DOI: 10.1111/cgf.14181

COMPUTER GRAPHICS *forum*
*Volume 00 (2020), number 00 pp. 1–15*

# ECHO: Extended Convolution Histogram of Orientations for Local Surface Description

Thomas W. Mitchel,[1] Szymon Rusinkiewicz,[2] Gregory S. Chirikjian[3,1] and Michael Kazhdan[1]

[1]Johns Hopkins University
tmitchel@jhu.edu, misha@cs.jhu.edu
[2]Princeton University
[3]National University of Singapore

**Abstract**

*This paper presents a novel, highly distinctive and robust local surface feature descriptor. Our descriptor is predicated on a simple observation: instead of describing the points in the vicinity of a feature point relative to a reference frame at the feature point, all points in the region describe the feature point relative to their own frames. Isometry invariance is a byproduct of this construction. Our descriptor is derived relative to the extended convolution – a generalization of the standard convolution that allows the filter to adaptively transform as it passes over the domain. As such, we name our descriptor the Extended Convolution Histogram of Orientations (ECHO). It exhibits superior performance compared to popular surface descriptors in both feature matching and shape correspondence experiments. In particular, the ECHO descriptor is highly stable under near-isometric deformations and remains distinctive under significant levels of noise, tessellation, complex deformations and the kinds of interference commonly found in real data.*

**Keywords:** 3D shape matching, modeling, curves and surfaces, modeling, computer vision – shape recognition, methods and applications

## 1. Introduction

Local feature descriptors play a critical role in both image and shape recognition applications. Generally, the initial step in such paradigms involves identifying a number of keypoints on a 2D image or 2D manifold. The purpose of local feature descriptors is to provide a distinctive characterization of the region surrounding each keypoint, which can then be compared to establish point-to-point correspondences between images or surfaces. Successful image descriptors, such as SIFT [Low99, Low04] and SURF [BTVG06, BETVG08], are both highly descriptive, in that characterizations of different neighbourhoods are sufficiently unique so as to differentiate between the two without ambiguity, repeatable, in that descriptions of regions that are fundamentally the same are nearly identical, and robust under nuisance parameters including noise and affine transformations. Similarly, popular surface descriptors, *e.g.* SHOT [TSDS10b, STDS14] and RoPS [GSB*13], are insensitive to noise, mesh resolution and rigid transformations.

For both images and surfaces, the majority of successful descriptors rely on a local frame to encode the neighbourhood about a keypoint. Typically, the construction of these descriptors consists of first defining a rotationally equivariant frame at the keypoint and then describing the neighbouring region relative to that frame; such approaches ensure that the region can be encoded without discarding discriminating information and that the descriptor is itself rotationally invariant. For shapes, the use of repeatable and noise-robust local frames significantly improves descriptor performance [PD11].

In this paper we present a novel, highly descriptive and robust surface feature descriptor based on a simple observation: instead of describing the local region relative to the frame at the feature point, we have all points in a local region describe the feature point relative to their own frames.

To this end, we leverage an operation called *extended convolution* which allows the convolution filter to adaptively rotate as it passes over a Euclidean domain. Extending the operation to 2D manifolds, we derive an optimal surface filter, *Extended Convolution Histogram of Orientations* (**ECHO**). Our code is freely available at https://github.com/mkazhdan/EchoDescriptors.

**Contribution.** The proposed surface descriptor is highly descriptive and is invariant to isometric deformations. It exhibits superior performance, in terms of both descriptiveness and insensitivity to noise and tessellation, relative to SHOT and RoPS.

## 2. Related Work

### Image descriptors

Owing to its high descriptiveness, insensitivity to changes in both illumination and viewpoint, and remarkable success in a variety of applications, the SIFT descriptor [Low99, Low04] has distinguished itself as one of the premier image feature descriptors. One of the key contributions of SIFT was the achievement of invariance under the action of 2D similarity transformations. In the SIFT pipeline, the scale corresponding to an image keypoint is determined by the point's location in scale-space and the orientation is determined by the direction of the gradient at the point. The point's neighbourhood is then encoded relative to the frame corresponding to the assigned scale and orientation to achieve invariance.

### Surface descriptors

The success of SIFT has helped to establish the construction of a descriptor relative to a local reference frame as the *de facto* standard amongst techniques used to achieve rotation invariance. However, unlike images, surfaces have no inherent signal to facilitate the construction of frames. To achieve invariance under rigid transformations, prior surface descriptors defined relative to intrinsic parameterizations, such as ISC [KBLB12], have sacrificed descriptive potential. In this context, it is not surprising that surface descriptors able to define frames generally exhibit superior overall descriptor performance [PD11]. Of these descriptors, SHOT [TSDS10b] RoPS [GSB*13], and USC [TSDS10a] are the most popular, and have been shown to outperform competing methods in terms of descriptiveness and robustness under a variety of nuisance parameters [STDS14, GBS*16].

The effectiveness of these descriptors is underpinned by the construction of frames based on the surface's principal curvature directions. Specifically, SHOT, RoPS, and USC compute a weighted covariance matrix centred at the point of interest. The eigenvectors of the resulting matrix can be interpreted as a smoothed version of the principal curvature directions. As long as the principal curvature values are distinct, a rigid frame can be constructed from the eigenvectors, though it is unique only up to sign. Each of these descriptors employ techniques to eliminate this ambiguity so as to produce a single repeatable frame.

A number of contemporaneous learned surface descriptors have been shown to significantly outperform SHOT and other handcrafted descriptors in certain applications [KZK17, WGY*18, DBI18, DBI19, SSS19, CPK19]. However, most of these methods learn local descriptors from existing handcrafted techniques [KZK17, DBI18], rather than input data [SSS19]. More generally, many state-of-the art pipelines for shape registration and correspondence directly incorporate "deterministic" descriptors such as SHOT in some capacity [VLB*17, LYLG18, DSL*19] and outperform learned approaches with the proper settings [DSL*19, SSDS19].

### Beyond standard convolution

Convolution incorporating rotational steering has been recognized as an effective tool for template matching since the early work of Freeman and Adelson [FA91, SF96, THO99], and has been extended to incorporate the actions of larger groups [KC99a, KC99b, CK16]. More recently, the advent of deep learning in imaging, vision and graphics has coincided with the development of group equivariant [CW16a, WHS18, BLV*18, WB18, WW19] and steerable [CW16b, WGTB17, WGW*18, EABMD18, WC19] convolutional neural networks (CNNs), which present a general notion of equivariant convolutions on homogeneous spaces. The majority of these techniques are concerned with convolutions over either flat Euclidean spaces or the transformation groups themselves, and most relevant to our work are a number of contemporary approaches for defining equivariant convolutions on surfaces. Generally speaking, these methods can be grouped into two categories: convolutions defined with respect to local geodesic parameterizations of the surface and graph-based convolutions.

The primary challenge in performing convolution on surfaces lies in resolving the ambiguity between the orientation of the local surface and that of the filter or kernel. Convolutions defined in terms of geodesic parameterizations have addressed this issue by either pooling over the responses corresponding to a discrete set of filter orientations [MBBV15, MBM*17, PO18] or defining local reference frames relative to the extrinsic properties of the surface [BMRB16, SDL18]. The former approach limits discriminative potential while the latter can be unstable under deformations.

Most graph-based convolutions eliminate orientation ambiguity by utilizing radially isotropic filters [PDKS19, ZXW*19]. An exception are gauge equivariant convolutions [CWKW19] and their generalization to triangular meshes [dHWCW20], which achieve equivariance to transformations of the frame field by placing specific constraints on the filter which, when applied to higher-order tensors, are not as restrictive as the constraint of radial symmetry. While graph-based methods can be computed efficiently [KW17, DBV16], their applications are limited as the size of the filter cannot exceed that of the immediate graph neighbourhood.

## 3. Method Overview

The proposed descriptor is derived from an operation called *extended convolution* [MBK*20]. This is an extension of standard convolution that allows the filter to adaptively transform as it travels over a Euclidean domain. The extended convolution is similar to earlier works on steerable filters [FA91, SF96, THO99] though it differs in that its formulation is defined independently of the choice of filter. Here, we expand the extended convolution framework to 2D manifolds, fixing the class of transformations to rotations.

Given a point of interest on a surface and surrounding neighbourhood, our proposed surface descriptor corresponds to the filter that maximizes the response of the extended convolution at the keypoint. We derive the optimal filter for the keypoint and describe how the

optimal filter can be discretized as a two-dimensional array. The construction is straightforward: we attach frames to all points inside the neighbourhood using the gradient of an intrinsic signal defined on the surface. Then, we describe the position of the keypoint in the tangent space of each neighbour point using the logarithm map.

## 4. Extended Convolution

We begin by reviewing extended convolution for processing images and then describe how it can be extended to process signals on surfaces.

### 4.1. Extended convolution on images

Extended convolution generalizes the standard convolution by allowing the filtering process to take into account local transformations. Consider functions $h,\ f : \mathbb{R}^2 \to \mathbb{R}$, which we call the 'signal' and 'filter', respectively, and a *frame field* $\mathfrak{R}$, which we define to be a mapping such that $\mathfrak{R}_{\mathbf{q}}$ specifies a rotation in SO(2), for all $\mathbf{q} \in \mathbb{R}^2$. The planar *extended convolution* of a signal $h$ and frame field $\mathfrak{R}$ with a filter $f$, evaluated at $\mathbf{p} \in \mathbb{R}^2$ is given by [MBK*20]

$$(\{h, \mathfrak{R}\} * f)(\mathbf{p}) = \int_{\mathbb{R}^2} h(\mathbf{q})\big[\mathfrak{R}_{\mathbf{q}}(f)\big](\mathbf{p} - \mathbf{q})\, d\mathbf{q}. \tag{1}$$

The expression $[\mathfrak{R}_{\mathbf{q}}(f)](\mathbf{p} - \mathbf{q})$ denotes the evaluation of the function $\mathfrak{R}_{\mathbf{q}}(f)$ at the point $\mathbf{p} - \mathbf{q}$ and the rotation of $f$ by $R \in \mathrm{SO}(2)$ is defined in the usual manner as

$$R(f) \equiv f \circ R^{-1}.$$

Equation (1) can be interpreted as setting the value at $\mathbf{p}$ by iterating over all points $\mathbf{q}$ in the domain and summing the value of the signal $h(\mathbf{q})$ weighted by the evaluation of the filter $f$ at the position of $\mathbf{p}$ *in the frame of* $\mathbf{q}$.

### 4.2. Extended convolution on surfaces

Here, we introduce an analogous definition of the extended convolution for surfaces, which will be the focus of our subsequent analysis. The key idea is to note that for points $\mathbf{p}$ and $\mathbf{q}$ on a surface, one can express the position of $\mathbf{p}$ *relative to* $\mathbf{q}$ by using the logarithm at $\mathbf{q}$.

Given an oriented Riemannian manifold $S$, we define a frame field $\mathfrak{R}$ to be a mapping associating to each point $\mathbf{q} \in S$ an orthonormal map from Euclidean 2-space to the tangent space at $\mathbf{q}$. Formally, if $\mathcal{F}S$ is the fiber bundle with $\mathcal{F}S_{\mathbf{q}}$ the group of orthonormal transformations from $\mathbb{R}^2$ to $T_{\mathbf{q}}S$, then $\mathfrak{R} \in \Gamma(\mathcal{F}S)$ is a section of this bundle.

Then, given a signal $h : S \to \mathbb{R}$, a frame field $\mathfrak{R} \in \Gamma(\mathcal{F}S)$, and a filter $f : \mathbb{R}^2 \to \mathbb{R}$ we define the extended convolution of $h$ and $\mathfrak{R}$ with $f$ as the function on $S$ with

$$(\{h,\ \mathfrak{R}\} * f)(\mathbf{p}) = \int_{\mathbf{q} \in S} h(\mathbf{q})\ \big[\mathfrak{R}_{\mathbf{q}}(f)\big]\big(\log_{\mathbf{q}}(\mathbf{p})\big)\ d\mathbf{q}, \tag{2}$$

with $\log_{\mathbf{q}}(\mathbf{p})$ the logarithm of $\mathbf{p}$ with respect to $\mathbf{q}$, giving the position of the point $\mathbf{p}$ relative to its neighbour $\mathbf{q}$.

As the logarithm function is one to many, we use the tangent vector defining the shortest geodesic from $\mathbf{q}$ to $\mathbf{p}$. This makes the integrand well-defined as long as $\mathbf{p}$ is not on the cut-locus of $\mathbf{q}$. In practice we have not found this to be a problem as the cut-locus is a set of measure zero and the integrand is bounded, so the integral is well-defined.

#### Relationship to frame-less methods

Prior approaches for defining convolutions on surfaces are predicated on the assumption that there is no way to define a repeatable frame field on a surface. As a result, the convolution operation is designed so that the response is independent of the choice of reference orientation. In contrast, the extended convolution is specifically designed such that the response depends on the choice of frame field $\mathfrak{R}$, which is treated as an input parameter. This approach provides it with a degree of flexibility that makes it well-suited for a variety of different applications in image processing [MBK*20].

Critically, we show it is relatively straightforward to define a repeatable frame field $\mathfrak{R}$ on a given surface under which the response of the extended convolution remains consistent between different locations and meshes. Moreover, the extended convolution can be made invariant to isometries of the local surface by appropriate choices of the signal $h$ and transformation field $\mathfrak{R}$, without any restrictions on the filter $f$.

## 5. Application to Surface Feature Detection

The matching of local surface descriptors can be characterized as a feature detection problem; our strategy is to design a filter whose extended convolution will have a high response wherever the local surface resembles the target. We then define the ECHO descriptor to be (the discretization of) this filter.

### 5.1. Deriving the optimal filter

Given a surface $S$ and an associated frame field $\mathfrak{R}$ and signal $h$, we consider a feature point $\mathbf{p} \in S$. Our first objective is to define a filter $f^{\mathbf{p}} \in L^2(\mathbb{R}^2)$ which, of all possible filters, maximizes the extended convolution (up to scale) at $\mathbf{p}$.

With $\mathfrak{R}$ and $h$ fixed, the extended convolution can be thought of as a map $\mathcal{E} : L^2(\mathbb{R}^2) \to L^2(S)$ from the space of filters to the space of functions on $S$. From Equation (2) it is clear that $\mathcal{E}$ is linear, as must be the map $\mathcal{E}^{\mathbf{p}}(f) \equiv [\mathcal{E}(f)](\mathbf{p}) = (\{h,\ \mathfrak{R}\} * f)(\mathbf{p})$, obtained by evaluating the function returned by $\mathcal{E}$ at a point $\mathbf{p} \in S$.

Using the fact that the space of filters, $L^2(\mathbb{R}^2)$, is an inner-product space (with the inner-product of two filters defined by integrating their product) and applying the Riesz Representation Theorem, there exists a filter $\omega^{\mathbf{p}} \in L^2(\mathbb{R}^2)$ such that

$$\mathcal{E}^{\mathbf{p}}(f) = \int_{\mathbb{R}^2} f \cdot \omega^{\mathbf{p}}, \quad \forall f \in L^2\big(\mathbb{R}^2\big).$$

In particular, up to scale, the filter $f^{\mathbf{p}}$ that maximizes the response of the extended convolution at $\mathbf{p}$ must be $\omega^{\mathbf{p}}$.

This naturally leads to the question:

*What is the value of* $\omega^{\mathbf{p}}$ *at different points* $\mathbf{x} \in \mathbb{R}^2$?

To answer this question, we note that one way to evaluate $\omega^{\mathbf{p}}$ at a point $\mathbf{x}$ is to integrate it against the delta function at $\mathbf{x}$, which we showed above is the evaluation of the extended convolution with the delta function, at the point $\mathbf{p}$:

$$\omega^{\mathbf{p}}(\mathbf{x}) = \int_{\mathbb{R}^2} \delta_{\mathbf{x}} \cdot \omega^{\mathbf{p}} = \mathcal{E}^{\mathbf{p}}(\delta_{\mathbf{x}}) = (\{h,\ \mathfrak{R}\} * \delta_{\mathbf{x}})(\mathbf{p})$$
$$\overset{(2)}{=} \int_{\mathbf{q}\in S} h(\mathbf{q}) \cdot \left[\mathfrak{R}_{\mathbf{q}}(\delta_{\mathbf{x}})\right]\left(\log_{\mathbf{q}}(\mathbf{p})\right) d\mathbf{q}. \tag{3}$$

In practice, we would like the descriptor of the keypoint $\mathbf{p}$ to be local. To this end we note that if we restrict ourselves to filters that are supported within a disk or radius $\varepsilon$, the filter maximizing the response to the extended convolution at $\mathbf{p}$ is still, up to scale, $\omega^{\mathbf{p}}$, restricted to the disk of radius $\varepsilon$.

Using compactly supported filters, we are only interested in evaluating Equation (3) at points $|\mathbf{x}| \leq \varepsilon$. Since the evaluation of $\delta_{\mathbf{x}}(\log_{\mathbf{q}}(\mathbf{p}))$ will only be non-zero if the geodesic distance between $\mathbf{p}$ and $\mathbf{q}$ is less than or equal to $\varepsilon$, we can write out the value of the maximizing filter for $|\mathbf{x}| \leq \varepsilon$, up to scale, as:

$$f^{\mathbf{p}}(\mathbf{x}) = \int_{\{\mathbf{q}\in S \,|\, d_g(\mathbf{p},\,\mathbf{q})\leq\varepsilon\}} h(\mathbf{q}) \cdot \delta_{\mathbf{x}}\left(C_g^{\mathbf{p}}(\mathbf{q})\right) d\mathbf{q}, \tag{4}$$

where $d_g : S \times S \to \mathbb{R}_{\geq 0}$ is the function giving the geodesic distance between pairs of points on $S$ and $C_g^{\mathbf{p}} : S \to \mathbb{R}^2$ is the function giving the coordinates of the logarithm of $\mathbf{p}$ in the tangent space of points on $S$, expressed relative to the frames at those points:

$$C_g^{\mathbf{p}}(\mathbf{q}) = \mathfrak{R}_{\mathbf{q}}^{-1} \cdot \log_{\mathbf{q}}(\mathbf{p}).$$

Equation (4) can be interpreted as a recipe for constructing the optimal filter $f^{\mathbf{p}}$ as a histogram. Specifically, $f^{\mathbf{p}}$ can be obtained by having every point on the surface $\mathbf{q}$ lying in the geodesic $\varepsilon$-disk around $\mathbf{p}$ "cast a vote" into the histogram bin corresponding to the position of $\mathbf{p}$ in the tangent space at $\mathbf{q}$ with a weight equal to $h(\mathbf{q})$.

#### Maximizing convolution vs. minimizing $L_2$-difference

Given points $\mathbf{p}$ and $\mathbf{q}$ on $S$, and the associated optimal filters $f^{\mathbf{p}}$ and $f^{\mathbf{q}}$, we can compare the regions about $\mathbf{p}$ and $\mathbf{q}$ in one of two ways: We can compute $\mathcal{E}^{\mathbf{p}}(f^{\mathbf{q}})$, the evaluation of the extended convolution with $f^{\mathbf{q}}$ at the point $\mathbf{p}$; or, we can compute $\|f^{\mathbf{p}} - f^{\mathbf{q}}\|$, the $L_2$-difference between the filters. Using the above derivation, the former can be expressed as:

$$\mathcal{E}^{\mathbf{p}}(f^{\mathbf{q}}) = \langle f^{\mathbf{p}},\ f^{\mathbf{q}} \rangle$$

while the latter can be expanded as:

$$\|f^{\mathbf{p}} - f^{\mathbf{q}}\|^2 = \|f^{\mathbf{p}}\|^2 + \|f^{\mathbf{q}}\|^2 - 2\langle f^{\mathbf{p}},\ f^{\mathbf{q}} \rangle.$$

Thus, if the filters were normalized to have unit length, finding the filter maximizing the extended convolution would be equivalent to finding the filter minimizing the $L_2$-difference.

In our implementation, we leave the filters un-normalized. For example, this implies that different scalar multiples of the signal $h$ give rise to different optimal filters. As such, the value of the extended convolution is not a meaningful measure of regional similarity as a large value of $\mathcal{E}^{\mathbf{p}}(f^{\mathbf{q}})$ could reflect the fact that the signal around $\mathbf{q}$ has large magnitude, not that the signals about $\mathbf{p}$ and $\mathbf{q}$ are well-correlated. Instead, we use the $L_2$-difference.

### 5.2. Local isometry invariance

The form of $f$ in Equation (4) does not by itself ensurRelationship to frame-less methodse invariance of the filter under intrinsic isometries of the local surface, as the expression for $f$ depends on the definition of the signal $h$ and frame field $\mathfrak{R}$. To this end we follow the approach of [MBK*20], choosing an isometry invariant signal $\phi : S \to \mathbb{R}$ on the surface and using the gradient and gradient magnitudes of $\phi$ to define a frame field $\mathfrak{R}$ and signal $h$.

As in the Euclidean domain, given a vector field $\vec{V} \in \Gamma(TS)$, we may define a frame field $\mathfrak{R}^{\vec{V}} \in \Gamma(\mathcal{F}S)$ on the surface via

$$\mathfrak{R}_{\mathbf{q}}^{\vec{V}} \equiv \frac{1}{\|\vec{V}(\mathbf{q})\|}\left[\vec{V}(\mathbf{q})\ \mathbf{J} \cdot \vec{V}(\mathbf{q})\right], \tag{5}$$

where $\mathbf{J}$ is the counter-clockwise rotation by $90^\circ$.

Since the gradient of the signal $\phi$ is covariant, it provides a natural reference for defining $h$ and $\mathfrak{R}$. As in [MBK*20], we define $h : S \to \mathbb{R}$ to be the magnitude and set $\mathfrak{R}$ to be the frame field defined by the gradient of $\phi$:

$$h(\mathbf{q}) \equiv \|\nabla\phi(\mathbf{q})\| \qquad \text{and} \qquad \mathfrak{R}_{\mathbf{q}} \equiv \mathfrak{R}_{\mathbf{q}}^{\nabla\phi}. \tag{6}$$

The choice of $h$ ensures that at points where the gradient vanishes there is no contribution to the extended convolution, avoiding problems resulting from the singularity in the definition of $\mathfrak{R}^{\vec{V}}$.

Choosing $\phi$ to be an intrinsically defined function, the frame field $\mathfrak{R}$ and signal $h$ will be as well. Since the logarithm map is also intrinsic, it follows from Equation (4) that if the signal $\phi$ is intrinsically defined, the optimal filter will be invariant under global isometries. An exposition of the properties $h$ and $\mathfrak{R}$ must satisfy to ensure *local* isometry invariance can be found in the appendix.

### 5.3. Choosing $\phi$

To define the extended convolution, we need to select a function $\phi$ whose gradient and gradient magnitudes will define the frame field and signal. We would like to choose $\phi$ so that the derived descriptor is both stable and distinctive.

In our applications, we have found that the Heat Kernel Signature (HKS) [SOG09] provides a good balance between responsiveness and robustness; it captures subtle changes in the surface and is insensitive to sources of interference commonly found in mesh representations of surfaces such as noise and tessellation quality. In addition, the HKS provides a description of the intrinsic properties of the surface, ensuring that the proposed descriptor is invariant under isometric deformations of the surface.

### 5.4. Expression of the keypoint in the frames of its neighbours

The major computational step in the construction of our proposed surface descriptor is evaluating the logarithm, giving an expression of the keypoint as a vector in the tangent spaces of its neighbours. State-of-the-art algorithms for computing the logarithm

map parameterize the region about a given point through approaches based either on heat diffusion [SSC19, HA19] or on Dijkstra-like traversal [MR12]. A naive incorporation of one of these methods into our descriptor would entail computing a parameterization about every point in the support region, which is obviously undesirable.

To avoid this, we follow [SGW06, Rus10, HA19] and exploit a convenient relationship between the gradient of the geodesic distance function and the logarithm map. Setting $d_g^{\mathbf{p}}(\mathbf{q}) \equiv d_g(\mathbf{p}, \mathbf{q})$ to be the geodesic distance from $\mathbf{p}$ and using the symmetry of geodesic distances, the logarithm at $\mathbf{q}$ and geodesic distance from $\mathbf{p}$ are related by

$$\log_{\mathbf{q}}(\mathbf{p}) = -d_g^{\mathbf{p}}(\mathbf{q}) \cdot \left(\nabla d_g^{\mathbf{p}}(\mathbf{q})\right) = -d_g^{\mathbf{p}}(\mathbf{q}) \cdot \left(\frac{\nabla d_g^{\mathbf{p}}(\mathbf{q})}{\|\nabla d_g^{\mathbf{p}}(\mathbf{q})\|}\right), \quad (7)$$

where the last equation follows from the fact that the the distance function $d_g^{\mathbf{p}}$ satisfies the Eikonal equation. Thus we can compute a single (local) geodesic distance function at $\mathbf{p}$ and use the distance function $d_g^{\mathbf{p}}$ and its gradient to determine the logarithm of $\mathbf{p}$ in the tangent spaces of all neighbours $\mathbf{q}$.

More generally, letting $d : S \times S \to \mathbb{R}_{\geq 0}$ denote any distance function on $S$, and setting $d^{\mathbf{p}}(\mathbf{q}) \equiv d(\mathbf{p}, \mathbf{q})$, Equation (7) can be used to compute coordinate function giving the *position of* $\mathbf{p}$ *in the coordinate frame of* $\mathbf{q}$ as:

$$C^{\mathbf{p}}(\mathbf{q}) = -d^{\mathbf{p}}(\mathbf{q}) \cdot \mathfrak{R}_{\mathbf{q}}^{-1} \cdot \left(\frac{\nabla d^{\mathbf{p}}(\mathbf{q})}{\|\nabla d^{\mathbf{p}}(\mathbf{q})\|}\right). \quad (8)$$

This gives the feature descriptor a remarkable degree of flexibility in that the distance function can be treated as an input parameter that is chosen based on its suitability for the desired application.

## 6. The ECHO Surface Descriptor

The construction of the ECHO surface descriptor at a keypoint $\mathbf{p} \in S$ follows from the discretization of the optimal filter $f^{\mathbf{p}}$, as defined in Equation (4), relative to the signal and frame field given in Equation (6).

### 6.1. Discretization

We address the computation of the ECHO descriptor with respect to a discrete representation of surfaces as triangle meshes, $S = (\mathcal{V}, \mathcal{T})$. As is standard, a signal $\mathbf{f} : \mathcal{V} \to \mathbb{R}$ is represented by its values at the vertices and is extended by linear interpolation to the interior of triangles. Vector fields and frame fields are represented using a constant value per triangle. We represent a tangent vector at triangle $\mathbf{t} = (\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2) \in \mathcal{T}$ intrinsically using a pair of values, $\alpha$ and $\beta$, corresponding to the tangent vector $\alpha \cdot (\mathbf{v}_1 - \mathbf{v}_0) + \beta \cdot (\mathbf{v}_2 - \mathbf{v}_0)$. A frame field is then represented by the map $\mathbf{R} : \mathcal{T} \to \mathbb{R}^{2\times 2}$, with $\mathbf{R}_{\mathbf{t}}$ an orthogonal transformation from $\mathbb{R}^2$ (with the standard Euclidean inner-product) into the tangent space at triangle $\mathbf{t} \in \mathcal{T}$ (with inner-product given by the Riemannian metric).

### 6.2. Implementation

Given a vertex $\mathbf{p} \in \mathcal{V}$, we would like to compute the discretized descriptor $\mathbf{f}^{\mathbf{p}} \in \mathbb{R}^{(2n+1)\times(2n+1)}$, sampled on a $(2n+1) \times (2n+1)$ grid. To do this, we need to compute the discretized coordinate function $\mathbf{C}^{\mathbf{p}} : \mathcal{V} \to \mathbb{R}^2$ and discretized signal $\mathbf{h} : \mathcal{V} \to \mathbb{R}_{\geq 0}$, and we need to estimate the integral in Equation (4).

#### Defining the coordinate function and signal

To define a discretized coordinate function and signal, we need to represent $\mathbf{C}^{\mathbf{p}}$ and $\mathbf{h}$ by their values at vertices. As both functions are defined in terms of the gradients of scalar function, which are represented as constant values per triangle, we use area-weighted averaging to define the values at the vertices.

Specifically, letting $\mathcal{T}_{\mathbf{q}} = \{\mathbf{t} \in \mathcal{T} \mid \mathbf{t} \ni \mathbf{q}\}$ denote the subset of triangles incident on $\mathbf{q} \in \mathcal{V}$, we define the coordinates of the position of vertex $\mathbf{p}$ in the tangent space of vertex $\mathbf{q}$ (relative to the prescribed fame field) as:

$$\mathbf{C}^{\mathbf{p}}(\mathbf{q}) = -\mathbf{d}^{\mathbf{p}}(\mathbf{q}) \cdot \frac{\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{q}}} |\mathbf{t}| \cdot \mathbf{R}_{\mathbf{t}}^{-1} \cdot \left(\frac{\nabla \mathbf{d}^{\mathbf{p}}(\mathbf{t})}{\|\nabla \mathbf{d}^{\mathbf{p}}(\mathbf{t})\|}\right)}{\left\|\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{q}}} |\mathbf{t}| \cdot \mathbf{R}_{\mathbf{t}}^{-1} \cdot \left(\frac{\nabla \mathbf{d}^{\mathbf{p}}(\mathbf{t})}{\|\nabla \mathbf{d}^{\mathbf{p}}(\mathbf{t})\|}\right)\right\|}, \quad (9)$$

where $|\mathbf{t}|$ is the area of triangle $\mathbf{t}$.

Similarly, we define the vertex-based signal $\mathbf{h} : \mathcal{V} \to \mathbb{R}_{\geq 0}$ by taking the area-weighted average of the magnitudes of tangent vectors adjacent to a vertex $\mathbf{q}$:

$$\mathbf{h}(\mathbf{q}) = \frac{\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{q}}} |\mathbf{t}| \cdot \|\nabla\phi(\mathbf{t})\|}{\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{q}}} |\mathbf{t}|}. \quad (10)$$

#### Estimating the integral

Discretizing the integral in Equation (4) requires making two approximations. First, since we will be sampling at discrete positions on a regular grid, we replace the delta function with a wider kernel. Second, as the integral does not have a closed-form expression, we approximate it using discrete sampling.

To replace the delta function, we use a compactly supported kernel that approximates a Gaussian with deviation $\boldsymbol{\sigma}$:

$$k_{\mathbf{x},\sigma}(\mathbf{y}) = \begin{cases} \exp(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{\sigma^2}) & \text{if } \|\mathbf{x}-\mathbf{y}\| \leq 2\sigma \\ 0 & \text{otherwise} \end{cases}.$$

Then, we approximate the integral using $m$-th degree Gaussian quadrature samples [Cow73]. Specifically, setting $\mathcal{Q}_{\mathbf{t}}^m \subset \mathbf{t} \times \mathbb{R}$ to be the (finite) set of quadrature points and quadrature weights of degree $m$, we approximate the integral by first writing it out as a sum of integrals over the individual triangles of the mesh, and then approximating each per-triangle integral by a weighted summation over the quadrature samples. Combining the approximations gives:

$$\begin{aligned}
\mathbf{f}^{\mathbf{p}}(\mathbf{x}) &\approx \int_{\{\mathbf{q}\in S \mid d(\mathbf{p},\,\mathbf{q})\leq\varepsilon\}} \mathbf{h}(\mathbf{q}) \cdot k_{\mathbf{x},\sigma}\left(\mathbf{C}^{\mathbf{p}}(\mathbf{q})\right) d\mathbf{q} \\
&= \sum_{\mathbf{t}\in\mathcal{T}} \int_{\{\mathbf{q}\in\mathbf{t} \mid d(\mathbf{p},\,\mathbf{q})\leq\varepsilon\}} \mathbf{h}(\mathbf{q}) \cdot k_{\mathbf{x},\sigma}\left(\mathbf{C}^{\mathbf{p}}(\mathbf{q})\right) d\mathbf{q} \\
&\approx \sum_{\mathbf{t}\in\mathcal{T}} \sum_{\{(\mathbf{q},\,w_{\mathbf{q}})\in\mathcal{Q}_{\mathbf{t}}^m \mid \mathbf{d}(\mathbf{p},\,\mathbf{q})\leq\varepsilon\}} \mathbf{h}(\mathbf{q}) \cdot\ w_{\mathbf{q}} \cdot k_{\mathbf{x},\sigma}\left(\mathbf{C}^{\mathbf{p}}(\mathbf{q})\right) d\mathbf{q}.
\end{aligned}$$

**Algorithm 1.** ECHO Surface Descriptor

**Input:**
triangle mesh $S = (\mathcal{V}, \mathcal{T})$,
keypoint $\mathbf{p} \in \mathcal{V}$,
coordinate function $\mathbf{C}^{\mathbf{p}} : \mathcal{V} \to \mathbb{R}^2$,
signal $\mathbf{h} : \mathcal{V} \to \mathbb{R}_{\geq 0}$,
distance map $\mathbf{d}^{\mathbf{p}} : \mathcal{V} \to \mathbb{R}_{\geq 0}$,
support radius $\varepsilon \in \mathbb{R}_{>0}$,
descriptor radius $n \in \mathbb{Z}_{>0}$,
Gaussian variance $\sigma \in \mathbb{R}_{>0}$,
quadrature degree $m \in \mathbb{Z}_{>0}$
**Output:**
ECHO descriptor $\mathbf{f}^{\mathbf{p}} \in \mathbb{R}^{(2n+1)\times(2n+1)}$

$\alpha = \varepsilon \,/\, n$ ▷ Surface to descriptor scale

$\mathbf{f}^{\mathbf{p}} = \mathbf{0}^{(2n+1)\times(2n+1)}$ ▷ Initialize descriptor

$\forall\, \mathbf{t} = (\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2) \in \mathcal{T}$ such that $\exists\, i$ with $\mathbf{d}^{\mathbf{p}}(\mathbf{v}_i) \leq \varepsilon$

$\mathcal{Q}^m_{\mathbf{t}} = \text{QuadratureSamples}(m,\ \mathbf{t}) \subset \mathbf{t} \times \mathbb{R}$

$\forall\, (\mathbf{q},\ w_{\mathbf{q}}) \in \mathcal{Q}^m_{\mathbf{t}}$ ▷ Integrate over the triangle

$\mathbf{x}_{\mathbf{p}} = \alpha \cdot \mathbf{C}^{\mathbf{p}}(\mathbf{q})$ ▷ Position of $\mathbf{p}$ in the frame of $\mathbf{q}$

$\mathcal{B}_{\mathbf{q}} = \{\mathbf{x} \in \mathbb{Z}^2 \ \mid\ \|\mathbf{x}\| \leq n \text{ and } \|\mathbf{x} - \mathbf{x}_{\mathbf{p}}\| \leq 2\sigma\}$

$\forall\, \mathbf{x} \in \mathcal{B}_{\mathbf{q}}$ ▷ Splat signal into vicinity of $\mathbf{x}_{\mathbf{p}}$

$\mathbf{f}^{\mathbf{p}}(\mathbf{x}) \mathrel{+}= \mathbf{h}(\mathbf{q}) \cdot w_{\mathbf{q}} \cdot k_{\mathbf{x},\sigma}(\mathbf{x}_{\mathbf{p}})$

Pseudocode for the computation of the discrete descriptor is shown in **Algorithm** 1: we iterate over all triangles $\mathbf{t} \in \mathcal{T}$ which have at least one vertex within a distance of $\varepsilon$ of the keypoint; for each triangle $\mathbf{t}$, We compute the set of $m$-th degree quadrature samples $\mathcal{Q}^m_{\mathbf{t}}$; for each quadrature point $\mathbf{q}$, we compute $\mathbf{x}_{\mathbf{p}}$, the position of $\mathbf{p}$ in the coordinate frame of $\mathbf{q}$, scaled to the descriptor resolution; we also compute $\mathcal{B}_{\mathbf{q}}$ – the set of grid points that fall within a disk of radius $n$ of the origin and whose kernel function supports $\mathbf{x}_{\mathbf{p}}$; for each grid point $\mathbf{x} \in \mathcal{B}_{\mathbf{q}}$, we increment the value of the descriptor at $\mathbf{x}$ by the value of the signal at the quadrature point, $\mathbf{h}(\mathbf{q})$, weighted by the quadrature weight, $w_{\mathbf{q}}$, and the value of the kernel function centred at $\mathbf{x}$, evaluated at the position of $\mathbf{p}$ in the coordinate frame of quadrature point, $k_{\mathbf{x},\sigma}(\mathbf{x}_{\mathbf{p}})$.

### Defining the distance

In our implementation, we consider three different distance functions: The geodesic distance, the diffusion distance [CL06], and the biharmonic distance [LRF10].

To compute the distance map $\mathbf{d}^{\mathbf{p}}$ it is desirable to choose a method that allows the calculation to be truncated to exclude points whose distance from $\mathbf{p}$ exceeds $\varepsilon$. Assuming that the set of points in $S$ within a distance of $\varepsilon$ of the keypoint $\mathbf{p}$ is connected, the truncated distance function can be computed using a flood-fill approach.

For the the diffusion distance and biharmonic distance, the implementation is straight-forward. (We reuse the spectrum computed for the Heat Kernel Signature.) For the geodesic distance, we use the Dijkstra-like implementation in [MR12] originally proposed in [Rei04]. The authors' publicly available implementation provides a fast and accurate approximation of geodesic distances inside a local neighbourhood. An example of the computed geodesic distances and derived logarithms is shown on the right in Figure 1.

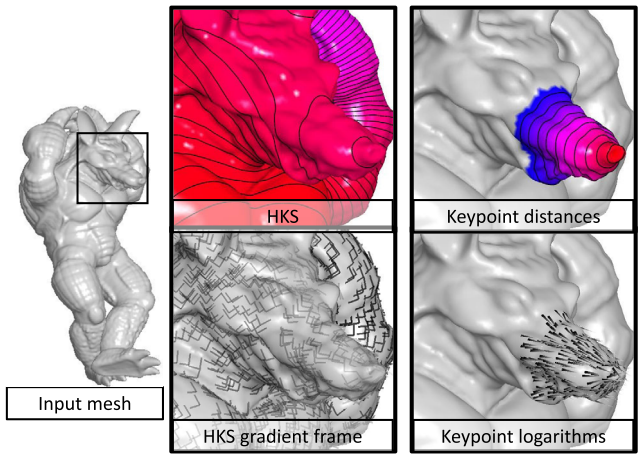



**Figure 1:** *Visualizations of the input geometry (left), the Heat Kernel Signature and derived frame field (middle), and the geodesic distances from the keypoint and the keypoint's logarithm in its neighbours' tangent frames (right). For the visualization of the signals, red corresponds to lower values and blue to larger ones. Frames defined by the gradients of the HKS are visualized by showing the directions of the positive x- and y-axes. Frames generated from larger magnitude gradients are rendered with higher opacity.*

## 7. Evaluation

We compare ECHO against popular surface feature descriptors in the context of feature matching and sparse shape correspondence. Descriptors are evaluated in terms of overall descriptiveness and robustness to rigid articulations, near isometric and non-isometric deformations, Gaussian noise, varying mesh tessellation, and topological and geometric changes. We consider the performance of ECHO – using geodesic, biharmonic [LRF10], and diffusion [CL06] distances – and several descriptors introduced in the last decade: SHOT [TSDS10b, STDS14], RoPS [GSB*13], USC [TSDS10a], and ISC [KBLB12]. The first three have been shown to be among the most effective descriptors currently in the literature [GBS*16]. Of these, SHOT has seen wide adoption in the context of 3D object retrieval, recognition and correspondence [VLB*17, MBM*17, BBL*17]. We include ISC in our comparisons as unlike SHOT, RoPS, and USC, it, like ECHO, is intrinsic. Moreover, it is similar to ECHO in that votes are weighted and binned with respect to

the HKS and geodesic distances, though it does not incorporate the use of frame fields.

To perform the evaluations, we use two datasets consisting of triangular meshes: the TOSCA dataset [BBK08] and the SHREC 2019 Shape Correspondence with Isometric and Non-Isometric Deformations benchmark dataset [DSL*19]. The former consists of a set of collections of synthetic humanoid and animal figures in different, near-isometric poses; the latter is made up of 3D scans of real-world objects exhibiting a wide variety of deformations.

We use the publicly available implementations of SHOT, RoPS, and USC in the PCL library [RC11] with the default parameters. For ISC, we use our own C++ implementation based on the Matlab implementation made available by the authors. We attempt to remain as faithful as possible to the authors' original implementation, though we replace their method for computing geodesics with that used in ECHO as discussed in Section 6.

The size of the support radius $\varepsilon$ depends on the mesh, and is proportional for all descriptors. Specifically, we follow [ZBH12, GBS*16] and set the support radius for all keypoints to be

$$\varepsilon = 0.08\sqrt{A\,/\,\pi}\,, \tag{11}$$

where $A$ is the area of the mesh. For the biharmonic and diffusion ECHO descriptors, $A$ is computed by first using the distance function to assign lengths to edges and then using Heron's Formula to compute the triangle areas from these lengths. All support regions contain a similar number of vertices.

ECHO descriptors are un-normalized, computed with a descriptor radius of $n = 5$, a Gaussian deviation of $\sigma = 1.3\,/\sqrt{-\log(0.05)}$, and a quadrature degree of $m = 5$ (corresponding to 7 samples within each triangle) [Cow73]. The HKS is computed once for each mesh as a pre-processing step using the first 200 eigenvalue-vector pairs of the Laplace-Beltrami operator and with a diffusion time of 0.1. The spectral decomposition is reused in the calculation of biharmonic and diffusion distances, the latter of which is computed with a diffusion time of 0.1. (Models are rescaled to have unit area prior to computing the spectrum so that the parameters used for computing the HKS and the diffusion distances are consistent with respect to scaling.) We use the same HKS calculation for both ECHO and ISC, as we find that the latter sees better performance using our selected parameters than those suggested by the authors in [BK10]. An example HKS and the corresponding frame field are shown in Figure 1 (middle).

### 7.1. Feature matching

The TOSCA dataset is used to evaluate the descriptors in terms of overall descriptiveness and robustness to increasing levels of Gaussian noise and mesh decimation. Experiments are performed by evaluating the performance of the descriptors in matching features from a set of *scene* meshes to those from a smaller set of *models*.

The *model* meshes consist of the nine 'null' meshes from each shape class (those numbered `0` and the `gorilla1` mesh). All other meshes in the dataset constitute the *scenes*, which are identical to the models up to near-isometric deformations and share the same triangulations. Examples of biharmonic ECHO descriptors computed at corresponding points on two scene meshes from the `centaur` class are shown in Figure 2.

To avoid the influence of keypoint detection algorithms in our experiments, we randomly generate corresponding points for both the models and the scenes in the following manner: First, we randomly select 1000 keypoints lying on each model mesh. Then, for each scene, we randomly select 1000 points that match at least one keypoint in the corresponding model. A keypoint on a scene is considered to match a point on a model if the two belong to the same class and if, after mapping the scene keypoint to the model, the geodesic distance between the two is less than the $1/4$ the support radius in Equation (11). Then, for each method, descriptors are computed at all model and scene keypoints.

#### Descriptiveness

To evaluate descriptiveness, we compute precision-recall curves for each descriptor at every scene keypoint, an approach that has been demonstrated to be well-suited to this task [KS04, MS05]. Given a scene keypoint, $\mathbf{p}$ and corresponding descriptor $\mathbf{f}^{\mathbf{p}}$, all keypoints from across all models are sorted based on the descriptor distance, giving $\{\mathbf{q}_1, \ldots, \mathbf{q}_M\}$ with

$$\left\|\mathbf{f}^{\mathbf{p}} - \mathbf{f}^{\mathbf{q}_i}\right\| \leq \left\|\mathbf{f}^{\mathbf{p}} - \mathbf{f}^{\mathbf{q}_{i+1}}\right\|.$$

We define $\mathcal{M}_{\mathbf{p}}$ to be the set of all model keypoints that are valid matches with $\mathbf{p}$. (Specifically, the points in $\mathcal{M}_{\mathbf{p}}$ belong to the same class as $\mathbf{p}$.) Following [SMKF04], the precision $P_{\mathbf{p}}$ and recall $R_{\mathbf{p}}$ assigned to $\mathbf{p}$ are defined as functions of the top $r$ model keypoints,

$$P_{\mathbf{p}}(r) = \frac{\left|\mathcal{M}_{\mathbf{p}} \cap \{\mathbf{q}_i\}_{i\leq r}\right|}{r} \quad \text{and} \quad R_{\mathbf{p}}(r) = \frac{\left|\mathcal{M}_{\mathbf{p}} \cap \{\mathbf{q}_i\}_{i\leq r}\right|}{\left|\mathcal{M}_{\mathbf{p}}\right|}. \tag{12}$$

**Note.** Our definition of a "match" is conservative in that it can exclude valid correspondences such as matching the right index finger from the `michael` model with the right index finger from a `victoria` scene.) This has the effect of creating more false negatives, which reduces the overall precision.

#### Robustness to noise

We test the robustness of each method with respect to various levels of Gaussian noise applied to all meshes in the scenes. Similar to [BBC*10, BBB*11, GBS*16], we add five levels of Gaussian noise with variances of $\sqrt{i\cdot\varepsilon\,/\,200}$, $1 \leq i \leq 5$ to the vertices of each mesh in the scenes. However, we scale the magnitude of the noise relative to local edge lengths. Namely, the noise added at a vertex $\mathbf{p}$ is scaled by a factor of $E_{\mathbf{p}}\,/\,E$ where $E_{\mathbf{p}}$ is the average length of the edges incident on $\mathbf{p}$ and $E$ is the average length of all edges in the mesh. This process produces five sets of scenes corresponding to each level of noise; the models are left unchanged. (See supplementary material for examples.)

#### Robustness to varying mesh Resolution

We also test the sensitivity of each descriptor with respect to changes in mesh resolution. Three new sets of scenes are constructed by decimating the original scene meshes by factors of 2, 4, and 8.

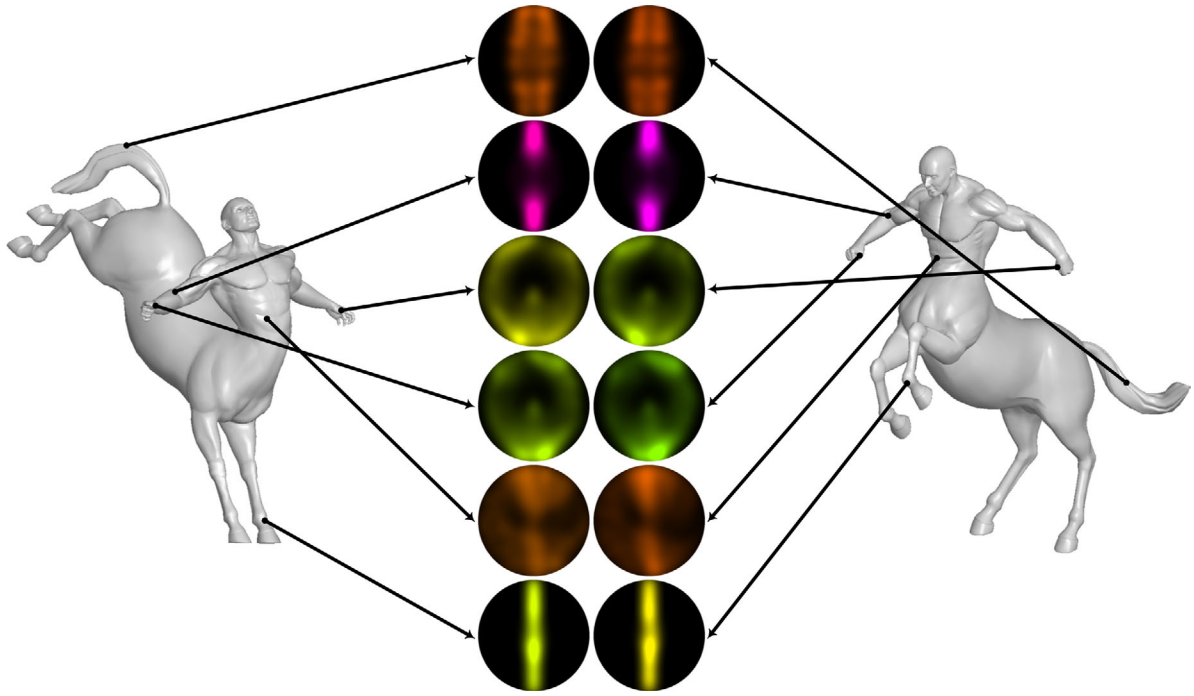


**Figure 2:** *Visualizations of biharmonic ECHO descriptors computed at corresponding points on two meshes from the* `centaur` *class of the TOSCA dataset [BBK08]. ECHO's intrinsic construction and use of frame fields enables rich and distinctive characterizations that remain consistent in the presence of significant local deformations. Descriptors are drawn using the HSV scale – hue encodes the absolute magnitude (ranging from smaller descriptors drawn in red to larger descriptors drawn in blue) and value encodes the relative magnitude (darker colors correspond to smaller descriptor values). Saturation is fixed at one.*

Specifically, we use OpenFlipper's [MK10] incremental mesh decimation module with the decimation priority determined by the distance to the original mesh. (See supplementary material for examples.)

### 7.2. Sparse correspondences

We use the SHREC 2019 Shape Correspondence Benchmark dataset to evaluate the quality of each descriptor under rigid articulations, near-isometric and non-isometric deformations, and topological and geometric changes in a sparse shape correspondence regime. The dataset consists of fifty meshes constructed from 3D scans of real-world objects; as a byproduct, the real-world scans contain noise, varying trianglulations, occluded geometry and various other sources of interference [DSL*19]. The dataset contains 76 pre-defined pairs of meshes partitioned into four increasingly challenging test sets: (1) articulations and rigid deformations, (2) near-isometric deformations, (3) non-isometric deformations and (4) topological and geometric changes. The authors provided the dense ground-truth correspondences associated with each pair.

Experiments are performed such that each type of descriptor is used to compute sparse correspondences between the two meshes in each pair in a naive correspondence regime. Specifically, each pair in is split into a *model* and a *scene* mesh, where the former is in a relatively simple 'null' pose and the latter is in a more complex pose. As in the feature matching experiments, 1000 keypoints lying on the scene mesh are randomly chosen. Descriptors are computed at all scene keypoints and at *every* vertex on the model mesh. For a given scene keypoint $\mathbf{p}$, we follow [CRB*16, LRB*16, DSL*19] and evaluate the correspondence quality by computing the (area-normalized) geodesic distance between the ground-truth position of the keypoint on the model mesh, $\mathbf{q}^*$, and the model vertex $\mathbf{q}$ with the smallest descriptor distance,

$$d_g(\mathbf{q}^*, \mathbf{q})\sqrt{\pi / A}, \tag{13}$$

where $A$ is the total surface area of the model mesh.

**Table 1:** *Mean descriptor run time over all model keypoints in the feature matching experiments. The 'null' meshes contain between 5000 and 53,000 vertices, with an average of approximately 30,000 vertices.*

| | | | | | ECHO | | |
|---|---|---|---|---|---|---|---|
| | SHOT | RoPS | USC | ISC | Biharmonic | Geodesic | Diffusion |
| (ms) | **20** | 138 | 108 | 55 | 118 | 58 | 132 |

### 7.3. Complexity

The mean run time for each descriptor over all model keypoints in the feature matching experiments is shown in Table 1. The SHOT descriptor is fastest, followed by the ISC descriptor and the ECHO descriptor computed using geodesic distances. While slower, we

believe that the computational overhead of biharmonic ECHO descriptor is justified by its effectiveness, as we discuss next.

## 8. Results and Discussion

Here we discuss the results of our evaluations using the TOSCA and SHREC 2019 Shape Correspondence Benchmark datasets.

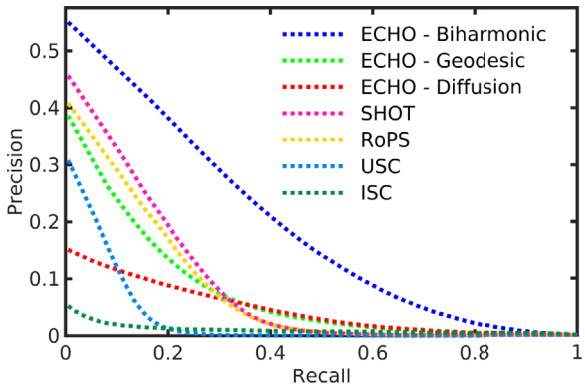
### 8.1. TOSCA

The results of the feature matching experiments using the TOSCA dataset are shown in Figure 3. The descriptiveness results are aggregated by computing the mean precision and recall across all keypoints in the scenes. The resulting curves for each descriptor are shown in Figure 3a.

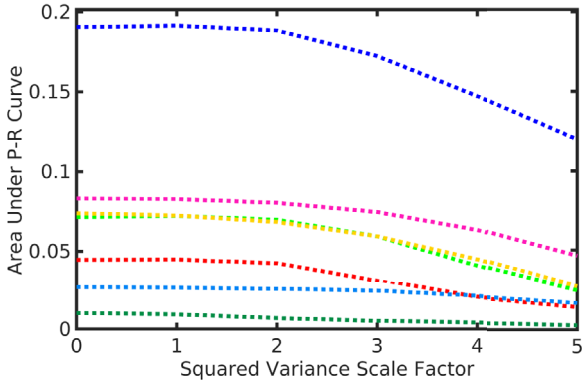
The biharmonic ECHO descriptor achieves the best performance by a significant margin, followed by SHOT and RoPS, though the difference between the latter is smaller. Generally speaking, ECHO, SHOT, RoPS and USC have approximately similar distinctive potential in the sense that all achieve rotation invariance without loss of information by incorporating frame fields in their construction. The superior performance of the ECHO descriptor is likely due to the stability of the biharmonic distance map and the fact that the descriptor is intrinsic and thus fully invariant to isometric deformations. Like ECHO, the ISC descriptor is also intrinsic, though its poor performance is likely a consequence of its lower descriptive ceiling as it independently discards per-frequency and per-radius phase information to achieve rotation invariance. While the mappings between the TOSCA models meshes and their corresponding meshes in the scenes are not perfect isometries, it is clear from the performance of ECHO that an intrinsic construction confers an advantage provided it can make use of a frame field. Results obtained using a normalized ECHO descriptor exhibit a similar trend, though performance was slightly worse compared to the un-normalized descriptors.

The feature matching robustness results are expressed by plotting the area under the mean precision-recall curves as a function of nuisance severity. Descriptor performance under increasing levels of Gaussian noise is shown in Figure 3b. The biharmonic ECHO descriptor achieves the best performance at all levels and remains stable relative to the other descriptors in the sense that it sees a proportional drop in performance at higher levels of noise. The geodesic and diffusion ECHO descriptors, SHOT, and RoPS all perform similarly and are less effective.

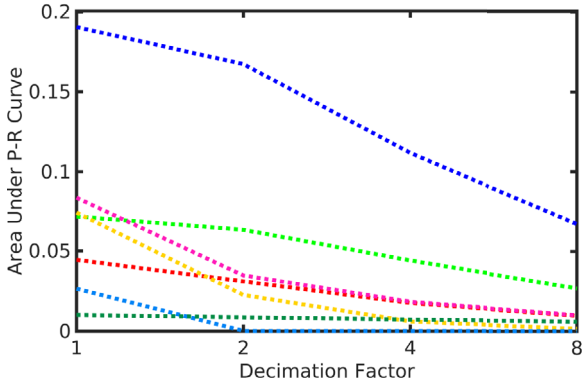
Descriptor matching performance relative to changes in mesh resolution is shown in Figure 3c, again using the area under the mean precision-recall curves. The biharmonic ECHO descriptor achieves the best performance, followed by the geodesic and diffusion ECHO descriptors. We are uncertain as to why ECHO is more stable when computed using geodesic distances than either biharmonic or diffusion distances.




(a) Overall descriptiveness under near-isometric deformations




(b) Robustness under increasing Gaussian noise

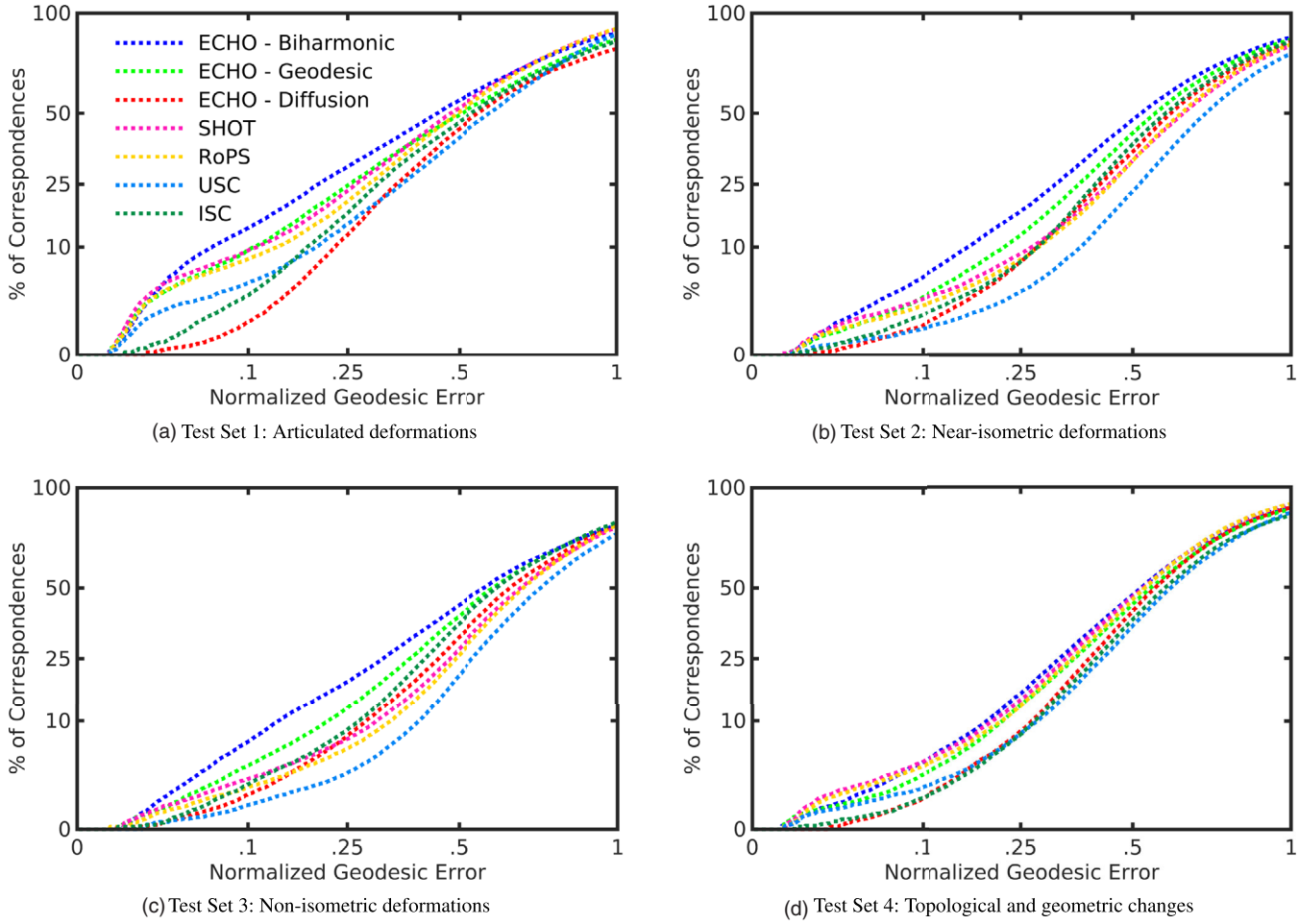



(c) Robustness under decreasing mesh resolution

**Figure 3:** *Results of feature matching evaluations using the TOSCA dataset. Top: Descriptiveness results in the form of the mean precision and recall curves for each descriptor. Middle and Bottom: Robustness results in the form of the areas under the mean precision-recall curves as functions of noise and decimation severity.*

### 8.2. SHREC '19 Shape Correspondence Benchmark

The results of the sparse correspondence experiments using the SHREC 2019 Shape Correspondence Benchmark dataset are shown in Figure 4. For each test set, the curve defined by plotting the percentage of the total number of correspondences for which the (normalized) geodesic distance between the model point with best-matching descriptor and the ground-truth model point is below a threshold value is used as an aggregate measure of descriptor correspondence quality. The resulting curves for test sets 1, 2, and 3,




(a) Test Set 1: Articulated deformations

(b) Test Set 2: Near-isometric deformations

(c) Test Set 3: Non-isometric deformations

(d) Test Set 4: Topological and geometric changes

**Figure 4:** *Results of the shape correspondence evaluations on each test set from the SHREC 2019 Shape Correspondence Benchmark [DSL*19]. For each descriptor, the percentage of total correspondences is expressed as a function of the normalized geodesic error. Both axes are plotted on a square root scale.*

which correspond to articulated, near-isometric, and non-isometric shape deformations, are shown in Figures 4a, 4b and 4c, plotted on a square root scale.

The biharmonic and geodesic ECHO descriptors achieve the best performance across the first three test sets, followed by SHOT. In particular, the differences in performance between the ECHO descriptors and the other methods on the second and third test sets, which concern near-isometric and non-isometric deformations, are especially significant. That ECHO sees little, if any, difference in performance between the two tests sets is unexpected. While ECHO is isometry invariant, it, like the other descriptors we consider, is not designed to be stable under non-isometric surfaces. Despite this, the results suggest that combing an intrinsic construction with a frame field can still be a powerful approach in the presence of more complex deformations. Examples of sparse correspondences found by the SHOT and biharmonic ECHO descriptors between mesh pairs from the first and third test sets are shown in Figure 5.

The error curves for test set 4, which considers topological and geometric changes, are shown in Figure 4d. SHOT and RoPS achieve a greater number of correspondences with lower errors, though the biharmonic ECHO descriptor begins to see more correspondences as the error increases. Here, the relatively poor performance of the ECHO descriptors might be explained by its use of the HKS, which has demonstrated instability under topological change [DSL*19].

Among other sources of interference we do not explicitly consider are matching and correspondence in the presence of partial shape data. It is not immediately obvious that the ECHO descriptor would struggle to a greater extent than popular extrinsic descriptors like SHOT and RoPS in the presence of occlusions or incomplete shapes. Regardless, we believe that our evaluations demonstrate that the ECHO descriptor is more informative than state-of-the-art methods and remains so under significant noise, changes in mesh resolution, complex deformations, and in the presence

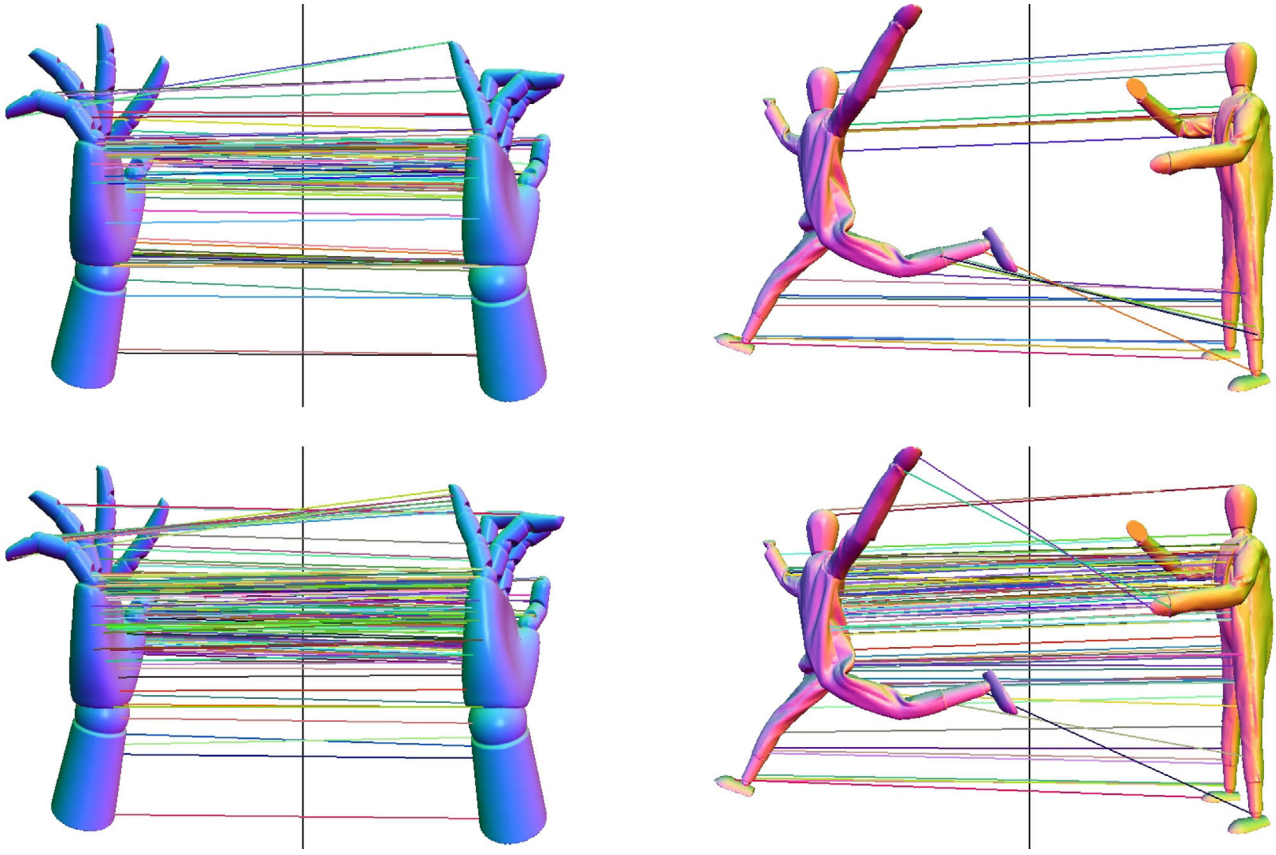


**Figure 5:** *Visualizations of sparse correspondences with a normalized geodesic error* $\leq 0.08$ *found by the SHOT (top row) and biharmonic ECHO (bottom row) descriptors between two example mesh pairs from the first (left column) and third (right column) test sets in the SHREC 2019 Shape Correspondence Benchmark dataset. Both SHOT and ECHO perform well in finding correspondences between meshes differing by locally rigid articulations (left). However, SHOT's performance sharply deteriorates in the presence of complex, non-isometric deformations (right), while ECHO remains relatively stable.*

of a variety of challenging nuisance factors commonly found in real data.

### 8.3. Discussion

The local shape descriptor we compute at the point $\mathbf{p}$ can be viewed as a specific instance of a more general family of descriptors. To this end consider two intrinsic functions on a surface, $d^{\mathbf{p}}$ and $h$, and an intrinsic vector field $\vec{V}$. Here $d^{\mathbf{p}}$ should depend on the point $\mathbf{p}$ being described. We define a histogram characterizing the point $\mathbf{p}$ by aggregating information from neighbouring points $\mathbf{q}$. Each point $\mathbf{q}$ contributes a vote with weight $h(\mathbf{q})$ into the bin describing the position of $\mathbf{p}$ relative to $\mathbf{q}$. Expressing the position in polar coordinates, the radius is given by $d^{\mathbf{p}}(\mathbf{q})$ and the angle is defined to be the angle between the tangent vectors $\nabla d^{\mathbf{p}}(\mathbf{q})$ and $\vec{V}(\mathbf{q})$.

The paper considers several choices of $d^{\mathbf{p}}$ including geodesic, diffusion, and biharmonic distances from $\mathbf{p}$. It takes $h = \phi$ and $\vec{V} = \nabla\phi$, where $\phi$ is the Heat Kernel Signature. It is obvious that this descriptor is isometry invariant because all of the involved functions and operators are intrinsic. In addition, the improved performance with the use of the biharmonic distance becomes clear as the biharmonic distance is more stable in the presence of noise than the geodesic distance. The motivation for using the HKS is also exposed. The feature points of a shape are usually local extrema of the HKS, leading to an anisotropic distribution of weights in the histogram which produces a more discriminating characterization. (Alternative choices for $\phi$ could include locally averaged Gaussian curvature, which is qualitatively similar to the HKS for small time-steps but does not require a spectral decomposition, or the Average Geodesic Distance Function [ZMT05].)

One could, of course, consider other choices of $d^{\mathbf{p}}$, $h$, and $\vec{V}$ so long as the derived polar coordinates are well-defined. In particular, we require that $h(\mathbf{q}) = 0$ whenever $\mathbf{q} \neq \mathbf{p}$ and either $\nabla d^{\mathbf{p}}$ or $\vec{V}$ vanishes, so that the descriptor remains well-defined even when the angular component of the polar coordinates of $\mathbf{p}$ relative to $\mathbf{q}$ is not.

Finally, we note that this work focuses on the evaluation of the ECHO descriptor as a stand-alone characterization of local geometry. A natural extension of this work is to incorporate the ECHO descriptor within a non-rigid registration pipeline, akin to the way in which SHOT is used to either initialize [LYLG18, DLRT19] or regularize [VLB*17] the registration process.

## 9. Conclusion

In this paper we propose a surface descriptor. We generalize the concept of *extended convolution* [MBK*20] to surfaces, which allows a filter to transform as it travels over the surface, and show that the ECHO descriptors are optimal relative to this framework. We evaluate the performance of our proposed descriptor against that of premier surface descriptors. Using biharmonic distances, the ECHO surface descriptor significantly outperforms the SHOT, RoPS, USC and ISC descriptors in terms of overall descriptiveness and remains more distinctive under significant levels of Gaussian noise, changes in tessellation quality, and complex deformations.

In the future, we would like to extend the ECHO descriptor to include invariance to local conformal transformations by baking scale invariance into the construction in a similar manner as local rotation invariance. To this point, we are also interested in designing a surface keypoint detection algorithm as a companion to ECHO with the goal of significantly improving its performance. In particular, the extended convolution on surfaces provides a natural framework for such an endeavor.

## Acknowledgements

We thank the authors of [BBK08] and [DSL*19] for making available the TOSCA and SHREC 2019 Shape Correspondence Benchmark datasets; a special thanks is offered to Roberto Dyke, Drs. Paul L. Rosin and Yu-Kun Lai for providing us with the ground-truth correspondences for the latter. We would also like to thank Drs. Phillip Herholz, Eivind Lyche Melvær and Martin Reimers for providing us with their respective code for geodesic maps. Finally, the authors would like to thank reviewers for their invaluable recommendations for improving the paper. This work was performed under National Science Foundation grant IIS-1619050 and Office of Naval Research Award N00014-17-1-2142.

## Appendix A: Proof of Local Isometry Invariance

As discussed in Section 5.2, invariance of the optimal filter $f^{\mathbf{p}}$ in Equation (4) under intrinsic isometries of the local surface depends on the choice of signal $h$ and frame field $\mathfrak{R}$.

Suppose the surface $S$ contains two features, centered at $\mathbf{p}_0$ and $\mathbf{p}_1$ that are locally isometric. That is, there exists a map $\Psi : S \to S$ taking $\mathbf{p}_0$ to $\mathbf{p}_1$ and satisfying:

$$d_g(\mathbf{q}_0,\ \mathbf{q}_1) = d_g(\Psi(\mathbf{q}_0),\ \Psi(\mathbf{q}_1)), \quad \forall\ d_g(\mathbf{p}_0,\ \mathbf{q}_i) \leq \varepsilon.$$

It follows that for all $\mathbf{q} \in S$ with $d_g(\mathbf{p}_0,\ \mathbf{q}) \leq \varepsilon$,

$$d\Psi|_{\mathbf{q}} \cdot \log_{\mathbf{q}}(\mathbf{p}_0) = \log_{\Psi(\mathbf{q})}(\mathbf{p}_1). \tag{A.1}$$

To ensure local isometry invariance, we need to define $h$ and $\mathfrak{R}$ so that the filters $f^{\mathbf{p}_0}$ and $f^{\mathbf{p}_1}$ defined about $\mathbf{p}_0$ and $\mathbf{p}_1$ are equal. In particular, we have equality if, for all $\mathbf{q} \in S$ satisfying $d_g(\mathbf{p}_0,\ \mathbf{q}) \leq \varepsilon$,

$$h(\mathbf{q}) = h(\Psi(\mathbf{q})) \quad \text{and} \quad \mathfrak{R}_{\mathbf{q}} = \left[d\Psi|_{\mathbf{q}}\right]^{-1} \cdot \mathfrak{R}_{\Psi(\mathbf{q})}. \tag{A.2}$$

To see this, we expand the expression for $f^{\mathbf{p}_0}$, getting:

$$\begin{aligned}
f^{\mathbf{p}_0} &\overset{(4)}{=} \int_{\{\mathbf{q}\in S \,|\, d_g(\mathbf{p}_0,\,\mathbf{q})\leq\varepsilon\}} h(\mathbf{q}) \cdot \delta_{\mathbf{x}}\left(\mathfrak{R}_{\mathbf{q}}^{-1} \cdot \log_{\mathbf{q}}(\mathbf{p}_0)\right) d\mathbf{q} \\
&\overset{(A.2)}{=} \int_{\{\mathbf{q}\in S \,|\, d_g(\mathbf{p}_0,\,\mathbf{q})\leq\varepsilon\}} h(\Psi(\mathbf{q})) \cdot \delta_{\mathbf{x}}\left(\mathfrak{R}_{\Psi(\mathbf{q})}^{-1} \cdot d\Psi|_{\mathbf{q}} \cdot \log_{\mathbf{q}}(\mathbf{p}_0)\right) d\mathbf{q} \\
&\overset{(A.1)}{=} \int_{\{\mathbf{q}\in S \,|\, d_g(\mathbf{p}_0,\,\mathbf{q})\leq\varepsilon\}} h(\Psi(\mathbf{q})) \cdot \delta_{\mathbf{x}}\left(\mathfrak{R}_{\Psi(\mathbf{q})}^{-1} \cdot \log_{\Psi(\mathbf{q})}(\mathbf{p}_1)\right) d\mathbf{q} \\
&= \int_{\{\mathbf{q}\in S \,|\, d_g(\mathbf{p}_1,\,\mathbf{q})\leq\varepsilon\}} h(\mathbf{q}) \cdot \delta_{\mathbf{x}}\left(\mathfrak{R}_{\mathbf{q}}^{-1} \cdot \log_{\mathbf{q}}(\mathbf{p}_1)\right) d\mathbf{q} \\
&\overset{(4)}{=} f^{\mathbf{p}_1}.
\end{aligned}$$

where the second to last equality follows from the chain rule and the fact that $d\Psi$ is an orthogonal transformation. Thus, $f^{\mathbf{p}_0}$ and $f^{\mathbf{p}_1}$ are equal.

Furthermore, it is not hard to show that if the signal $\phi : S \to \mathbb{R}$ defined on the surface is such that for all $\mathbf{q} \in S$ with $d_g(\mathbf{p}_0,\ \mathbf{q}) \leq \varepsilon$,

$$\phi(\mathbf{q}) = \phi(\Psi(\mathbf{q})),$$

*i.e.* it is locally invariant to the isometry $\Psi$, then $h$ and $\mathfrak{R}$ as defined in Equation (6) satisfy the conditions in Equation (A.2). This follows from taking the gradient of both sides:

$$\begin{aligned}
\nabla\phi|_{\mathbf{q}} &= \left[d\Psi|_{\mathbf{q}}\right]^{\dagger} \cdot \nabla\phi|_{\Psi(\mathbf{q})} \\
&= \left[d\Psi|_{\mathbf{q}}\right]^{-1} \cdot \nabla\phi|_{\Psi(\mathbf{q})},
\end{aligned}$$

where $[d\Psi|_{\mathbf{q}}]^{\dagger}$ is the adjoint of $d\Psi|_{\mathbf{q}}$ and the second equality follows from the fact that the adjoint of an orthogonal transformation is its inverse.

## References

[BBB*11] Boyer E., Bronstein A. M., Bronstein M. M., Bustos B., Darom T., Horaud R., Hotz I., Keller Y., Keustermans J., Kovnatsky A., et al.: SHREC 2011: Robust feature detection and description benchmarkk. In *Eurographics Workshop on 3D Object Retrieval* (2011).

[BBC*10] Bronstein A., Bronstein M., Castellani U., Dubrovina A., Guibas L., Horaud R., Kimmel R., Knossow D., Von Lavante E., Mateus D., et al.: SHREC 2010: Robust feature detection and description benchmark. In *Eurographics Workshop on 3D Object Retrieval* (2010).

[BBK08] Bronstein A. M., Bronstein M. M., Kimmel R.: *Numerical Geometry of Non-Rigid Shapes*. Springer Verlag, 2008.

[BBL*17] Bronstein M. M., Bruna J., LeCun Y., Szlam A., Vandergheynst P.: Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine 34* (2017), 18–42.

[BETVG08] Bay H., Ess A., Tuytelaars T., Van Gool L.: Speeded-up robust features (SURF). *Computer Vision and Image Understanding 110* (2008), 346–359.

[BK10] Bronstein M. M., Kokkinos I.: Scale-invariant heat kernel signatures for non-rigid shape recognition. In *Computer Vision and Pattern Recognition* (2010), pp. 1704–1711.

[BLV*18] Bekkers E. J., Lafarge M. W., Veta M., Eppenhof K. A., Pluim J. P., Duits R.: Roto-translation covariant convolutional networks for medical image analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2018), Springer, pp. 440–448.

[BMRB16] Boscaini D., Masci J., Rodolà E., Bronstein M.: Learning shape correspondence with anisotropic convolutional neural networks. In *Advances in Neural Information Processing Systems* (2016), pp. 3189–3197.

[BTVG06] Bay H., Tuytelaars T., Van Gool L.: SURF: Speeded up robust features. In *European Conference on Computer Vision*. (2006), pp. 404–417.

[CK16] Chirikjian G. S., Kyatkin A. B.: *Harmonic Analysis for Engineers and Applied Scientists: Updated and Expanded Edition*. Courier Dover Publications, 2016.

[CL06] Coifman R. R., Lafon S.: Diffusion maps. *Applied and Computational Harmonic Analysis 21*, 1 (2006), 5–30.

[Cow73] Cowper G.: Gaussian quadrature formulas for triangles. *International Journal for Numerical Methods in Engineering 7*, 3 (1973), 405–408.

[CPK19] Choy C., Park J., Koltun V.: Fully convolutional geometric features. In *International Conference on Computer Vision* (2019), pp. 8958–8966.

[CRB*16] Cosmo L., Rodolà E., Bronstein M. M., Torsello A., Cremers D., Sahillioglu Y.: SHREC '16: Partial matching of deformable shapes. *Eurographics Workshop on 3D Object Retrieval* (2016).

[CW16a] Cohen T., Welling M.: Group equivariant convolutional networks. In *International Conference on Machine Learning* (2016), pp. 2990–2999.

[CW16b] Cohen T. S., Welling M.: Steerable CNNs. *arXiv preprint arXiv:1612.08498* (2016).

[CWKW19] Cohen T., Weiler M., Kicanaoglu B., Welling M.: Gauge equivariant convolutional networks and the icosahedral CNN. In *International Conference on Machine Learning* (2019), pp. 1321–1330.

[DBI18] Deng H., Birdal T., Ilic S.: PPFnet: Global context aware local features for robust 3d point matching. In *Computer Vision and Pattern Recognition* (2018), pp. 195–205.

[DBI19] Deng H., Birdal T., Ilic S.: 3D local features for direct pairwise registration. In *Computer Vision and Pattern Recognition* (2019), pp. 3244–3253.

[DBV16] Defferrard M., Bresson X., Vandergheynst P.: Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems* (2016), pp. 3844–3852.

[dHWCW20] de Haan P., Weiler M., Cohen T., Welling M.: Gauge equivariant mesh CNNs: Anisotropic convolutions on geometric graphs. *arXiv preprint arXiv:2003.05425* (2020).

[DLRT19] Dyke R. M., Lai Y.-K., Rosin P. L., Tam G. K.: Non-rigid registration under anisotropic deformations. *Computer Aided Geometric Design 71* (2019), 142–156.

[DSL*19] Dyke R. M., Stride C., Lai Y.-K., Rosin P. L., Aubry M., Boyarski A., Bronstein A. M., Bronstein M. M., Cremers D., Fisher M., Groueix T., Guo D., Kim V. G., Kimmel R., Lähner Z., Li K., Litany O., Remez T., Rodolá E., Russell B. C., Sahillioğlu Y., Slossberg R., Tam G. K. L., Vestner M., Wu Z., Yang J.: Shape correspondence with isometric and non-isometric deformations. In *Eurographics Workshop on 3D Object Retrieval* (2019).

[EABMD18] Esteves C., Allen-Blanchette C., Makadia A., Daniilidis K.: Learning SO(3) equivariant representations with spherical CNNs. In *European Conference on Computer Vision*. (2018), pp. 52–68.

[FA91] Freeman W. T., Adelson E. H.: The design and use of steerable filters. *Transactions on Pattern Analysis and Machine Intelligence 13* (1991), 891–906.

[GBS*16] Guo Y., Bennamoun M., Sohel F., Lu M., Wan J., Kwok N. M.: A comprehensive performance evaluation of 3D local feature descriptors. *International Journal of Computer Vision 116* (2016), 66–89.

[GSB*13] Guo Y., Sohel F., Bennamoun M., Lu M., Wan J.: Rotational projection statistics for 3D local surface description and object recognition. *International Journal of Computer Vision 105* (2013), 63–86.

[HA19] Herholz P., Alexa M.: Efficient computation of smoothed exponential maps. *Computer Graphics Forum 38* (2019), 79–90.

[KBLB12] Kokkinos I., Bronstein M. M., Litman R., Bronstein A. M.: Intrinsic shape context descriptors for deformable shapes. In *Computer Vision and Pattern Recognition* (2012), pp. 159–166.

[KC99a] Kyatkin A., Chirikjian G.: Computation of robot configuration and workspaces via the Fourier transform on the discrete-motion group. *The International Journal of Robotics Research 18*, 6 (1999), 601–615.

[KC99b] Kyatkin A. B., Chirikjian G. S.: Pattern matching as a correlation on the discrete motion group. *Computer Vision and Image Understanding 74*, 1 (1999), 22–35.

[KS04] Ke Y., Sukthankar R.: PCA-SIFT: A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition* (2004), vol. 2, IEEE, pp. 506–513.

[KW17] Kipf T. N., Welling M.: Semi-supervised classification with graph convolutional networks. *ICLR* (2017).

[KZK17] Khoury M., Zhou Q.-Y., Koltun V.: Learning compact geometric features. In *International Conference on Computer Vision* (2017), pp. 153–161.

[Low99] Lowe D. G.: Object recognition from local scale-invariant features. In *International Conference on Computer Vision* (1999), vol. 2, pp. 1150–1157.

[Low04] Lowe D. G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60* (2004), 91–110.

[LRB*16] Lähner Z., Rodolà E., Bronstein M. M., Cremers D., Burghard O., Cosmo L., Dieckmann A., Klein R., Sahillioglu Y.: SHREC'16: Matching of deformable shapes with topological noise. *Eurographics Workshop on 3D Object Retrieval* (2016).

[LRF10] Lipman Y., Rustamov R. M., Funkhouser T. A.: Biharmonic distance. *Transactions on Graphics 29* (2010), 1–11.

[LYLG18] Li K., Yang J., Lai Y.-K., Guo D.: Robust non-rigid registration with reweighted position and transformation sparsity. *IEEE Transactions on Visualization and Computer Graphics 25*, 6 (2018), 2255–2269.

[MBBV15] Masci J., Boscaini D., Bronstein M., Vandergheynst P.: Geodesic convolutional neural networks on riemannian manifolds. In *Proceedings of the International Conference on Computer Vision Workshops* (2015), pp. 37–45.

[MBK*20] Mitchel T. W., Brown B., Koller D., Weyrich T., Rusinkiewicz S., Kazhdan M.: Efficient spatially adaptive convolution and correlation, 2020. http://arxiv.org/abs/2006.13188.

[MBM*17] Monti F., Boscaini D., Masci J., Rodola E., Svoboda J., Bronstein M. M.: Geometric deep learning on graphs and manifolds using mixture model CNNs. In *Computer Vision and Pattern Recognition* (2017), pp. 5115–5124.

[MK10] Möbius J., Kobbelt L.: OpenFlipper: an open source geometry processing and rendering framework. In *International Conference on Curves and Surfaces* (2010), Springer, pp. 488–500.

[MR12] Melvær E. L., Reimers M.: Geodesic polar coordinates on polygonal meshes. *Computer Graphics Forum 31* (2012), 2423–2435.

[MS05] Mikolajczyk K., Schmid C.: A performance evaluation of local descriptors. *Transactions on Pattern Analysis and Machine Intelligence 27* (2005), 1615–1630.

[PD11] Petrelli A., Di Stefano L.: On the repeatability of the local reference frame for partial shape matching. In *International Conference on Computer Vision* (2011), pp. 2244–2251.

[PDKS19] Perraudin N., Defferrard M., Kacprzak T., Sgier R.: DeepSphere: Efficient spherical convolutional neural network with HEALPix sampling for cosmological applications. *Astronomy and Computing 27* (2019), 130–146.

[PO18] Poulenard A., Ovsjanikov M.: Multi-directional geodesic neural networks via equivariant convolution. *Transactions on Graphics 37*, 6 (2018), 1–14.

[RC11] Rusu R. B., Cousins S.: 3D is here: Point Cloud Library (PCL). In *International Conference on Robotics and Automation* (2011).

[Rei04] Reimers M.: *Topics in mesh based modeling*. PhD thesis, PhD thesis, University of Oslo, 2004.

[Rus10] Rustamov R. M.: Barycentric coordinates on surfaces. *Computer Graphics Forum 29* (2010), 1507–1516.

[SDL18] Schonsheck S. C., Dong B., Lai R.: Parallel transport convolution: A new tool for convolutional neural networks on manifolds. *arXiv preprint arXiv:1805.07857* (2018).

[SF96] Simoncelli E. P., Farid H.: Steerable wedge filters for local orientation analysis. *Transactions on Image Processing 5* (1996), 1377–1382.

[SGW06] Schmidt R., Grimm C., Wyvill B.: Interactive decal compositing with discrete exponential maps. *Transactions on Graphics 25* (2006), 605–613.

[SMKF04] Shilane P., Min P., Kazhdan M., Funkhouser T.: The Princeton shape benchmark. In *Proceedings Shape Modeling Applications* (2004), pp. 167–178.

[SOG09] Sun J., Ovsjanikov M., Guibas L.: A concise and provably informative multi-scale signature based on heat diffusion. In *Computer Graphics Forum* (2009), vol. 28, pp. 1383–1392.

[SSC19] Sharp N., Soliman Y., Crane K.: The vector heat method. *Transactions on Graphics 38* (2019), 1–19.

[SSDS19] Spezialetti R., Salti S., Di Stefano L.: Performance evaluation of learned 3D features. In *International Conference on Image Analysis and Processing* (2019), Springer, pp. 519–531.

[SSS19] Spezialetti R., Salti S., Stefano L. D.: Learning an effective equivariant 3D descriptor without supervision. In *International Conference on Computer Vision* (2019), pp. 6401–6410.

[STDS14] 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.

[THO99] Teo P. C., Hel-Or Y.: Design of multiparameter steerable functions using cascade basis reduction. *Transactions on Pattern Analysis and Machine Intelligence 21* (1999), 552–556.

[TSDS10a] Tombari F., Salti S., Di Stefano L.: Unique shape context for 3D data description. In *Proceedings of the ACM Workshop on 3D Object Retrieval* (2010), pp. 57–62.

[TSDS10b] Tombari F., Salti S., Di Stefano L.: Unique signatures of histograms for local surface description. In *European Conference on Computer Vision*. (2010), pp. 356–369.

[VLB*17] Vestner M., Lähner Z., Boyarski A., Litany O., Slossberg R., Remez T., Rodola E., Bronstein A., Bronstein M., Kimmel R., et al.: Efficient deformable shape correspondence via kernel matching. In *3D Vision* (2017), pp. 517–526.

[WB18] Worrall D., Brostow G.: Cubenet: Equivariance to 3D rotation and translation. In *European Conference on Computer Vision*. (2018), pp. 567–584.

[WC19] Weiler M., Cesa G.: General E (2)-equivariant steerable CNNs. In *Advances in Neural Information Processing Systems* (2019), pp. 14334–14345.

[WGTB17] Worrall D. E., Garbin S. J., Turmukhambetov D., Brostow G. J.: Harmonic networks: Deep translation and rotation equivariance. In *Computer Vision and Pattern Recognition* (2017), pp. 5028–5037.

[WGW*18] Weiler M., Geiger M., Welling M., Boomsma W., Cohen T. S.: 3D steerable CNNs: Learning rotationally equivariant features in volumetric data. In *Advances in Neural Information Processing Systems* (2018), pp. 10381–10392.

[WGY*18] Wang H., Guo J., Yan D.-M., Quan W., Zhang X.: Learning 3D keypoint descriptors for non-rigid shape matching. In *European Conference on Computer Vision*. (2018), pp. 3–19.

[WHS18] Weiler M., Hamprecht F. A., Storath M.: Learning steerable filters for rotation equivariant CNNs. In *Computer Vision and Pattern Recognition* (2018), pp. 849–858.

[WW19] Worrall D., Welling M.: Deep scale-spaces: Equivariance over scale. In *Advances in Neural Information Processing Systems* (2019), pp. 7364–7376.

[ZBH12] Zaharescu A., Boyer E., Horaud R.: Keypoints and local descriptors of scalar functions on 2D manifolds. *International Journal of Computer Vision 100* (2012), 78–98.

[ZMT05] Zhang E., Mischaikow K., Turk G.: Feature-based surface parameterization and texture mapping. *Transactions on Graphics 24* (2005), 1–27.

[ZXW*19] Zhao F., Xia S., Wu Z., Duan D., Wang L., Lin W., Gilmore J. H., Shen D., Li G.: Spherical U-Net on cortical surfaces: Methods and applications. In *International Conference on Information Processing in Medical Imaging* (2019), Springer, pp. 855–866.

**Supporting Information**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary material