, smooth *heat kernel* $h_t(p,q)$ which is the fundamental solution to the heat equation. For compact $\mathcal{M}$, the heat kernel can be expressed as the convergent series

$$h_t(p,q) = \sum_{k=0}^{\infty} e^{-\lambda_k t} \phi_k(p) \phi_k(q). \qquad (2)$$

The heat kernel can be given an interpretation as the transition probability density of a random walk of length t from point p to point q on $\mathcal{M}$. Following from the invariance of $\Delta_{\mathcal{M}}$ under isometric transformations of $\mathcal{M}$, the kernel $h_t(p,q)$ is an intrinsic quantity of the shape.

2.2. Global Point Signatures

Given a point $p \in \mathcal{M}$, the *Global Point Signature* (GPS) [Rus07] at p is defined as the infinite-dimensional vector

$$g(p) = \left(\frac{1}{\sqrt{\lambda_1}} \phi_1(p), \frac{1}{\sqrt{\lambda_2}} \phi_2(p), \frac{1}{\sqrt{\lambda_3}} \phi_3(p), \dots \right),$$

which corresponds to the vector of (rescaled) coefficients of an indicator function around $p \in \mathcal{M}$, when represented in the basis $\{\phi_k\}_{k=1}^{\infty}$. The GPS shares with the heat kernel the property of being an *intrinsic* quantity, *i.e.* two isometric surfaces have the same image under g (up to orthogonal transformations of the ϕ_k).

The above definitions give rise to a meaningful notion of inner product among GPS signatures, which we will now elucidate. Consider the Green's function of a general weighted manifold $(\mathcal{M}, \mu)$, which is defined by

$$G(p,q) = \int_0^{\infty} h_t(p,q) dt. \qquad (3)$$

From the relation above, it follows that estimates of the Green's function can be obtained from estimates of the heat kernel. We get the closed form expression

$$G(p,q) = \sum_{k=1}^{\infty} \frac{1}{\lambda_k} \phi_k(p) \phi_k(q), \qquad (4)$$

The expression above is sometimes referred to as the *commute-time kernel* for $\mathcal{M}$. In analogy with the heat kernel, it can be interpreted as the transition probability density of a random walk of *any* length over $\mathcal{M}$. In addition, the value of $G(p,q)$ is now a *scale-invariant* quantity, *i.e.* it does not depend on the global scale of the shape. Note that, due to the scale-space properties of the basis $\{\phi_k\}$ [Ber03], in practice one can truncate the series (4) to the first $\bar{k}$ terms (a typical value is $\bar{k} = 100$). The key observation in this section is that Eq. (4) corresponds to the *dot product* of two (infinite-dimensional) GPS signatures at points $p, q \in \mathcal{M}$ [Rus07]. This brings us to the following simple remark:

Remark 1 The GPS embedding of a shape $\mathcal{M}$ is endowed with the inner product defined by $\langle g(p), g(q) \rangle = G(p,q)$.

Recall that, for any compact Riemannian manifold, the Green's function $G(p,q)$ is such that $u(p) = \int_{\mathcal{M}} G(p,q) \eta(q) dq$ solves the Poisson equation $\Delta u = \eta$, where η is a prescribed function on the manifold. Poisson's equation arises in numerous application areas, including mesh editing [YZX*04] and surface reconstruction just to name a few. Common to these methods is the requirement to satisfy linear modelling constraints on the surface, while preserving differential properties of the original geometry in the least squares sense. Intuitively, $G(p,q)$ then measures the degree of "coupling" of two points p and q on a shape $\mathcal{M}$ undergoing geometric deformations.

2.3. An ensemble of segmentations

The intrinsic properties of GPS make it a natural candidate to produce robust segmentations of a given shape. In this context, the notion of *robustness* corresponds to the stability (or repeatability) of the segmentation across non-rigid deformations of the manifold.

A straightforward approach towards isometry-invariant segmentation is to run a clustering process on the GPS embedding of $\mathcal{M}$. Consider the partition function $C : g(p) \mapsto i$ with $p \in \mathcal{M}$, assigning to each point in $\text{Im}(g)$ its "cluster identity" $i \in \{1, \dots, k\}$. The goal of clustering is to find a good partition based on the sample $g(p_1), \dots, g(p_n)$ of n points, distributed independently according to an unknown probability measure. The quality of C can be quantified via the scatter measure

$$w(C) = \frac{1}{n} \sum_{\ell=1}^{k} \sum_{C(g(p_i))=\ell} \|g(p_i) - c_\ell\|^2 da(p_i), \qquad (5)$$

where $da(p_i)$ denotes the discrete area element of point

$p_i \in \mathcal{M}$, c_ℓ is the mean of the ℓth cluster based on the assignment C, and $\|\cdot\|$ denotes the standard Euclidean metric in $\mathbb{R}^\infty$ as induced by the inner product (4).

Definition 1 A *segmentation* of $\mathcal{M}$ is a minimizer of the within-point scatter (5).

The k-means clustering algorithm is an alternating procedure for minimizing (5). It starts from an initial seeding $\{c_\ell\}_{\ell=1}^k$ of cluster centres, followed by the assignment of each $g(p_i)$ to the nearest centre c_ℓ. Centres are then recomputed, and the procedure is repeated. The algorithm can get trapped in local minima, and its stability is characterized by the uniqueness of the minimum of the objective function with respect to the true (unknown) distribution. Indeed, there is no guarantee of existence of a unique minimizer for (5), and in fact the number of minimizers could be infinite.

Example. We show an example of this procedure in Fig. 1 (middle). We segmented a cat shape starting from different initial seeds, chosen randomly. The instability of the clustering process is revealed by the different segments we get after each minimization. Note, however, the presence of *consistent* regions across the different segmentations (notice, for instance, the raised leg).

One crucial criterion in evaluating the quality of a clustering scheme is its consistency under transformations of the data set. Even assuming an optimal seeding scheme is available, significant deviations from isometry may still affect the quality and stability of this direct approach across different deformations of the shape (recall that our samples are drawn from mappings of the shape points under g, which is an *isometry* invariant). Our key insight is that regions that are consistent across different segmentations of the shape (*e.g.* from different seeds) *remain consistent under non-isometric deformations of the manifold*. In the next section we show how to formalize this intuition in order to elicit the stable components of a given shape. We conclude this section with the following useful observation:

Remark 2 Every baseline segmentation of $\mathcal{M}$ is *scale-invariant*. In other words, starting from a fixed seed $\{c_\ell\}_{\ell=1}^k$ with $c_\ell \in \mathcal{M}$, it holds $C(g(\mathcal{M})) = C(g(\alpha\mathcal{M}))$ for $\alpha \in \mathbb{R}$.

The remark is a direct consequence of Eq. (4) and Def. 1.

3. Consensus segmentation of a shape

In this section we focus on the problem of deriving a *consensus* segmentation from a heterogeneous ensemble of segmentations of a shape. Our intuition is that the consensus segmentation, which relies on aggregate statistics gathered from the segmentations in the ensemble, can reveal segments (*a.k.a.* components) in the shape that are more stable to deformations than the single baseline segmentations. Formally, our approach can be categorized as an unsupervised learning method; the theory behind it is inspired by recent work on consensus clustering [LRBR*13b, LRBR*13a].

3.1. Preliminaries

A shape $\mathcal{N} = \{q_i\}_{i=1}^n$ consists of a set of n nodes, which can be regarded as faces (*e.g.* triangles) in case of meshes or points in case of point clouds. An *ensemble* of m segmentations over a given shape $\mathcal{N}$ can be represented as a matrix $E \in [n]^{m \times n}$, where we denote the set $\{1, \dots, n\}$ with $[n]$ for the sake of convenience. Each entry of E denoted by E_{ui} represents the index of the segment in the uth segmentation that contains shape node $q_i \in \mathcal{N}$. Rows of E are referred to as *segmentations* and the uth row is denoted by $E_{u*} \in [n]^n$, whereas columns of E are referred to as *signatures* and the ith signature is denoted by $E_{*i} \in [n]^m$.

Definition 2 Given an ensemble of segmentations E and a maximum number of segments $0 < k \le n$, a *consensus segmentation* is a segmentation $\mathbf{x}^* \in [k]^n$ that is as close as possible to all segmentations in E.

The consensus segmentation can be regarded as the *Fréchet sample mean* with samples in E defined with respect to the semi-metric space $([n]^n, d)$, where $[n]^n$ is the set of segmentations and $d : [n]^n \times [n]^n \to \mathbb{R}_+$ is a semi-metric:

$$\mathbf{x}^* \in \arg\min_{\mathbf{x} \in [k]^n} \sum_{u=1}^m d^2(E_{u*}, \mathbf{x}) . \qquad (6)$$

Since two segmentations are the same under a permutation of the segments' label, we have that $d(\mathbf{x}, \mathbf{y}) = 0$ holds also for some $\mathbf{y} \ne \mathbf{x}$. This is the reason why we assume d to be a *semi*-metric.

Measuring the distance between segmentations is in general troublesome, for segmentations might have different number of segments and the correspondences between segments are not known. An elegant way of sidestepping these problems consists in defining the distance in terms of second-order quantities. Specifically, in this work we define d in a way to penalize when two shape nodes are not consistently segmented in the two segmentations, *i.e.* when they belong to the same segment in one segmentation but to different ones in the other. Specifically, we adopt

$$d^2(\mathbf{x}, \mathbf{z}) = \sum_{i,j \in [n]} w_{ij} (\mathbb{1}_{x_i=x_j} - \mathbb{1}_{z_i=z_j})^2 , \qquad (7)$$

where $\mathbb{1}_P$ is an indicator function returning 1 if property P is true and 0 otherwise. The w_{ij}'s are non-negative weights that allow to finely control the incidence of inconsistent segment assignments for specific pairs of shape nodes. They are set to a non-zero constant if no prior preferences are available. In the case that triangles are used as shape nodes, the weight of a pair of nodes can be set, *e.g.* to the product of the areas of the respective triangles.

Remark 3 The consensus segmentation depends on a parameter k, which should be regarded as an upper bound to the actual number of segments. In fact, the number of distinct segment indices appearing in the solution to (6) might be less than k. Similarly to other approaches [SOCG10, LBB11], determining the optimal number of components to be extracted
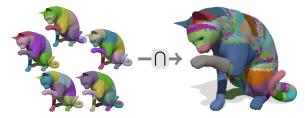
Figure 2: *Example of base segments for the cat shape, obtained from the intersection of* 50 *segmentations in the ensemble (only* 5 *are shown for visualization purposes).*

is not our focus in this work, but rather we show that stable components can be obtained with our approach under different choices of k.

3.2. Consensus with base segments

The distance measure (7) has an inherent quadratic complexity, which prevents its direct application to large shapes. In the context of data clustering, to overcome this scalability issue, [LRBR*13b, LRBR*13a] have proposed to artificially sparsify the weights w_{ij}'s, by randomly setting most of them to zero. By doing so, finding the consensus solution to (6) becomes scalable at the price of optimizing a lower bound of the original objective. Nevertheless, experiments conducted in the aforementioned works show that the quality of the solution is substantially preserved.

In this work, we pursue a different idea to overcome the scalability issue related to the computation of the consensus segmentation in (6). Our solution consists in reducing the number of elements to be segmented, yet preserving a principled relation to the original formulation. Before moving into details, we give some more definitions. Given a segmentation ensemble E, consider the following equivalence relation Ξ defined on the set [n] of indices to shape nodes, which considers two nodes equivalent if they share the same signature: $\Xi = \{(i,j) \in [n]^2 : E_{*i} = E_{*j}\}$.

Definition 3 Let $\mathcal{B} = [n]/\Xi = \{\mathcal{B}_h\}_{h\in[\hat{n}]}$ be the quotient set of [n], having $\hat{n}$ equivalence classes. A *base segment* is an equivalence class in $\mathcal{B}$.

Intuitively, $\mathcal{B}$ represents the intersection of all segmentations in the ensemble, as depicted in Fig. 2.

Remark 4 The number of base segments is in general considerably lower than the number of nodes in the shape and, by construction, they preserve the segment co-occurrence statistics of the shape nodes.

We will consider the base segments as the atomic elements to be segmented in the consensus segmentation, instead of the shape nodes. First, we derive from E a new segmentation ensemble $F \in [n]^{m \times \hat{n}}$ involving base segments as follows: for all $u \in [m]$ and $h \in [\hat{n}]$, we select an arbitrary element $i \in \mathcal{B}_h$ and set $F_{uh} = E_{ui}$. Note that the selection of i here has no importance, since E_{ui} is constant by construction for all $i \in \mathcal{B}_h$. Then, we replace the segmentation variable

$\mathbf{x} \in [k]^n$ involving shape nodes with a segmentation variable $\mathbf{y} \in [k]^{\hat{n}}$ involving base segments. Finally, we obtain the new formulation of the consensus segmentation in terms of base segments as

$$\mathbf{y}^* \in \arg\min_{\mathbf{y}\in[k]^{\hat{n}}} \sum_{u\in[m]} \hat{d}^2(F_{u*}, \mathbf{y}), \qquad (8)$$

where $\hat{d} : [\hat{n}]^{\hat{n}} \times [\hat{n}]^{\hat{n}} \to \mathbb{R}_+$ is defined as

$$\hat{d}^2(\mathbf{y}, \mathbf{z}) = \sum_{h,\ell\in[\hat{n}]} \hat{w}_{h\ell}\left(\mathbb{1}_{y_h=y_\ell} - \mathbb{1}_{z_h=z_\ell}\right)^2,$$

and $\hat{w}_{h\ell} = \sum_{i\in\mathcal{B}_h}\sum_{j\in\mathcal{B}_\ell} w_{ij}$.

The following theorem establishes some important relations between (8) and (6). It shows that the optimal objective value of (8) is an upper bound to the one of (6) and provides a sufficient condition for the bound to be tight.

Theorem 1 Let $\mathbf{y}^* \in [k]^{\hat{n}}$ be a global solution to (8) and let $\mathbf{x}^* \in [k]^n$ be a global solution to (6). Then

$$\sum_{u\in[m]} d^2(E_{u*}, \mathbf{x}^*) \le \sum_{u\in[m]} \hat{d}^2(F_{u*}, \mathbf{y}^*), \qquad (9)$$

with equality if there exists $\omega_{h\ell}$ such that $\omega_{h\ell} = w_{ij}$ for all $h, \ell \in [\hat{n}]$, $i \in \mathcal{B}_h$, $j \in \mathcal{B}_\ell$.

Proof See supplementary material. $\square$

Note that, due in part to the presence of intrinsic symmetries in the shapes, the segments partaking in the final consensus are not necessarily connected (for example, the two legs of the cat shape may end up in the same segment). For this reason, we split the segments a posteriori into maximal connected components. This operation can be carried out very efficiently in linear time.

3.3. Complexity issues

By making the shift to base segments, we render the computational complexity of the consensus segmentation computation independent from the number of shape nodes, but dependent on the size of the ensemble and the segmentations therein. Clearly, the ensemble construction does still depend on the actual shape size, but one can adopt scalable, fast shape segmentation algorithms (*e.g. k*-means). As a consequence, our method can become inefficient if the ensemble comprises several thousands of base segments. However, this is unlikely to happen in general as the amount of data required to produce stable segmentations is very limited in practice (see Sect. 4.1). By exploiting the base segments, we are able to keep the shapes in full resolution, without the need for downsampling strategies as adopted in, *e.g.* [LBB11, LBB12, GF08].

Remark 5 In the case that the number of base segments is too large, one can still resort to an additional sampling strategy geared towards sparsifying the weights $\hat{w}_{h\ell}$, similarly to what done in [LRBR*13b, LRBR*13a]. In Sect. 4.3 (super-segmentation) we show a practical example of this sampling approach.

3.4. Algorithmic issues

The optimization problem in (8) can be written in terms of the following, equivalent matrix-factorization problem (see Thm. 2 in the supplementary material):

$$Y^* \in \arg\min_{Y \in \mathcal{S}} \|C - Y^\top Y\|_M^2, \qquad (10)$$

where Y is taken from the set $\mathcal{S}$ of left-stochastic, binary $k \times \hat{n}$ matrices, and $\|A\|_M = \sqrt{\sum_{h\ell} M_{h\ell} A_{h\ell}^2}$ denotes the weighted Frobenius matrix norm with weight matrix M. The weight matrix has entries $M_{h\ell} = \hat{w}_{h\ell}$, whereas C is a $\hat{n} \times \hat{n}$ matrix with entries

$$C_{h\ell} = \frac{1}{m} \sum_{u \in [m]} \mathbb{1}_{F_{uh} = F_{u\ell}}. \qquad (11)$$

This matrix, which is known as *co-association matrix* in the consensus clustering literature, summarizes the co-occurrence statistics of the base segments within the ensemble. In other words, $C_{h\ell}$ holds the fraction of segmentations in the ensemble, where the two base segments h and ℓ belong both to the same segment.

The feasible sets of (8) and (10) are isomorphic, as the left-stochastic binary matrices in $\mathcal{S}$ are an alternative representation for the segmentations in $[k]^{\hat{n}}$. Indeed, any segmentation $\mathbf{x} \in [k]^{\hat{n}}$ is identified by a matrix $Y \in \mathcal{S}$ with components $Y_{kh} = \mathbb{1}_{k=x_h}$, while for any Y we can construct a segmentation $\mathbf{x} \in [k]^{\hat{n}}$ such that $x_h = \arg\max_k Y_{kh}$. In Thm. 2 of the supplementary material we show that the two optimization problems are actually equivalent.

To optimize (10), we have implemented the approach proposed in [LRBR*13b, LRBR*13a], properly adapted to cope also with weight matrices having non-zero diagonal. In fact, the solution proposed in the aforementioned works assumes for technical reasons a weight matrix with null diagonal. The optimization procedure works on a relaxation of (10), where the binary constraints are replaced with nonnegativity constraints, and finds in general a local solution to (10). The problem is indeed non-convex and finding the global solution is NP-hard. We refer the reader to [LRBR*13b, LRBR*13a] for details about the optimization algorithm.

4. Experimental results

We performed a wide range of experiments to assess the validity of our approach. The Laplace-Beltrami operator on all shapes was computed using the cotangent weight scheme [MDSB02]. The first set of experiments, aimed at investigating detector repeatability and the sensitivity of our method to different parametrizations, was performed on the TOSCA dataset [BBK08]. This dataset consists of 9 shape classes, represented as triangular meshes, undergoing nearly isometric (*i.e.* almost inelastic) transformations. Shape sizes range from approximately 4000 to 50,000 vertices, and

Figure 3: *Example of consensus segmentation on different shape classes. The detected regions remain stable even under inter-class variations.* david, centaur *and* victoria *are from the TOSCA dataset, the* kid *is from [RRBW*14].*

ground-truth point-to-point correspondences among all the shapes belonging to the same class are available. For fair comparisons with [LBB11] we downsampled all meshes to at most 10,000 vertices. Since all the shapes from the adopted datasets are represented as triangle meshes, following Sect. 3 we model the associated segmentation problem directly on the face elements, corresponding to approximately 20,000 nodes per shape after decimation. The GPS function g is projected onto mesh faces via the linear combination $g(t) = \frac{1}{3} \sum_i g(t_i)$, where t is a mesh triangle and $i \in \{1, 2, 3\}$ ranges over its vertices. In Fig. 3 we show segmentations obtained on a subset of the dataset.

Comparisons with the state of the art were performed on a different dataset, namely the SHREC'10 feature detection and description benchmark [BBB*10]. Differently from the TOSCA dataset, where shapes are only transformed via near isometries, shapes from the SHREC'10 benchmark undergo nine different types of deformations, namely: *isometry* (similar to TOSCA), *noise* and *shotnoise* (additive positional noise on the surface points), *holes* and *microholes* (simulating partiality and occlusions), *sampling* (mesh simplification down to 20% of the original points), *scale* and *localscale*, and *topology* (drastic perturbations of mesh connectivity resulting in large metric distortion). Each transformation appears in 5 intensity levels on a total of 3 shape classes (*man*, *dog*, *horse*). Again, all models are decimated and ground-truth correspondences between shapes belonging to the same class are given. See Fig. 6 for examples.

Detector repeatability. Following previous studies [MTS*05, LBB11], the main measure we adopt for the quantitative assessment of our method is the *repeatability* of the consensus segmentation when seen as a region-wise feature detector. Let $\mathcal{M}$ and $\mathcal{N}$ be respectively a transformed shape and its "null" counterpart (*i.e.* in a canonical pose with no other transformation applied), and let $\mathcal{M}^{(1)}, \dots, \mathcal{M}^{(k_m)}$ and $\mathcal{N}^{(1)}, \dots, \mathcal{N}^{(k_n)}$ denote regions
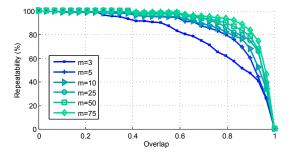
Figure 4: *Sensitivity of consensus segmentation w.r.t. number of segmentations (*m*) in the initial ensemble. Performance does not increase significantly for* m > 75.

detected in the two shapes. Since the ground-truth correspondence is available, for each region $\mathcal{M}^{(j)}$ we can compute its image in $\mathcal{N}$ and obtain $\tilde{\mathcal{N}}^{(j)}$. Given two regions $\mathcal{N}^{(i)}$ and $\mathcal{M}^{(j)}$, we can now define their *overlap* as the area ratio

$$O(\mathcal{N}^{(i)}, \mathcal{M}^{(j)}) = \frac{A(\mathcal{N}^{(i)} \cap \tilde{\mathcal{N}}^{(j)})}{A(\mathcal{N}^{(i)} \cup \tilde{\mathcal{N}}^{(j)})} \qquad (12)$$

$$= \frac{A(\mathcal{N}^{(i)} \cap \tilde{\mathcal{N}}^{(j)})}{A(\mathcal{N}^{(i)}) + A(\tilde{\mathcal{N}}^{(j)}) - A(\mathcal{N}^{(i)} \cap \tilde{\mathcal{N}}^{(j)})},$$

where $A(\mathcal{R}) = \sum_{p \in \mathcal{R}} da(p)$ defines the area of region $\mathcal{R}$.

Given an overlap value $o \in \mathbb{R}$, the *repeatability at o* is defined as the percentage of regions detected in both shapes at corresponding locations, and having overlap greater than o. Note that, since all the shapes in both datasets exhibit global (bilateral) intrinsic symmetries, we can give two different definitions for (12) depending on whether we use the ground-truth map $\mathcal{M}^{(j)} \mapsto \tilde{\mathcal{N}}^{(j)}$ or its symmetric counterpart. Following standard practice [OBCS*12, LBB11], all our plots report the best results over the two definitions.

4.1. Sensitivity analysis

In this section we evaluate the robustness of our method under different parametrizations as well as different initial requirements on the data (*e.g.* lack of a mesh structure). Unless stated otherwise, all the experiments in this section are performed on the TOSCA dataset.

Size of the ensemble. The first experiment is aimed at evaluating the sensitivity of the consensus segmentation when fed with ensembles E of different sizes (m). Let k_u be the number of segments in the uth segmentation. We fix $k_u = 10$ segments for all $u = 1, \ldots, m$, and at most $k = 5$ regions in the final consensus. Since our approach is based on evidence accumulation, in theory we can expect more accurate results as the number of provided clusterings increases. Fig. 4 shows the results of this experiment for various values of m. As indicative values, with the other parameters fixed, the total number of base segments (Sect. 3) range from 20 (for m = 3) to 800 (for m = 75) on average. As we can see from the plots, the segmentations quickly become stable as we push

the size of the ensemble towards moderately larger values. Note that with only m = 5 initial segmentations we already reach very stable results, with around 80% repeatability at overlap 0.8 (*i.e.* 80% of the detected regions are exactly reproducible to within 80% of their area). This result contrasts with the few previous attempts at evidence-based shape analysis (*e.g.* [GF08] required as many as 1200 random cuts in order to produce sufficiently good partitions).

Initial number of segments. Due to the randomized seeding scheme that we follow in the *k*-means clustering step, increasing the total number of segments in each initial clustering is likely to produce noisy segmentations, unless enough data is provided. One may then ask how much additional evidence is needed in order to produce stable segmentations. We performed this analysis by computing, for different values of k_u, the number of segmentations required to obtain 80% repeatability at overlap 0.8. We observed a linear dependency among the two quantities, with m = 5 segmentations required for $k_u = 10$, and around m = 40 for $k_u = 20$. Note, however, that setting the initial number of segments to values larger than 10 is rarely needed; we refer to Sect. 4.3 for one possible scenario.

Final number of segments. Given the results from the previous experiments, one may still wonder about the general validity of our analysis for different values of k (*i.e.* the maximum number of detected components). In other words, is there a correlation between the optimal size of the initial ensemble and k? To answer this question, we performed a similar experiment to the previous section, substituting k_u by k. We could not observe any evident correlation among the two quantities, with m $\geq$ 25 yielding consistently high repeatability for all $k = 3, \ldots, 10$. Together with the previous experiments, these results demonstrate the stability of the consensus estimate under different parametrizations and limited amount of data.

Representation and sampling. In Fig. 5 we demonstrate the applicability of our method for the task of stable component detection on point clouds. For this experiment we used two nearly-isometric deformations of a human shape from the SHREC'10 dataset, and removed the triangulation from one of the two. The Laplacian on the resulting point cloud was approximated using the Gaussian-weighted construction of [BSW09]. Note that our general framework does not require any modification as we change representation from triangulated meshes to point clouds (Sect. 3). While spurious components still appear in very noisy settings (rightmost point cloud in Fig. 5), the overall segmentation remains quite stable even under aggressive downsampling. In a separate set of experiments, we investigated the influence of the number of eigenfunctions used for constructing the descriptors (Table 1). While the dimensionality of the descriptor has a direct influence on the computational time required by the clustering step, the segmentation itself is not significantly affected.

Figure 5: *Consensus segmentation on point clouds of increasing sparsity. A shape of 52k vertices (top left) is transformed to a different pose and then, after removing the triangulation, subsampled to 20%, 10%, and 2% of its original points (left to right). Segmentations remain stable under the nearly isometric deformation and the change of representation. Note that the consensus segmentation is performed on triangles in the first case, whereas it is vertex-based in the other cases where connectivity information is lost.*

Indeed, small values already give a good trade-off between efficiency and quality of the detector.

	15	**50**	**100**	**200**
overlap 0.5	96.47	97.45	97.97	98.44
overlap 0.7	84.96	90.34	92.82	95.31
overlap 0.9	71.35	79.56	83.53	85.23

Table 1: *Detector repeatability as a function of descriptor size (in bold).*

4.2. Comparisons

In Fig. 7(a)(b) we compare with the state-of-the-art methods of Litman *et al.* [LBB11] and Sipiran and Bustos [SB13]. Note that repeatability of the latter method (red curve) is only measured on those portions of the mesh that are repeatable against non-rigid transformations (by definition of key component [SB13]), and thus a fair comparison cannot be made (hence the dashed stroke).

In these experiments we set the size of the ensemble to $m = 50$ (number of segmentations) and $k_u = 10$ (number of regions per segmentation). These values were chosen to give a good trade-off between efficiency and stability, based on the previous experiments. We set the final number of segments to $k = 5$ for all shape classes, resulting in approximately the same number of regions as the other methods. Being based on the analysis we performed on a subset of the TOSCA dataset (which only covers nearly isometric deformations), the parameters were not optimized for benchmark performance in any way. The plots in Fig. 7 show quality improvement over the other methods across the whole benchmark. In particular, our detector yields significantly better results on the *sampling* and *scale* classes, where we get close to ideal performance (detailed curves from [LBB11, SB13] are not reproduced here for space reasons). Visual results on

several deformations of a human shape from the SHREC'10 dataset are depicted in Fig. 6.

It is worth mentioning that, differently from Litman *et al.*, our method produces a set of *non*-overlapping regions. Even though overlapping regions might be more stable in principle, the conducted experiments demonstrate that our solution yields superior results over a wide range of data transformations. Nonetheless, the applicability of a consensus-seeking scheme for overlapping regions might be an interesting direction for future investigations.

4.3. Shape matching

Arguably one of the main applications of stable feature detection on deformable shapes is shape retrieval and matching [BBB*10]. Recent state-of-the-art approaches base their success upon the ability to detect repeatable regions on the given shapes [PBB*13, OBCS*12]. The assumption of repeatable regions implies that there exists some unknown permutation relating the two sets of detected components; once a solution to this "coarse" matching problem is found, a dense correspondence can be inferred over the entire shapes.

The simplest way to define a *region descriptor* for a given component $\mathcal{R} \subset \mathcal{M}$ is by computing the area-weighted average of its q-dimensional pointwise descriptors $\beta : \mathcal{M} \to \mathbb{R}^q$,

$$\hat{\beta}(\mathcal{R}) = \sum_{p \in \mathcal{R}} \beta(p) \mathrm{d}a(p). \tag{13}$$

Even though other choices are also possible [LBB11], in this paper we will consider the previous definition for the sake of simplicity.

Comparisons. In Fig. 7(c) we evaluate the performance of our feature detector when employed within the following
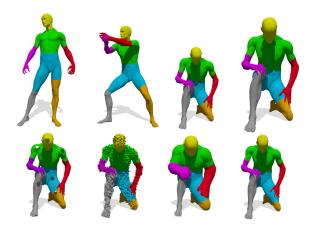

Figure 6: *Stable regions detected via consensus segmentation on a human shape (SHREC'10 dataset). The figure shows the detected regions under different transformations, namely: near-isometry and global rescaling (first row), holes, noise, localscale, and microholes (second row).*
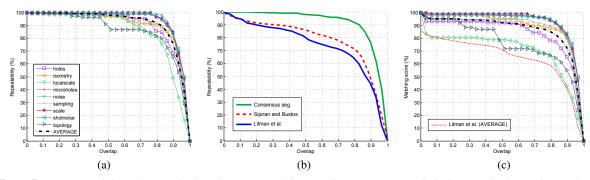
Figure 7: *(a) Repeatability of our method on the SHREC'10 dataset. (b) Comparison with the best results from other methods; plotted are the* average *scores over all deformations. The red curve is dashed due to differences in the evaluation (see text). (c) Matching score attained by our method (all deformations + average) and Litman* et al. *(average) on the SHREC'10 dataset.*
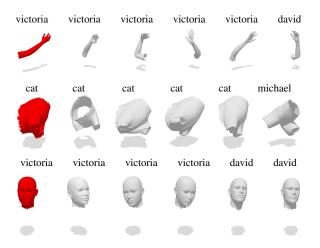

Figure 8: *Example of shape retrieval on the TOSCA dataset using our region detector and SI-HKS as a descriptor. Depicted in red is the query region, from left to right we show the 1st, 2nd, 4th, 10th, and 15th closest matches.*

matching pipeline. For each region $\mathcal{N}^{(i)}$ in the null shape, we define its first match in $\mathcal{M}$ as the region minimizing

$$\mathcal{M}^{(i*)} = \underset{\mathcal{M}^{(j)} \subset \mathcal{M}}{\arg\min} \, \|\hat{\beta}(\mathcal{N}^{(i)}) - \hat{\beta}(\mathcal{M}^{(j)})\|_2 , \qquad (14)$$

i.e. the nearest neighbour of $\mathcal{N}^{(i)}$ in descriptor space. Following [LBB11], we then define the *matching score* at overlap o as the ratio

$$\text{score}(o) = \frac{|\{O(\mathcal{N}^{(i)}, \mathcal{M}^{(i*)}) \geq o\}|}{k_n} . \qquad (15)$$

An ideal detector-descriptor pair has a constant matching score of 1. As in [LBB11], in this experiment we use the scale-invariant heat kernel signature (SI-HKS) [BK10] to construct the region descriptors according to Eq. (13). Some examples of region matching for a shape (segment) retrieval application are depicted in Fig. 8.

Super-segmentation. Similarly to other approaches [SOCG10, LBB11], our method requires the desired num-

ber of regions to be specified approximately. While in general it is desirable to determine the optimal number of segments automatically, this property could be exploited to *oversegment* a given shape, one possible application being the definition of "coarse" levels in a hierarchical matching pipeline. We explore this possibility by detecting $k = 50$ regions starting from an ensemble of $m = 150$ segmentations, each composed of $k_u = 100$ segments. With these parameters, the number of base segments grows up to 12k (a 40-fold increase in size with respect to the average scenario). In this situation, following the analysis of Sect. 3.3, our baseline method can perform poorly. In Fig. 9 we show the solution we obtain after applying the sampling strategy described in the same section, with sampling rate equal to 5%. Note that in this experiment we are keeping the shapes at *full resolution* (around 55k triangles).

Figure 9: *Oversegmenting nearly isometric deformations of a shape still produces repeatable regions. The smaller triangle-sized segments are a result of the sampling process and can be filtered out a posteriori.*

4.4. Performance

As described in Sect. 3.3, consensus segmentation is an efficient operation as its performance does not depend directly on shape size. Here we give an empirical evaluation of the computational times required by our algorithm. The creation of the starting ensemble E was written in Matlab using a standard *k*-means implementation, and trivially parallelized over 4 cores on an Intel Core i7 with 8GB memory (more efficient alternatives can be considered for this step). The main consensus algorithm was implemented in C++. In Table 2

we report the average time required to produce stable segmentations on the TOSCA dataset, using different ensemble sizes. We used the first 50 eigenfunctions of the LB operator to compute GPS descriptors. The other parameters are set to the default values of $k_u = 10$ and $k = 5$.

	m = 10	m = 25	m = 50	m = 75
build E	1.94	4.64	9.24	14.05
optimize	0.02	0.12	0.32	0.76

Table 2: *Average times (sec) required by our algorithm on shapes of 10k vertices.*

5. Conclusions

We have proposed a novel region-based feature detector for three-dimensional shapes, which can be regarded as a robust alternative to the widespread point-based detectors. Our approach builds around the idea, imported from the consensus clustering literature, that robust and stable segments can be obtained by eliciting a consensus among a set of putative segmentations. We have indicated an effective ensemble creation procedure together with a novel consensus segmentation algorithm, which is made practical by a shift of the clustering domain, from shape nodes to so-called base segments. A thorough quantitative evaluation of the proposed solution on standard datasets demonstrates its effectiveness compared to other state-of-the-art methods.

Acknowledgments

References

[AG12] ALUSH A., GOLDBERGER J.: Ensemble segmentation using efficient integer linear programming. *IEEE Trans. Pattern Analysis and Machine Intelligence 34*, 10 (2012), 1966–1977. 2

[ASC11] AUBRY M., SCHLICKEWEI U., CREMERS D.: Pose-consistent 3d shape segmentation based on a quantum mechanical feature descriptor. In *Proc. DAGM* (2011). 2

[BBB*10] BRONSTEIN A., BRONSTEIN M., BUSTOS B., ET AL.: SHREC 2010: robust feature detection and description benchmark. In *3DOR* (2010), pp. 79–86. 6, 8

[BBK08] BRONSTEIN A., BRONSTEIN M., KIMMEL R.: *Numerical Geometry of Non-Rigid Shapes*, 1 ed. Springer Publishing Company, Incorporated, 2008. 6

[Ber03] BERGER M.: *A panoramic view of Riemannian geometry*. Springer, Berlin, New York, Heidelberg, 2003. 3

[BK10] BRONSTEIN M., KOKKINOS I.: Scale-invariant heat kernel signatures for non-rigid shape recognition. In *Proc. CVPR* (2010), pp. 1704–1711. 9

[BPVR11] BENJAMIN W., POLK A. W., VISHWANATHAN S., RAMANI K.: Heat walk: Robust salient segmentation of non-rigid shapes. *Computer Graphics Forum 30*, 7 (2011), 2097–2106. 2

[BSW09] BELKIN M., SUN J., WANG Y.: Constructing Laplace operator from point clouds in Rd. In *ACM-SIAM Symposium on Discrete Algorithms* (2009), pp. 1031–1040. 7

[FJ05] FRED A., JAIN A.: Combining multiple clustering using evidence accumulation. *IEEE Trans. Pattern Analysis and Machine Intelligence 27*, 6 (2005), 835–850. 2

[GBAL09] GĘBAL K., BÆRENTZEN J. A., AANÆS H., LARSEN R.: Shape analysis using the auto diffusion function. In *Proc. SGP* (2009), pp. 1405–1413. 2

[GF08] GOLOVINSKIY A., FUNKHOUSER T.: Randomized cuts for 3D mesh analysis. *ACM Trans. Graph. 27*, 5 (2008). 2, 5, 7

[LBB11] LITMAN R., BRONSTEIN A. M., BRONSTEIN M. M.: Diffusion-geometric maximally stable component detection in deformable shapes. *Computers and Graphics 35*, 3 (2011), 549 – 560. 2, 4, 5, 6, 7, 8, 9

[LBB12] LITMAN R., BRONSTEIN A. M., BRONSTEIN M. M.: Stable volumetric features in deformable shapes. *Computers and Graphics 36*, 5 (2012), 569–576. 1, 2, 5

[LRBR*13a] LOURENÇO A., ROTA BULÒ S., REBAGLIATI N., FRED A. L. N., FIGUEIREDO M. A. T., PELILLO M.: Consensus clustering using partial evidence accumulation. In *IbPRIA* (2013), pp. 69–78. 4, 5, 6

[LRBR*13b] LOURENÇO A., ROTA BULÒ S., REBAGLIATI N., FRED A. L. N., FIGUEIREDO M. A. T., PELILLO M.: Probabilistic consensus clustering using evidence accumulation. *Machine Learning* (2013). online. 4, 5, 6

[MDSB02] MEYER M., DESBRUN M., SCHRÖDER P., BARR A. H.: Discrete differential-geometry operators for triangulated 2-manifolds. In *VisMath* (2002). 6

[MTS*05] MIKOLAJCZYK K., TUYTELAARS T., SCHMID C., ZISSERMAN A., ET AL.: A comparison of affine region detectors. *Int. J. Comput. Vis. 65*, 1-2 (2005), 43–72. 1, 6

[OBCS*12] OVSJANIKOV M., BEN-CHEN M., SOLOMON J., BUTSCHER A., GUIBAS L.: Functional maps: a flexible representation of maps between shapes. *ACM Trans. Graph. 31*, 4 (2012), 30:1–30:11. 1, 7, 8

[PBB*13] POKRASS J., BRONSTEIN A. M., BRONSTEIN M. M., SPRECHMANN P., SAPIRO G.: Sparse modeling of intrinsic correspondences. *Computer Graphics Forum 32*, 2pt4 (2013), 459–468. 1, 8

[Reu10] REUTER M.: Hierarchical shape segmentation and registration via topological features of laplace-beltrami eigenfunctions. *Int. J. Comput. Vis. 89*, 2-3 (2010), 287–308. 2

[RRBW*14] RODOLÀ E., ROTA BULÒ S., WINDHEUSER T., VESTNER M., CREMERS D.: Dense non-rigid shape correspondence using random forests. In *Proc. CVPR* (2014). 6

[Rus07] RUSTAMOV R. M.: Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Proc. SGP* (2007), pp. 225–233. 3

[SB13] SIPIRAN I., BUSTOS B.: Key-components: detection of salient regions on 3d meshes. *The Visual Computer 29*, 12 (2013), 1319–1332. 2, 8

[SG02] STREHL A., GHOSH J.: Cluster ensembles - a knowledge reuse framework for combining multiple partitions. *J. of Machine Learning Research*, 3 (2002), 583–617. 2

[SOCG10] SKRABA P., OVSJANIKOV M., CHAZAL F., GUIBAS L.: Persistence-based segmentation of deformable shapes. In *Proc. NORDIA* (June 2010). 2, 4, 9

[YZX*04] YU Y., ZHOU K., XU D., SHI X., BAO H., GUO B., SHUM H.-Y.: Mesh editing with poisson-based gradient field manipulation. *ACM Trans. Graph. 23*, 3 (2004), 644–651. 3