EXTENDING CONVOLUTION THROUGH SPATIALLY ADAPTIVE ALIGNMENT

by
Thomas Worth Mitchel

A dissertation submitted to The Johns Hopkins University in conformity
with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
May, 2022

Abstract

Convolution underlies a variety of applications in computer vision and graphics, including efficient filtering, analysis, simulation, and neural networks. However, convolution has an inherent limitation: when convolving a signal with a filter, the filter orientation remains fixed as it travels over the domain, and convolution loses effectiveness in the presence of deformations that change alignment of the signal relative to the local frame. This problem metastasizes when attempting to generalize convolution to domains without a canonical orientation, such as the surfaces of 3D shapes, making it impossible to locally align signals and filters in a consistent manner.

This thesis presents a unified framework for transformation-equivariant convolutions on arbitrary homogeneous spaces and 2D Riemannian manifolds called *extended convolution*. This approach is based on the the following observation: to achieve equivariance to an arbitrary class of transformations, we only need to consider how the positions of points as seen in the frames of their neighbors deform. By defining an equivariant frame operator at each point with which we align the filter, we correct for the change in the relative positions induced by the transformations. This construction places no constraints on the filters, making extended convolution highly descriptive. Furthermore, the framework can handle arbitrary transformation groups, including higher-dimensional non-compact groups that act non-linearly on the domain. Critically, extended convolution naturally generalizes to arbitrary 2D Riemannian manifolds as it does not need a canonical coordinate system to be applied.

The power and utility of extended convolution is demonstrated in several applications. A unified framework for image and surface feature descriptors called *Extended Convolution Histogram of Orientations* (ECHO) is proposed, based on the optimal filters maximizing the response of the extended convolution at a given point. Using the generalization of extended convolution to surface vector fields, state-of-the-art surface convolutional neural networks (CNNs) are constructed. Last, we move beyond rotations and isometries and use extended convolution to design spherical CNNs equivariant to Möbius transformations, representing a first step toward conformally-equivariant surface networks.

Thesis Readers

Dr. Michael Kazhdan (Primary Advisor)
Professor
Department of Computer Science
Johns Hopkins University

Dr. Gretar Tryggvason
Professor
Department of Mechanical Engineering
Johns Hopkins University

Dr. Noah J. Cowan
Professor
Department of Mechanical Engineering
Johns Hopkins University

Attribution

The figures, tables, algorithms, and portions of the text in this thesis come from several papers published by myself and collaborators. The following publications represent the original source for this content, re-used here with permission:

- Thomas W. Mitchel, Benedict Brown, David Koller, Tim Weyrich, Szymon Rusinkiewicz, and Michael Kazhdan. Efficient Spatially Adaptive Convolution and Correlation. *arXiv preprint arXiv:2006.13188*, 2020 (Chapters 1, 3, 5, 8)

- Thomas W. Mitchel, Szymon Rusinkiewicz, Gregory S. Chirikjian, and Michael Kazhdan. ECHO: Extended Convolution Histogram of Orientations for Local Surface Description. *Computer Graphics Forum*, 40(1):180–194, 2021 (Chapters 1, 4, 5, 8)

- Thomas W. Mitchel, Vladimir G. Kim, and Michael Kazhdan. Field Convolutions for Surface CNNs. In *International Conference on Computer Vision*, pages 10001–10011, 2021 (Chapters 1, 4, 6, 8)

- Thomas W. Mitchel, Noam Aigerman, Vladimir G. Kim, and Michael Kazhdan. Möbius Convolutions for Spherical CNNs. *arXiv preprint arXiv:2201.12212*, 2022 (Chapters 1, 3, 7, 8, Appendix I)

Contents

Tables

Figures

Algorithms

Chapter 1

Introduction

Convolution underpins a number of applications in vision and geometry processing, including pattern recognition and compression of images [KJM05, Wal91], symmetry detection in 2D images [KS06] and 3D models [KFR04], reconstruction of 3D surfaces [SBS06], inversion of the Radon transform for medical imaging [KS01, Nat01], and convolutional neural networks (**CNNs**) [FM82, LBD+89]. Convolution takes two inputs: a signal ψ and a filter f. The signal ψ is the input data – an image, volumetric data, network features, really any reasonable function. The filter f is a function that assigns weights to points based on their position relative to their neighbors. A simple example of a filter is a Gaussian,

$$f(z) = e^{-|z|^2}.$$

As a weight function, it assigns the largest value to the origin, and increasingly smaller values to points farther away. Traditionally, the signal and filter belong to either $L^2(\mathbb{R}^n, \mathbb{R})$ or $L^2(\mathbb{R}^n, \mathbb{C})$, the space of either real- or complex-valued square-integrable functions on $n-$dimensional Euclidean space.

Formally, the *convolution* of ψ with f is the function in $L^2(\mathbb{R}^n, \mathbb{R})$ with

$$(\psi * f)(y) = \int_{\mathbb{R}^n} \psi(z)\, f(y - z)\, dz. \tag{1.1}$$

Here, the filter also slides across the signal, and at each point y the value of the signal is replaced by the linear combination of the neighboring values and filter weights determined by $y - z$, the position of the point *as seen from its neighbors*.

Convolution is effective in applications because it responds to a contextualized window on the signal, forcing the task to be *translation-equivariant*. In other words, for any signal ψ, filter f, and translation t, convolution commutes with the action of t by left shifts, $[t\,\psi](z) = \psi(z - t)$ with

$$t\,(\psi * f) = (t\,\psi * f),$$

which follows easily from a change of variables in Equation (1.1). In particular, this property has been fundamental to the success of CNNs due to the translational symmetries inherent in most vision tasks [FM82, LBD+89, LB+95], as the image data can usually be expected to share a consistent alignment relative to the image plane. For example, in a series of images captured by a car's dashboard camera, the surface of the road may always align with the horizontal axis of image plane, so the motion of pedestrians crossing the street can be well-described by lateral shifts.

However, perception and analysis frameworks based on convolution run into trouble when data isn't consistently aligned, as is the case when salient objects of features appear in a collection of images in different orientations. While convolution commutes with translations, it does not commute with rotations, dilations, or more complex types of transformations – convolving an image with a filter will produce a different response at rotated or scaled version of the same pattern in the image. To overcome these limitations, a simple approach may consist of quantizing the rotation angle and pooling responses over different filter orientations. For CNNs, training data can be extended by applying randomly sampled rotations to network inputs, forcing it to learn a measure of rotation-equivariance. That said, these approaches aren't readily generalizable to more complex groups of transformations that are better representative of the kinds of deformations found in real-world data, which cannot be parameterized on compact domains.

These problems metastasize when attempting to generalize convolution to domains without either global or local, repeatable coordinate systems, such as the surfaces of 3D shapes. The lack of a such a coordinate system makes it impossible to locally align signals and filters in a consistent manner, precluding the simple transposition of Euclidean notions of convolution to curved surfaces. To circumvent this problem, various approaches have tried to impose a regular grid structure, by either inducing a planar parameterization via projections [SMKLM15], cuts [SBR16], and tilings [MGA+17] or volumetric rasterization [WSK+15]. Unfortunately, neither method provides a compelling solution – planar projections induce distortion and volumetric approaches are computationally expensive and tend to lose effectiveness in the presence of non-rigid shape deformations.

1.1 Equivariant convolutions

The advent of deep learning in imaging and vision has coincided with an increased awareness of the limitations of standard convolutional techniques, and has facilitated the development of more general convolutional frameworks equivariant to transformation groups. These methods can be broadly categorized based on whether convolution is integrated over the group itself or the *homogeneous space* – the domain on which it acts. Rotation-equivariant convolutions based on the former approach were integrated into CNNs by [CW16, CGW19], where kernels are parameterized in terms of equivariant basis functions on the group and convolution is performed by lifting features from the domain and searching over all possible transformations of the features or kernels. This approach is highly effective when considering the action of discrete groups on features sampled on a regular lattice, and has since been extended to handle the continuous group of rotations in both two and three dimensions [CGKW18, LW21]. However, this approach isn't readily generalizable – either theoretically or computationally – to higher-dimensional or non-compact groups, where there are more parameters to integrate over, the domains of integration are unbounded, and the representations are infinite-dimensional.

Equivariant CNNs that integrate over the domain on which the group acts can trace their lineage to earlier work on steerable filters [FA91, SF96, THO99]. Kernels are parameterized in terms of equivariant basis functions *on the domain* that rotate or dilate with the local coordinate system [WGTB17, WC19, WW19, SSS19a], and this approach has been extended to both volumetric domains [WC19] and point clouds [QSMG17]. Unfortunately, finite-dimensional equivariant bases often don't exist for non-commutative and non-compact transformation groups of interest, limiting the practical scope of these approaches.

Critically, the notion of rotation-equivariance has facilitated the generalization of convolutional frameworks to domains without canonical coordinate systems such as the sphere [CW16, EMD20] and arbitrary 2D surfaces. In the latter case, convolution is defined intrinsically over the Riemannian manifold and is *isometry-equivariant* – providing a repeatable response in the presence of distance-preserving transformations. These approaches can generally be classified in relation to two emerging paradigms: *diffusive* convolutions and *transporting* convolutions. In the former, convolution operations are closely related to heat diffusion on surfaces wherein heat (e.g. Gaussian) kernels are used to propagate scalar features. Despite their success in a variety of scenarios, most notably in dense shape correspondence, these methods face an intractable problem: radially symmetric filters are individually undiscriminating and diffusive frameworks are not naturally suited to handle the orientation ambiguity problem introduced by the use of more descriptive, anisotropic kernels.

Recently, several techniques have been introduced for surface convolutions based on parallel transport [PO18, dHWCW20, WEH20]. In contrast to diffusive approaches, transporting convolutions are designed specifically to address the rotation ambiguity problem by propagating tangent vector features that transform with local coordinate systems. However, to make the convolution independent of the choice of local coordinate frame, most existing methods strongly constrain the class of filters that can be used.

1.2 Beyond rotations, dilations, and isometries

Despite their success, rotation- and isometry-equivariant CNNs can fail to achieve adequate performance in the presence of the kinds of complex deformations commonly found in real-world image and shape data [MKK21]. Such deformations may potentially be better modeled by higher-dimensional transformation groups. For example, homographies (projective transformations) better approximate changes in camera viewpoints than similarities (rotations and dilations) [HZ03] and, for spherical images, can be represented using conformal (angle-preserving) transformations [EMSJB14, SS16]. For geometry processing, conformal transformations encompass a broader class of deformations than isometries that still preserve the sense of 'shape' [LPRM02, GWC$^+$04, CPS11].

While the importance of these transformation groups is well known, there exists little work generalizing equivariant convolutions to handle them. This is likely due to several factors: 1). The majority of successful existing approaches formulate convolution as an integral over the group of transformations as opposed to the domain itself [CW16, CGW19, LW21]; or 2). rely on finite-dimensional group representations to parameterize kernels [WGTB17, WW19]; and 3) expect the transformations to act linearly on the domain [FSIW20]. None of these approaches are readily extended to handle higher-dimensional, non-compact groups of transformations where the domain of integration is unbounded, the representations are infinite-dimensional, and the action of the group is nonlinear.

1.3 Contributions and outline

This thesis presents a unified framework for transformation-equivariant convolutions on arbitrary homogeneous spaces and 2D Riemannian manifolds, which we call *extended convolution*. Our approach is based on the following observation: to achieve equivariance to an arbitrary class of transformations, we only need to consider how the positions of points as seen in the frames of their neighbors deform. By defining an equivariant frame operator at each point with which we align the filter, we correct for the change in the relative positions induced by the transformations

The resulting framework is highly flexible and descriptive - the construction places no constraints on the kinds of filters that can be used. Furthermore, the framework can handle arbitrary transformation groups, including higher-dimensional non-compact groups that act non-linearly on the domain, such as Möbius transformations of the sphere. Critically, extended convolution naturally generalizes to arbitrary 2D Riemannian manifolds – such as the surfaces of 3D shapes – as it does not need a canonical coordinate system to be applied.

This thesis is divided into two parts, focusing on theory and applications, respectively. In Part I, we first review the relevant mathematical background (Chapter 2), including transformation groups, diffeomorphisms, and their actions on homogeneous spaces. Chapter 3 develops a general theory of extended convolution on arbitrary homogeneous spaces, and makes the framework concrete by realizing equivariant extended convolutions on three canonical domains – the plane, the sphere, and the disk. In Chapter 4, the framework is generalized to construct isometry-equivariant convolutional operators on 2D Riemannian manifolds.

In Part II, we demonstrate the power and flexibility of extended convolution in several applications. In Chapter 5, we use extended convolution to develop a unified framework for image and surface feature descriptors called *Extended Convolution Histogram of Orientations* (**ECHO**). In Chapter 6, we use the generalization of extended convolution to surface vector fields to construct state-of-the art surface CNNs. Last, we move beyond rotations and isometries and use extended convolution to construct spherical CNNs equivariant to Möbius transformations (Chapter 7).

Part I

Extended Convolution

Chapter 2

Background

2.1 Transformation groups

Many interesting classes of transformations form groups. A *group* is a set G, equipped with a product operation $\circ : G \times G \longrightarrow G$ satisfying the following properties:

1. **Associativity**: $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3),\ \forall g_1, g_2, g_3 \in G$

2. **Unit Element**: There exists an element $e \in G$ such that $e \circ g = g = g \circ e$ for all $g \in G$

3. **Inverses Exist**: For each $g \in G$, there exists an element $h \in G$ such that $h \circ g = e = g \circ h$, denoted as $g^{-1} \equiv h$.

A *transformation group* is a group G that acts on a set S such that the mapping

$$\begin{aligned} g : S &\longrightarrow S \\ x &\longmapsto gx \end{aligned}$$

is a bijection with the properties

$$(g_1 \circ g_2)x = g_1(g_2 x) \qquad \text{and} \qquad ex = x$$

for all $x \in S$ and $g, g_1, g_2 \in G$. Here we consider transformation groups that are *matrix Lie groups* whose elements belong either to $\mathbb{R}^{n\times n}$ or $\mathbb{C}^{n\times n}$ and have the special property that the set G is a smooth manifold. That is, a $d-$dimensional matrix

Lie group can be parameterized on a subset U of $\mathbb{R}^d$ or $\mathbb{C}^d$ such that $g = g(\mathbf{q}) = [g_{ij}(\mathbf{q})],\ \mathbf{q} \in U$ and each matrix entry is an analytic function.

2.1.1 Homogeneous spaces

Given a (Lie) transformation group G acting on a set S, the latter is called a *homogeneous space* if it is a smooth manifold and G acts transitively – for any $x, y \in S$ there exists $g \in G$ such that $gx = y$.

Homogeneous spaces can be viewed as coset spaces. A subgroup H of a group G is a subset $H \subseteq G$ that forms a group under the same product operation. Given a subgroup $H \subseteq G$ and an element $g \in G$, the associated *left coset* is the set $gH = \{g \circ h \,|\, h \in H\}$. Cosets are disjoint, so for any $g_1, g_2 \in G$, $g_1H \cap g_2H \neq \emptyset$ if and only if $g_1H = g_2H$. The set of all left cosets is the called the *coset space*, denoted $G/H = \{gH \,|\, g \in G\}$, and the natural map

$$\begin{aligned} \pi : G &\longrightarrow G/H \\ g &\longmapsto gH \end{aligned} \tag{2.1}$$

sending $g \in G$ to $gH \in G/H$ is called the *quotient map*. The quotient map intertwines the actions of G on itself and on G/H, from which it follows that G acts naturally on G/H through left multiplication,

$$g_1(\pi(g_2)) \equiv \pi(g_1 \circ g_2) \iff g_1(g_2H) \equiv (g_1 \circ g_2)H$$

for all $g_1, g_2 \in G$. The choice of H induces a natural "origin", denoted $0 \in G/H$ with $0 = eH$, which is preserved under the action of H. That is, $g(0) = 0$ if an only if $g \in H$.

Similarly, given a point $x \in S$, the set of all elements in G mapping x to itself forms a subgroup called the *stabilizer* of x, denoted $H_x = \{g \in G \,|\, gx = x\}$. The connection between the homogeneous space S and the coset space G/H_x can be made concrete via the map

$$\begin{aligned} \kappa : G/H_x &\to S \\ gH_x &\mapsto gx \end{aligned} \tag{2.2}$$

which is well-defined, bijective, and intertwines the actions of G on G/H_x and S [Lee12]. It follows that S and G/H_x are isomorphic, and we write $S \cong G/H_x$, denoting the natural mapping from G to S by

$$\widetilde{\pi} \equiv \kappa \circ \pi. \tag{2.3}$$

Note that $\kappa(0) = x$, so in some sense viewing the homogeneous space S as the coset space G/H_x provides something similar to a “global parameterization" of S about x. While any $x \in S$ can be stabilized, both the geometry of S and the specific action of G often motivate a choice of x which simplifies the form of H_x.

2.1.2 Canonical domains

In this thesis we will pay special attention to three *canonical domains* that are of practical interest in graphics and vision – the plane, the Riemann sphere, and the disk. In what follows, we will show how each of these can be realized as homogeneous spaces under the action of transformation groups.

$\mathbb{C}$ as a homogeneous space

Here we identify the plane $\mathbb{R}^2$ with the complex line $\mathbb{C}$ via the isomorphism $(x, y) \mapsto x + iy$. The two-dimensional special Euclidean group $\mathrm{SE}(2)$ comprises all rotations and translations of the plane. A planar rotation by an angle θ can be expressed as multiplication by the complex number $e^{i\theta}$. The set of all unit complex numbers under multiplication forms the group $\mathrm{U}(1)$ which is isomorphic to $\mathrm{SO}(2)$. Similarly, a translation by a vector $\mathbf{t} = [t_1, t_2]^\top \in \mathbb{R}^2$ is equivalent to a shift by the complex number $t = t_1 + it_2 \in \mathbb{C}$. In this context, elements of SE(2) can be expressed as

$$g(\theta, t) \equiv \begin{bmatrix} e^{i\theta} & t \\ 0 & 1 \end{bmatrix} \in \mathbb{C}^{2\times 2}$$

where $e^{i\theta} \in \mathrm{U}(1)$ is a rotation by an angle θ and $t \in \mathbb{C}$ is a translation. SE(2) acts transitively on $\mathbb{C}$ with

$$g(\theta, t)\, z \equiv e^{i\theta}\, z + t.$$

It is easy to see that stabilizer subgroup of the origin $0 \in \mathbb{C}$ is the subset of SE(2) consisting of all elements such that $t = 0$ and is isomorphic to U(1). It follows that the map

$$\begin{aligned} \kappa : \mathrm{SE}(2)/\mathrm{U}(1) &\longrightarrow \mathbb{C} \\ g(\theta, t)\, \mathrm{U}(1) &\mapsto t \end{aligned} \tag{2.4}$$

is an isomorphism so $\mathbb{C} \cong \mathrm{SE}(2)/\mathrm{U}(1)$.

$\widehat{\mathbb{C}}$ as a homogeneous space

The two-sphere S^2 can be associated with the *Riemann sphere* $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ via the stereographic projection taking the north pole to $0 \in \widehat{\mathbb{C}}$. The two-dimensional complex special linear group $\mathrm{SL}(2, \mathbb{C})$ consists of all matrices in $\mathbb{C}^{2\times 2}$ with unit determinant. Elements of $\mathrm{SL}(2, \mathbb{C})$ are called *Möbius transformations* and act transitively on $\widehat{\mathbb{C}}$ by fractional linear transformations. That is, for any $g = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{SL}(2, \mathbb{C})$ and $z \in \widehat{\mathbb{C}}$,

$$gz \equiv \frac{az + b}{cz + d}. \tag{2.5}$$

Observing that

$$g\,0 = \frac{b}{d} = 0 \iff b = 0,$$

it is clear that the subgroup $\mathrm{L} \subset \mathrm{SL}(2,\mathbb{C})$ consisting of the lower-triangular elements of $\mathrm{SL}(2,\mathbb{C})$ stabilizes $0 \in \widehat{\mathbb{C}}$ and the map

$$\begin{aligned} \kappa : \mathrm{SL}(2,\mathbb{C})/\mathrm{L} &\to \widehat{\mathbb{C}} \\ g\,\mathrm{L} &\mapsto \frac{b}{d} \end{aligned} \tag{2.6}$$

is an isomorphism between $\mathrm{SL}(2,\mathbb{C})/\mathrm{L}$ and $\widehat{\mathbb{C}}$ up to multiplication by -1.

$\widehat{\mathbb{C}}$ can also be realized as a homogeneous space under the action of two-dimensional special unitary group $\mathrm{SU}(2) \subset \mathrm{SL}(2,\mathbb{C})$ via fractional linear transformations. Elements $g \in \mathrm{SU}(2)$ are of the form

$$g = \begin{bmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{bmatrix}, \quad |\alpha|^2 + |\beta|^2 = 1, \tag{2.7}$$

and can be parameterized in terms of $z{-}y{-}z$ Euler angle triplets $(\theta,\phi,\psi) \in [0,2\pi)\times [0,\pi] \times [-2\pi,2\pi)$ corresponding to a factorization as a product of one-parameter subgroups

$$g(\theta,\phi,\psi) \equiv \begin{bmatrix} e^{-\frac{i\theta}{2}} & 0 \\ 0 & e^{\frac{i\theta}{2}} \end{bmatrix} \begin{bmatrix} \cos\frac{\phi}{2} & \sin\frac{\phi}{2} \\ -\sin\frac{\phi}{2} & \cos\frac{\phi}{2} \end{bmatrix} \begin{bmatrix} e^{-\frac{i\psi}{2}} & 0 \\ 0 & e^{\frac{i\psi}{2}} \end{bmatrix}.$$

Noting that

$$g(\theta,\phi,\psi)\,0 = \tan\frac{\phi}{2}\,e^{-i\theta} = 0 \iff \phi = 0,$$

it follows that the subgroup stabilizing the origin consists of the elements of the form

$$\begin{bmatrix} e^{-\frac{i\theta}{2}} & 0 \\ 0 & e^{\frac{i\theta}{2}} \end{bmatrix}, \tag{2.8}$$

which is isomorphic to $\mathrm{U}(1)$, the group of unit complex numbers under multiplication. We note that the isomorphism between $\mathrm{SU}(2)/\mathrm{U}(1)$ and $\widehat{\mathbb{C}}$

$$\begin{aligned} \kappa : \mathrm{SU}(2)/\mathrm{U}(1) &\to \widehat{\mathbb{C}} \\ g\,\mathrm{U}(1) &\mapsto \frac{\beta}{\bar{\alpha}} = \tan\frac{\phi}{2}\,e^{-i\theta} \end{aligned} \tag{2.9}$$

induces a natural parameterization of $\widehat{\mathbb{C}}$ in spherical coordinates $(\theta,\ \phi) \in [0,2\pi)\times[0,\pi]$.

$\mathbb{D}$ as a homogeneous space

The open complex unit disk

$$\mathbb{D} = \{z \in \mathbb{C} \mid |z| < 1\},$$

is a homogeneous space under the action of the subgroup $\mathrm{SU}(1,1) \subset \mathrm{SL}(2,\mathbb{C})$, which consists of the elements of $\mathrm{SL}(2,\mathbb{C})$ of the form

$$\begin{bmatrix} \alpha & \beta \\ \overline{\beta} & \overline{\alpha} \end{bmatrix}, \quad |\alpha|^2 - |\beta|^2 = 1. \tag{2.10}$$

Similar to $\mathrm{SU}(2)$, $\mathrm{SU}(1,1)$ can be parameterized in terms of the triplets $(\theta, \tau, \psi) \in [0, 2\pi) \times \mathbb{R}_{\geq 0} \times [-2\pi, 2\pi)$ corresponding to the factorization

$$g(\theta,\tau,\psi) \equiv \begin{bmatrix} e^{-\frac{i\theta}{2}} & 0 \\ 0 & e^{\frac{i\theta}{2}} \end{bmatrix} \begin{bmatrix} \cosh\tau & \sinh\tau \\ \sinh\tau & \cosh\tau \end{bmatrix} \begin{bmatrix} e^{-\frac{i\psi}{2}} & 0 \\ 0 & e^{\frac{i\psi}{2}} \end{bmatrix}.$$

By a similar argument as above it can be shown that the origin preserving subgroup again consists of the diagonal elements in Equation (2.8) which is isomorphic to $\mathrm{U}(1)$ and that the isomorphism

$$\begin{aligned} \kappa : \mathrm{SU}(1,1)/\mathrm{U}(1) &\longrightarrow \mathbb{D} \\ g\,\mathrm{U}(1) &\mapsto \frac{\beta}{\overline{\alpha}} = \tanh\tau\, e^{-i\theta} \end{aligned} \tag{2.11}$$

induces a parameterization of $\mathbb{D}$ in the coordinates $(\theta,\ \tau) \in [0, 2\pi) \times \mathbb{R}_{\geq 0}$.

2.2 Riemannian manifolds

In this thesis we are primarily interested in defining convolutions of functions on smooth 2D Riemannian manifolds such as the plane, the sphere, and the surfaces of 3D shapes. Generally speaking, a smooth 2D manifold M is a space that is locally Euclidean, *i.e.* the local neighborhood about any point "looks" like a copy of the plane.

2.2.1 Tangent spaces and Riemannian metrics

This idea is formalized in the concept of the tangent space. At any point p on a 2D manifold M, we can attach at 2D vector space T_pM called the *tangent space*. As its name suggests, T_pM is the space of all tangent vectors at p – the velocities at p of all curves on M passing through p. A 2D *Riemannian manifold* a smooth 2D manifold M equipped with a Riemannian metric s, which is a smooth map assigning a symmetric 2D tensor s_p at each point $p \in M$. Assigning to each point $p \in M$ a basis $\{\mathbf{e}_1, \mathbf{e}_2\}_p$ in the tangent space T_pM, the metric tensor s_p defines an inner product in tangent space

$$\langle \mathbf{v}_1, \mathbf{v}_2 \rangle_p = \mathbf{v}_1^T s_p \mathbf{v}_2, \quad \forall \mathbf{v}_1, \mathbf{v}_2 \in T_pM,$$

which can be used to define the lengths of and angles between tangent vectors,

$$|\mathbf{v}|_p^2 = \langle \mathbf{v}, \mathbf{v} \rangle_p \qquad \text{and} \qquad \cos\theta = \frac{\langle \mathbf{v}_1, \mathbf{v}_2 \rangle_p}{|\mathbf{v}_1|_p \, |\mathbf{v}_2|_p}. \tag{2.12}$$

Moving forward, we will refer to smooth 2D Riemannian manifolds (M, s) interchangeably as *surfaces*.

Tangent vectors as complex numbers

Following the approach of Knoppel *et al.* [KCPS13], we represent tangent vectors as complex numbers. For each point $p \in M$, T_pM can be associated with $\mathbb{C}$ such that for any $\mathbf{v} \in T_pM$, we have $\mathbf{v} \equiv r\, e^{i\theta}$, with $r = |\mathbf{v}|_p$ and θ the angle between $\mathbf{v}$ and $\mathbf{e}_1$.

Logarithm and exponential maps

The notions of distances and angles admitted by a Riemannian metric in Equation (2.12) enable a natural parameterization of the local surface about a point. Specifically, given a point $p \in M$ and a neighboring point $q \in M$, the shortest length geodesic curve along the surface from p to q gives the information needed to describe the "position" of q in the tangent space at p. That is, the length of the curve gives a radial (geodesic) distance r_{pq} and the initial direction at p an angle θ_{pq}. Together, these give a point T_pM corresponding to the relative position of q as seen from p, called the *logarithm of q with respect to p*

$$\log_p q \equiv r_{pq}\, e^{i\theta_{pq}}. \tag{2.13}$$

Similarly, the *exponential* is the inverse of the logarithm map, taking the vector $\log_p q \in T_pM$ to the point $q \in M$. Typically, the logarithm map is only well-defined locally as for points farther away on a surface, the shortest length geodesic may not be unique.

Parallel transport

At certain points we will wish to view a vector $\mathbf{v}$ in the tangent space at a point $p \in M$ in the frame in the tangent space at a neighboring point q. To do so, we transport $\mathbf{v}$ along the shortest geodesic from p to q such that the angle between $\mathbf{v}$ and the tangent vector remains fixed as it travels along the curve. Intuitively, the vector is continuously rotated as it travels to maintain a fixed orientation relative to the tangent vector and the cumulative rotation undergone by the vector in moving along the geodesic from p to q corresponds to a linear map from T_pM to T_qM called *parallel transport*. Formally, we denote by φ_{qp} the change in angle resulting from the parallel transport $\mathcal{P}_{q\leftarrow p} : T_pM \to T_qM$ along the shortest geodesic from p to q, such that for any $\mathbf{v} \in T_pM$,

$$\mathcal{P}_{q\leftarrow p}(\mathbf{v}) \equiv e^{i\varphi_{qp}}\,\mathbf{v}. \tag{2.14}$$

2.2.2 Functions on surfaces

Given a surface M, we are broadly interested in scalar and tensor functions on M taking values in some domain D. The operation perhaps most fundamental to computing convolutions is integration. For a surface M with Riemannian metric s, the *area measure* dp at a point $p \in M$ is expressed in local coordinates $p \mapsto (z_1, z_2)$ as

$$dp \equiv \sqrt{\det s_p}\, dz_1\, dz_2. \tag{2.15}$$

and the integral of a function $\psi : M \to D$ over the surface M is the quantity

$$\int_M \psi\, dp.$$

In practice, we consider functions taking values in $D = \mathbb{R}, \mathbb{C}, \mathbb{R}^{m\times n}$, or $\mathbb{C}^{m\times n}$ belonging to the spaces of *square-integrable functions* on M

$$L^2(M, D) \equiv \left\{ \psi : M \to D \,\middle|\, \int_M |\psi|^2\, dp < \infty \right\}, \tag{2.16}$$

where $|\cdot|$ denotes either the modulus or Frobenius norm, depending on whether ψ is scalar- or tensor-valued. We note that $L^2(M, \mathbb{C})$ forms an inner-product space with the inner-product of two functions defined by integrating the product of the first with the conjugate of the second

$$\langle \psi_1,\, \psi_2 \rangle = \int_M \psi_1\, \overline{\psi_2}\, dp, \qquad \forall \psi_1,\, \psi_2 \in L^2(M, \mathbb{C}). \tag{2.17}$$

A special class of functions on surfaces are *vector fields*, maps $X : M \to TM$ assigning to each point $p \in M$ a tangent vector $X(p) \in T_pM$. We denote the space of vector fields on M as $\Gamma(TM)$.

2.2.3 Diffeomorphisms

A *diffeomorphism* $\boldsymbol{\gamma} : M \to M'$ is a smooth, bijective map between a surface M and a surface M'. At each point $p \in M$, the action of the diffeomorphism $\boldsymbol{\gamma}$ induces a linear map between the tangent spaces at corresponding points given by the differential of $\boldsymbol{\gamma}$, $d\boldsymbol{\gamma}|_p : T_pM \longrightarrow T_{\boldsymbol{\gamma}(p)}M'$. The action of an orientation-preserving diffeomorphism distorts the area measure such that

$$d\boldsymbol{\gamma}(p) = \lambda_{\boldsymbol{\gamma}}^2(p)\, dp, \tag{2.18}$$

where $\lambda_{\boldsymbol{\gamma}}^2 : M \longrightarrow \mathbb{R}_{>0}$ is a smooth map called the *scale factor*.

In fact, the set of all diffeomorphisms mapping a surface M to itself forms a group under composition, $\mathrm{Diff}(M)$, acting transitively on M. While we do not interpret general 2D surfaces as homogeneous spaces, we will see how the action of the transformation groups on the homogeneous spaces discussed in §2.1 can be viewed as diffeomorphisms on smooth manifolds. Here we will focus on two subgroups of diffeomorphisms that play an important role in vision and graphics: isometric and conformal diffeomorphisms.

Isometric diffeomorphisms

Isometric diffemorphims (which we refer to generally as isometries) preserve both orientation and distances. Formally, a diffeomorphism $\boldsymbol{\gamma} : M \longrightarrow M'$ is an *isometry* if for all $p \in M$ and $\mathbf{v}_1, \mathbf{v}_2 \in T_pM$,

$$\langle \mathbf{v}_1,\, \mathbf{v}_2 \rangle_p = \left\langle \left[\, d\boldsymbol{\gamma}|_p \,\right] \mathbf{v}_1,\, \left[\, d\boldsymbol{\gamma}|_p \,\right] \mathbf{v}_2 \right\rangle_{\boldsymbol{\gamma}(p)}. \tag{2.19}$$

Or, in other words, isometries are diffeomorphisms that preserve the inner product. Since an isometry $\boldsymbol{\gamma}$ preserves distances, it therefore must also preserve areas, so $\lambda_{\boldsymbol{\gamma}}^2(p) = 1$ for all $p \in M$ and the area measure in Equation (2.15) is invariant. We note that the distance-preservation property of isometries induces a natural notion of locality. Namely, a diffeomorphism $\boldsymbol{\gamma}$ is a *local isometry* if there exist neighborhoods $\mathcal{N} \subseteq M$ and $\mathcal{N}' = \boldsymbol{\gamma}(\mathcal{N}) \subseteq M'$ such that the restriction $\boldsymbol{\gamma} : \mathcal{N} \longrightarrow \mathcal{N}'$ is an isometry.

It follows from Equation (2.19) that if $\boldsymbol{\gamma} : M \longrightarrow M'$ is an isometry, then at each point $p \in M$, the differential $d\boldsymbol{\gamma}|_p$ is a special orthogonal transformation. If the bases in the tangent spaces are *orthonormal*, then the action of the differential $d\boldsymbol{\gamma}|_p : T_pM \longrightarrow T_{\boldsymbol{\gamma}(p)}M'$ can be expressed as a rotation by an angle γ_p, with

$$\left[\, d\boldsymbol{\gamma}|_p \,\right]\mathbf{v} \equiv e^{i\gamma_p}\,\mathbf{v}. \tag{2.20}$$

Similarly, the logarithm and transport operators in Equations (2.13-2.14) transform under isometries as [GQ20]

$$\log_{\boldsymbol{\gamma}(p)} \boldsymbol{\gamma}(q) = e^{i\gamma_p}\, \log_p q \qquad \text{and} \qquad \mathcal{P}_{\boldsymbol{\gamma}(q)\leftarrow\boldsymbol{\gamma}(p)} = e^{i(\gamma_q-\gamma_p)}\mathcal{P}_{q\leftarrow p}. \tag{2.21}$$

Conformal diffeomorphisms

Conformal diffeomorphisms preserve orientations and angles, encompassing a broader class of transformations than isometries that still preserve a sense of “shape”. Formally, a diffeomorphism $\boldsymbol{\gamma} : M \longrightarrow M'$ is *conformal* if for all $p \in M$ and $\mathbf{v}_1, \mathbf{v}_2 \in T_pM$

$$\lambda^2_{\boldsymbol{\gamma}}(p)\,\langle \mathbf{v}_1,\, \mathbf{v}_2\rangle_p = \left\langle \left[\, d\boldsymbol{\gamma}|_p \,\right]\mathbf{v}_1,\, \left[\, d\boldsymbol{\gamma}|_p \,\right]\mathbf{v}_2 \right\rangle_{\boldsymbol{\gamma}(p)}, \tag{2.22}$$

where $\lambda^2_{\boldsymbol{\gamma}}(p)$ is the scale factor as in Equation (2.18). It is easy to see that isometries are a special case of conformal transformations where $\lambda^2_{\boldsymbol{\gamma}} = 1$.

2.2.4 Homogeneous spaces as Riemannian manifolds

Here we reconsider the three canonical domains in §2.1.2 as Riemannian manifolds. The actions of the associated transformation groups can be viewed as diffeomorphisms, which we will classify as either isometric or conformal.

$\mathbb{C}$ as a Riemannian manifold

The plane – equivalently $\mathbb{C}$ – is a Riemannian manifold under the Euclidean metric, expressed at $z = z_1 + iz_2 \in \mathbb{C}$ as

$$s_z = dz_1^2 + dz_2^2.$$

Here, s_z corresponds to the standard dot product in Euclidean space, and in associating $T_z\mathbb{C}$ with $\mathbb{C}$ we can write

$$\langle \mathbf{v}_1, \mathbf{v}_2 \rangle_z = \frac{1}{2}(\bar{\mathbf{v}}_1\mathbf{v}_2 + \mathbf{v}_1\bar{\mathbf{v}}_2),$$

for all $\mathbf{v}_1, \mathbf{v}_2 \in T_z\mathbb{C}$. The area measure is the familiar

$$dz = dz_1 dz_2. \tag{2.23}$$

Given a function $\psi \in L^2(M, \mathbb{C})$ we can express the differential of ψ at $x \in \mathbb{C}$ in local coordinates $z = z_1 + iz_2$ as the complex number

$$d\,\psi|_x \equiv \frac{1}{2}\left(\frac{\partial\psi}{\partial z_1}\bigg|_x - i\,\frac{\partial\psi}{\partial z_2}\bigg|_x\right). \tag{2.24}$$

Recall that elements $g(\theta, t) \in \mathrm{SE}(2)$ act transitively on $\mathbb{C}$ via

$$g(\theta, t)\, z = e^{i\theta}\, z + t$$

and the differential of $g(\theta, t) \in \mathrm{SE}(2)$ is given by

$$d\,g(\theta, t)|_z = e^{i\theta} \tag{2.25}$$

Then, for any $z \in \mathbb{C}$, $\mathbf{v}_1, \mathbf{v}_2 \in T_z\mathbb{C}$, and $g = g(\theta, t) \in \mathrm{SE}(2)$ we have

$$\begin{aligned} \Big\langle \big[\, d\,g|_z \,\big]\, \mathbf{v}_1,\, \big[\, d\,g|_z \,\big]\, \mathbf{v}_2 \Big\rangle_{g\,z} &= \langle e^{i\theta}\, \mathbf{v}_1,\, e^{i\theta}\, \mathbf{v}_2 \rangle_{g\,z} \\ &= \frac{1}{2}(e^{-i\theta}\, \bar{\mathbf{v}}_1\, e^{i\theta}\, \mathbf{v}_2 + e^{i\theta}\, \mathbf{v}_1\, e^{-i\theta}\, \bar{\mathbf{v}}_2) \\ &= \frac{1}{2}(\bar{\mathbf{v}}_1\mathbf{v}_2 + \mathbf{v}_1\bar{\mathbf{v}}_2) \\ &= \langle \mathbf{v}_1, \mathbf{v}_2 \rangle_z, \end{aligned}$$

so it's clear that transformations in SE(2) are isometric diffeomorphisms of the plane.

$\widehat{\mathbb{C}}$ as a Riemannian manifold

We view $\widehat{\mathbb{C}}$ as a Riemannian manifold under the round metric, expressed at $z = z_1 + i\,z_2 \in \widehat{\mathbb{C}}$ as

$$s_z = \frac{4}{\left(1+|z|^2\right)^2}\,(dz_1^2 + dz_2^2), \tag{2.26}$$

and in associating $T_z\widehat{\mathbb{C}}$ with $\mathbb{C}$ we can write

$$\langle \mathbf{v}_1,\, \mathbf{v}_2\rangle_z = \frac{2}{\left(1+|z|^2\right)^2}(\bar{\mathbf{v}}_1\mathbf{v}_2 + \mathbf{v}_1\bar{\mathbf{v}}_2),$$

for all $\mathbf{v}_1,\, \mathbf{v}_2 \in T_z\widehat{\mathbb{C}}$. Here, the area measure takes the form

$$dz = \frac{4\,dz_1\,dz_2}{\left(1+|z|^2\right)^2}, \tag{2.27}$$

and the differential of a function $\psi \in L^2(\widehat{\mathbb{C}}, \mathbb{C})$ is defined in the same manner as in Equation (2.24). Recall that $\mathrm{SL}(2,\mathbb{C})$ acts transitively on $\widehat{\mathbb{C}}$ via fractional linear transformations as in Equation (2.5). Then, the differential of the transformation $g = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{SL}(2,\mathbb{C})$ at $z \in \widehat{\mathbb{C}}$ is given by

$$d\,g|_z = \frac{1}{(cz+d)^2}, \tag{2.28}$$

and for any $z \in \widehat{\mathbb{C}}$ and $\mathbf{v}_1,\, \mathbf{v}_2 \in T_z\widehat{\mathbb{C}}$ we have

$$\begin{aligned}
\big\langle [\,d\,g|_z\,]\,\mathbf{v}_1,\, [\,d\,g|_z\,]\,\mathbf{v}_2\big\rangle_{g\,z} &= \left\langle \frac{1}{(cz+d)^2}\,\mathbf{v}_1,\, \frac{1}{(cz+d)^2}\,\mathbf{v}_2 \right\rangle_{g\,z} \\
&= \frac{2}{\left(1+|g\,z|^2\right)^2}\,\frac{1}{|cz+d|^4}\,(\bar{\mathbf{v}}_1\mathbf{v}_2 + \mathbf{v}_1\bar{\mathbf{v}}_2) \\
&= \left[\frac{\left(1+|z|^2\right)^2}{\left(1+|g\,z|^2\right)^2\,|cz+d|^4}\right]\frac{2}{\left(1+|z|^2\right)^2}(\bar{\mathbf{v}}_1\mathbf{v}_2 + \mathbf{v}_1\bar{\mathbf{v}}_2) \\
&= \left[\frac{\left(1+|z|^2\right)^2}{\left(1+|g\,z|^2\right)^2\,|cz+d|^4}\right]\langle \mathbf{v}_1,\, \mathbf{v}_2\rangle_z
\end{aligned}$$

so elements of $\mathrm{SL}(2,\mathbb{C})$ are conformal diffeomorphisms on $\widehat{\mathbb{C}}$ with scale factor

$$\lambda_g^2(z) = \frac{(1+|z|^2)^2}{(1+|gz|^2)^2\,|cz+d|^4} = \frac{(1+|z|^2)^2}{\big(|az+b|^2+|cz+d|^2\big)^2}. \tag{2.29}$$

If $g = \left[\begin{smallmatrix} a & b \\ -\bar{b} & \bar{a}\end{smallmatrix}\right] \in \mathrm{SU}(2) \subset \mathrm{SL}(2,\mathbb{C})$, then

$$\begin{aligned}\Big(|az+b|^2+\big|\bar{b}z-\bar{a}\big|^2\Big)^2 &= (az+b)\big(\bar{a}\bar{z}+\bar{b}\big)+\big(\bar{b}z-\bar{a}\big)\big(b\bar{z}-a\big)\\ &= (|a|^2+|b|^2)\,(1+|z|^2)^2\\ &\overset{(2.7)}{=} (1+|z|^2)^2,\end{aligned}$$

which gives $\lambda_g^2(z) = 1$, so $\mathrm{SU}(2)$ is the group of isometries of $\widehat{\mathbb{C}}$.

Note that we can define the Hesssian of a function $\psi \in L^2(\widehat{\mathbb{C}},\mathbb{C})$ as a complex number whose coefficients are related to the covariant Hessian computed with respect to the round metric in Equation (2.26). The terms depend on the first and second partial derivatives of ψ in addition to the Christoffel symbols corresponding to the metric. At the origin, the terms depending on the first derivatives and Christoffel symbols vanish, and the Hessian becomes

$$\nabla d\,\psi\big|_0 = \frac{1}{4}\left(\frac{\partial^2\psi}{dz_1^2} - \frac{\partial^2\psi}{\partial z_2^2} - 2i\,\frac{\partial^2\psi}{\partial z_1\partial z_2}\right)\Bigg|_0, \tag{2.30}$$

with the Hessian of $g = \left[\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right] \in \mathrm{SL}(2,\mathbb{C})$ at the origin given by

$$\nabla d\,g\big|_0 = -\frac{2c}{d^3}. \tag{2.31}$$

$\mathbb{D}$ as a Riemannian manifold

The open unit disk $\mathbb{D} = \{z \in \mathbb{C} \mid |z| < 1\}$ is a Riemannian manifold under the hyperbolic metric, expressed at $z = z_1 + iz_2 \in \mathbb{D}$ as

$$s_z = \frac{4}{\big(1-|z|^2\big)^2}\,(dz_1^2+dz_2^2), \tag{2.32}$$

and in associating $T_z\mathbb{D}$ with $\mathbb{C}$ we can write

$$\langle \mathbf{v}_1,\ \mathbf{v}_2\rangle_z = \frac{2}{\left(1-|z|^2\right)^2}(\bar{\mathbf{v}}_1\mathbf{v}_2+\mathbf{v}_1\bar{\mathbf{v}}_2),$$

for all $\mathbf{v}_1,\ \mathbf{v}_2 \in T_z\mathbb{D}$. Here, the area measure is

$$dz = \frac{4\,dz_1\,dz_2}{\left(1-|z|^2\right)^2}, \tag{2.33}$$

with differential of a function $\psi \in L^2(\mathbb{D},\mathbb{C})$ defined in the same way as above.

SU$(1,1) \subset$ SL$(2,\mathbb{C})$ acts transitively on $\mathbb{D}$ via fractional linear transformations, and the differential of the transformation $g = \left[\begin{smallmatrix} a & b \\ \bar{b} & \bar{a}\end{smallmatrix}\right] \in$ SU$(1,1)$ is

$$d\,g|_z = \frac{1}{\left(\bar{b}z+\bar{a}\right)^2}.$$

Here, for any $z \in \mathbb{D}$ and $\mathbf{v}_1,\ \mathbf{v}_2 \in T_z\mathbb{D}$ we have

$$\begin{aligned}
\left\langle \left[\,d\,g|_z\,\right]\mathbf{v}_1,\ \left[\,d\,g|_z\,\right]\mathbf{v}_2\right\rangle_{g\,z} &= \left\langle \frac{1}{\left(\bar{b}z+\bar{a}\right)^2}\,\mathbf{v}_1,\ \frac{1}{\left(\bar{b}z+\bar{a}\right)^2}\,\mathbf{v}_2\right\rangle_{g\,z} \\
&= \frac{2}{\left(1-|g\,z|^2\right)^2}\,\frac{1}{\left|\bar{b}z+\bar{a}\right|^4}\,(\bar{\mathbf{v}}_1\mathbf{v}_2+\mathbf{v}_1\bar{\mathbf{v}}_2) \\
&= \frac{2}{\left(\left|\bar{b}z+\bar{a}\right|^2-\left|az+b\right|^2\right)^2}(\bar{\mathbf{v}}_1\mathbf{v}_2+\mathbf{v}_1\bar{\mathbf{v}}_2) \\
&= \frac{2}{(|a|^2-|b|^2)^2\,(1-|z|^2)^2}(\bar{\mathbf{v}}_1\mathbf{v}_2+\mathbf{v}_1\bar{\mathbf{v}}_2) \\
&\overset{(2.10)}{=} \frac{2}{(1-|z|^2)^2}(\bar{\mathbf{v}}_1\mathbf{v}_2+\mathbf{v}_1\bar{\mathbf{v}}_2) \\
&= \langle \mathbf{v}_1,\ \mathbf{v}_2\rangle_z,
\end{aligned}$$

so SU$(1,1)$ is the group of isometries of $\mathbb{D}$.

2.3 Representations by shift operators

A *representation* of a group G on a vector space V is a map from G to group of linear transformations over V

$$\begin{aligned} T : G &\to \mathrm{GL}(V) \\ g &\mapsto T(g) \end{aligned}$$

satisfying the property

$$T(g_1)T(g_2) = T(g_1 \circ g_2).$$

Here we are interested in representations of Lie group on spaces of functions. Given a Lie group G, Lie subgroup $H \subset G$, and homogeneous space $M = G/H$, G acts naturally on $L^2(M, \mathbb{C})$ by left shifts

$$g\,f = f \circ g^{-1}, \quad g \in G,\ f \in L^2(M, \mathbb{C}), \tag{2.34}$$

where $\circ$ denotes the composition of maps. Fixing g, we note that the map on $L^2(M, \mathbb{C})$ given by left-shifts is linear with

$$g(\alpha \cdot f_1 + \beta \cdot f_2) = \alpha \cdot \left[g\,f_1\right] + \beta \cdot \left[g\,f_2\right], \tag{2.35}$$

for all $f_1, f_2 \in L^2(M, \mathbb{C})$ and α, $\beta \in \mathbb{C}$. Furthermore for any $g_1, g_2 \in G$ and $f \in L^2(M, \mathbb{C})$ we have

$$g_1\left[g_2\,f\right] = g_1\left(f \circ g_2^{-1}\right) = f \circ g_2^{-1} \circ g_1^{-1} = (g_1 \circ g_2) f,$$

from which it follows that the action the action of G by left shifts forms a representation of G on $L^2(M, \mathbb{C})$.

Chapter 3

Extended Convolution on Homogeneous Spaces

In this chapter we present a general framework for transformation group-equivariant convolutions on arbitrary homogeneous spaces. Despite the pervasiveness of convolution and its twin correlation in vision and image processing, they have an inherent limitation: when convolving or correlating a signal with a filter, the filter remains fixed throughout operation, and cannot adapt to spatial information. We show how to remove this limitation by extending convolution with a frame operator, allowing the filter to adaptively transform as it travels over a domain.

We can make extended convolution equivariant to an arbitrary transformation group by observing that we only need to consider the lower-dimensional subgroup which transforms the positions of points as seen in the frames of their neighbors. By defining the frame operator in a manner that is equivariant under the action of the subgroup, we can align the filter to correct for the change in the relative positions. To compute convolutions, input features are mapped to a density distribution controlling for the change in area measure induced by the transformation, and integrated against the aligned filters over the homogeneous space, rather than the group itself.

3.1 Describing relative positions

To motivate our approach, we'll begin by recalling the familiar definition of convolution in the plane. The convolution of a signal $\psi \in L^2(\mathbb{C}, \mathbb{C})$ with a filter $\psi \in L^2(\mathbb{C}, \mathbb{C})$ is the function in $L^2(\mathbb{C}, \mathbb{C})$ with

$$(\psi * f)(z) = \int_{\mathbb{C}} \psi(y)\, f(z - y)\, dy. \tag{3.1}$$

We can see that the operation is inherently localized – filters distribute their weights with respect to parameterizations of the plane about different points. At each point z in the convolution of ψ with f in Equation (3.1), the value of ψ at each neighboring point y is weighted by the value of the filter at the *relative position* of z as seen in the frame at y, which is given by $z - y$. To extend the concept of filter localization to arbitrary homogeneous spaces, we need a generalized description of the relative positions between points.

3.1.1 Generalized logarithm and exponential

Suppose we are given a Lie group G and an associated homogeneous space M such that $M \cong G/H_0$, where $H_0 \subset G$ is the stabilizer subgroup of some arbitrary choice of origin $0 \in M$. Our goal is to define a parameterization of M about a given point $p \in M$. Recalling that M can be viewed as a Riemannian manifold under a given metric s, the Riemannian logarithm map (§2.2) initially appears to be a good candidate as it provides a natural notion of local parameterizations. However, it has several important drawbacks. It is only locally invertible and is inflexible in the sense that it can only be defined relative to the shortest, rather than arbitrary, geodesics. Generally speaking, these limitations are workable, and we will later use the logarithm map to define extended convolution and correlation on arbitrary 2D Riemannian manifolds. That said, while any homogeneous space can be viewed as a

Riemannian manifold, the converse is false. Here, we can leverage the added structure induced by viewing M as a homogeneous space to define a more flexible notion of the logarithm map which we call the generalized logarithm.

Formally, we define the *generalized logarithm* to be a lifting map

$$\begin{aligned} \mathrm{Log} : M &\longrightarrow G \\ p &\longmapsto \mathrm{Log}_p \end{aligned} \qquad \text{satisfying} \qquad \widetilde{\pi}\left(\mathrm{Log}_p^{-1}\right) = p, \;\; \forall p \in M, \tag{3.2}$$

where $\mathrm{Log}_p^{-1} \in G$ denotes the inverse element of $\mathrm{Log}_p \in G$ and $\widetilde{\pi} : G \longrightarrow M$ is the natural map from G to M as defined in Equation (2.3). It follows that for all $p \in M$

$$\mathrm{Log}_p^{-1} 0 = \mathrm{Log}_p^{-1} \widetilde{\pi}(e) = \widetilde{\pi}\left(\mathrm{Log}_p^{-1} e\right) = \widetilde{\pi}\left(\mathrm{Log}_p^{-1}\right) \overset{(3.2)}{=} p,$$

so

$$\mathrm{Log}_p\, p = 0, \, \forall p \in M.$$

Then, for any point $q \in M$, we can express the “position” of q in the frame of p as $\mathrm{Log}_p\, q \in M$. By analogy to Riemannian geometry, the generalized logarithm maps M to the “tangent space” at p. We make this explicit by denoting the image of the logarithm map at p as M_p:

$$\mathrm{Log}_p : M \longrightarrow M_p,$$

though formally M and M_p are the same space. Similarly, we define the *generalized exponential* at p as the inverse of the generalized logarithm,

$$\mathrm{Exp}_p \equiv \mathrm{Log}_p^{-1} : M_p \longrightarrow M,$$

mapping the “tangent space” at $p \in M$ to the “base space” M. We note that the transitive action of G on M ensures that at any point $p \in M$, there exists $g \in G$ such that $\widetilde{\pi}(g^{-1}) = p$, so it is always possible to define $\mathrm{Log} : M \longrightarrow G$.

We use the notation Log and Exp to distinguish the generalized logarithm and exponential from the Riemannian logarithm and exponential maps log and exp. Unlike the latter, Log_p and Exp_p are global bijections on M under the assumption that

G is a transformation group. Furthermore, the generalized logarithm and exponential are not necessarily uniquely defined nor continuous, though we will see that this is not a requirement for extended convolution to be well-defined.

3.1.2 Action of the origin-stabilizing subgroup

Following the above discussion, we will think of convolutional filters as functions defined on a canonical "tangent space" describing the weight with which a point contributes to its neighbor in terms of the position of the neighbor in the frame of that point. Our goal is to define a convolution operator on M equivariant to transformations in G. To this end we need to understand how the position of a point in the frame of its neighbor changes under the action of a transformation $g \in G$.

Describing the transformation from one coordinate frame to the other is straightforward: Beginning in M_p, we 1). map to M by applying Exp_p, 2). transform M by g, and 3). map back to M_{gp} using Log_{gp}. Composing these gives the transformation,

$$D_p^g \equiv \mathrm{Log}_{gp} \circ\ g \circ \mathrm{Exp}_p : M_p \longrightarrow M_{gp} \in G, \tag{3.3}$$

with the notation chosen to reflect dependence on both p and g, equivalently represented in the diagram

$$\begin{tikzcd}
M \arrow[r, "\mathrm{Log}_p"] \arrow[d, "g"'] & M_p \arrow[d, "D_p^g"] \\
M \arrow[r, "\mathrm{Log}_{gp}"] & M_{gp}
\end{tikzcd}.$$

Rearranging terms, it's easy to see that for all $q \in M$,

$$D_p^g \,\mathrm{Log}_p\, q = \mathrm{Log}_{gp}\, gq. \tag{3.4}$$

From Equation (3.3) and the definitions of the generalized logarithm and exponential, it can be shown that $D_p^g\, 0 = 0$ for all $p \in M$ as

$$D_p^g\, 0 \overset{(3.3)}{=} (\mathrm{Log}_{gp} \circ\ g \circ \mathrm{Exp}_p)\, 0 \overset{(3.2)}{=} (\mathrm{Log}_{gp} \circ g)\, p = \mathrm{Log}_{gp}\, gp \overset{(3.2)}{=} 0.$$

Thus, D_p^g must belong to the origin-preserving subgroup $H_0 \subset G$. Intuitively, this follows from the facts that p maps to gp under the action of g and that both p and gp are the origin in their respective tangent spaces. This simple but critical observation implies that in defining equivariant convolution and correlation on a homogeneous space $M \cong G/H_0$, we only need to consider the action of the origin-preserving subgroup H_0, rather than the full group G.

3.2 Extended convolution

Given a homogeneous space $M \cong G/H_0$, we view M as a Riemannian manifold under a given metric s, which defines an area measure dp at each point $p \in M$ as in Equation (2.15). We implement extended convolution by shifting a filter over the M, aligning the shifted filter using a frame field, and distributing the values of a density function to neighboring points, with distribution weights given by the aligned filter.

3.2.1 The frame and density operators

Extended convolution is defined with respect to a *frame operator* $\mathfrak{T}$ and *density operator* ρ that take in a function $\psi \in L^2(M, \mathbb{C})$ and return a frame field taking values in H_0 and a real- or complex-valued density field,

$$\begin{aligned} \mathfrak{T} : L^2(M, \mathbb{C}) &\to L^2(M, H_0) \\ \psi &\mapsto \mathfrak{T}_\psi \end{aligned} \qquad \text{and} \qquad \begin{aligned} \rho : L^2(M, \mathbb{C}) &\to L^2(M, \mathbb{C}) \\ \psi &\mapsto \rho_\psi \end{aligned} \tag{3.5}$$

Given a transformation $g \in G$ and a point $p \in M$, the frame operator corrects for the change in relative position resulting from g, as characterized by the origin-preserving transformation D_p^g from Equation (3.3). Similarly, the density operator adjusts for the change in the area measure used for integration, given by the scale factor $\lambda_g^2(p)$ as in Equation (2.18).

3.2.2 Extending convolution

Given operators $\mathfrak{T}$ and ρ, the *extended convolution* of a function ψ with with a filter f, both in $L^2(M,\mathbb{C})$, is formally expressed as the function in $L^2(M,\mathbb{C})$ with

$$(\psi * f)(p) = \int_M \rho_\psi(q) \left[\mathfrak{T}_\psi(q)\, f\right]\left(\operatorname{Log}_q p\right) dq, \tag{3.6}$$

where $\mathfrak{T}_\psi(q)\, f \equiv f \circ \left[\mathfrak{T}_\psi(p)\right]^{-1}$ denotes the standard action of transformations on f by left shifts. That is, to get the value at a point $p \in M$, we iterate over all neighbors q, compute the position of p in the frame $\mathfrak{T}_\psi(q)$ at q, evaluate the filter at that point, and accumulate the density $\rho_\psi(q)$ at q weighted by the filter value.

3.2.3 Equivariance

The transformation operator $\mathfrak{T}$ and the density operator ρ must satisfy certain conditions to ensure that extended convolutions are equivariant to transformations in G; *i.e.* that for any function ψ, filter f, and transformation $g \in G$, extended convolution commutes with the action of g by left shifts,

$$g\,(\psi * f) = (g\,\psi * f) \tag{3.7}$$

Claim 1 (Conditions for equivariance)**.** *If for all $\psi \in L^2(M,\mathbb{C})$ and $g \in G$, the operators $\mathfrak{T}$ and ρ satisfy*

$$D_p^g \circ \mathfrak{T}_\psi(p) = \mathfrak{T}_{g\,\psi}(gp) \qquad \textit{and} \qquad \lambda_g^{-2}(p)\, \rho_\psi(p) = \rho_{g\,\psi}(gp) \tag{3.8}$$

for all $p \in M$, then for any filter $f \in L^2(M,\mathbb{C})$ Equation (3.7) holds.

Proof. Suppose $\mathfrak{T}$ and ρ satisfy the condition and consider any $\psi \in L^2(M,\mathbb{C})$ and $g \in G$. For any filter $f \in L^2(M,\mathbb{C})$ and $q \in M$, we can relate the expression of the filter over M_q to the expression of the filter over M_{gq}:

$$\begin{aligned}\left[\mathfrak{T}_\psi(q)\, f\right] \circ \mathrm{Log}_q &= f \circ \left[\mathfrak{T}_\psi(q)\right]^{-1} \circ \mathrm{Log}_q \\ &\overset{(3.8)}{=} f \circ \left[\mathfrak{T}_{g\,\psi}(gq)\right]^{-1} \circ D_q^g \circ \mathrm{Log}_q \\ &\overset{(3.3)}{=} f \circ \left[\mathfrak{T}_{g\,\psi}(gq)\right]^{-1} \circ \mathrm{Log}_{gq} \circ g \\ &= \left[\mathfrak{T}_{g\,\psi}(gq)\, f\right] \circ \mathrm{Log}_{gq} \circ g. \end{aligned} \tag{3.9}$$

Using the relationship between the expression of the filters over M_q and M_{gq} it follows that for any $p \in M$,

$$\begin{aligned}\left[g\,(\psi * f)\right](p) &\overset{(3.6)}{=} \int_M \rho_\psi(q) \left[\mathfrak{T}_\psi(q)\, f\right]\left(\mathrm{Log}_q\, g^{-1} p\right) dq \\ &\overset{(3.9)}{=} \int_M \rho_\psi(q) \left[\mathfrak{T}_{g\,\psi}(gq)\, f\right]\left(\mathrm{Log}_{gq}\, p\right) dq \\ &\overset{(3.8)}{=} \int_M \lambda_g^2(q)\, \rho_{g\,\psi}(gq) \left[\mathfrak{T}_{g\,\psi}(gq)\, f\right]\left(\mathrm{Log}_{gq}\, p\right) dq \\ &= \int_M \rho_{g\,\psi}(q') \left[\mathfrak{T}_{g\,\psi}(q')\, f\right]\left(\mathrm{Log}_{q'}\, p\right) dq' \\ &\overset{(3.6)}{=} (g\,\psi * f)(p), \end{aligned} \tag{3.10}$$

where the second to last equality follows from the change of variables

$$\begin{aligned} q &\mapsto gq \\ dq &\overset{(2.18)}{\mapsto} \lambda_g^2(q)\, dq. \end{aligned} \tag{3.11}$$

Thus,

$$g\,(\psi * f) = (g\,\psi * f),$$

as desired. □

In practice, rather than have the user provide the frame and density operators directly, they can be derived from an input signal ψ in such a way so as to satisfy the conditions in Equation (3.8). However, this technique is a bit of an art and depends strongly on the specific transformation group G, stabilizer subgroup H_0, and homogeneous space $M \cong G/H_0$ of interest. As a result, it is not readily generalizeable to arbitrary homogeneous spaces and specific examples of its application to several domains of interest are provided in the latter half of this chapter.

3.2.4 Locally-supported filters

Extended convolution defined in Equation (3.6) implicitly assumes global filter support as integration is performed over the whole of M. However, in certain applications global support can be undesirable and in what follows we define a notion of locally-supported extended convolution.

Let $\mathcal{N}_0 \subset M$ be an open subset containing $0 \in M$. Given operators $\mathfrak{T}$ and ρ as in Equation (3.5), we can define the extended convolution of a function $\psi \in L^2(M, \mathbb{C})$ with a compactly-supported filter $f \in L^2(\mathcal{N}_0, \mathbb{C})$ by restricting the domain of integration at each point $p \in M$ to the region

$$\mathcal{N}_p^{\psi} = \left\{ q \in M \,\middle|\, p \in \left[\operatorname{Exp}_q \circ \mathfrak{T}_{\psi}(q)\right] \mathcal{N}_0 \right\}, \tag{3.12}$$

such that

$$(\psi * f)(p) = \int_{\mathcal{N}_p^{\psi}} \rho_{\psi}(q) \left[\mathfrak{T}_{\psi}(q)\, f\right] \left(\operatorname{Log}_q p\right) dq, \tag{3.13}$$

where $\left[\operatorname{Exp}_q \circ \mathfrak{T}_{\psi}(q)\right] \mathcal{N}_0$ is the image of $\mathcal{N}_0$ under the transformation $\operatorname{Exp}_q \circ \mathfrak{T}_{\psi}(q) \in G$. Note that for any $p \in M$ and $q \in \mathcal{N}_p^{\psi}$, it follows from Equation (3.12) that there exists $n \in \mathcal{N}_0$ such that $p = \left(\operatorname{Exp}_q \circ \mathfrak{T}_{\psi}(q)\right) n$ so $\left[\mathfrak{T}_{\psi}(q)\right]^{-1} \operatorname{Log}_q p = n$. Thus, $\left[\mathfrak{T}_{\psi}(q)\right]^{-1} \operatorname{Log}_q p \in \mathcal{N}_0$ for all $p \in M$ and $q \in \mathcal{N}_p^{\psi}$ so the convolution is well defined.

We now show that locally-supported extended convolution is equivariant in the sense of Equation (3.7) under the conditions on the operators in Equation (3.8).

Claim 2 (Equivariance of locally-supported extended convolution)**.** *Given an open subset $\mathcal{N}_0 \subset M$ containing $0 \in M$, if for all $\psi \in L^2(M, \mathbb{C})$ and $g \in G$, the operators $\mathfrak{T}$ and ρ satisfy the conditions in Equation (3.8) for all $p \in M$, then for any filter $f \in L^2(\mathcal{N}_0, \mathbb{C})$ Equation (3.7) holds for locally-supported extended convolution in Equation (3.13).*

Proof. Note that for any $g \in G$ and $p \in M$,

$$\begin{aligned}
\mathcal{N}^{\psi}_{g^{-1}p} &\overset{(3.12)}{=} \left\{ q \in M \,\middle|\, g^{-1}p \in \left[\mathrm{Exp}_q \circ \mathfrak{T}_{\psi}(q)\right] \mathcal{N}_0 \right\} \\
&= \left\{ q \in M \,\middle|\, p \in \left[g \circ \mathrm{Exp}_q \circ \mathfrak{T}_{\psi}(q)\right] \mathcal{N}_0 \right\} \\
&\overset{(3.3)}{=} \left\{ q \in M \,\middle|\, p \in \left[\mathrm{Exp}_{gq} \circ D^g_q \circ \mathfrak{T}_{\psi}(q)\right] \mathcal{N}_0 \right\} \\
&\overset{(3.8)}{=} \left\{ q \in M \,\middle|\, p \in \left[\mathrm{Exp}_{gq} \circ \mathfrak{T}_{g\psi}(gq)\right] \mathcal{N}_0 \right\} \\
&= g^{-1} \left\{ gq \in M \,\middle|\, p \in \left[\mathrm{Exp}_{gq} \circ \mathfrak{T}_{g\psi}(gq)\right] \mathcal{N}_0 \right\} \\
&\overset{(3.12)}{=} g^{-1}\, \mathcal{N}^{g\psi}_p .
\end{aligned} \tag{3.14}$$

Then, following the proof of Claim 1, we have

$$\begin{aligned}
\left[g\,(\psi * f)\right](p) &\overset{(3.13)}{=} \int_{\mathcal{N}^{\psi}_{g^{-1}p}} \rho_{\psi}(q) \left[\mathfrak{T}_{\psi}(q)\, f\right]\left(\mathrm{Log}_q\, g^{-1}p\right) dq \\
&\overset{(3.10)}{=} \int_{\mathcal{N}^{\psi}_{g^{-1}p}} \lambda^2_g(q)\, \rho_{g\,\psi}(gq) \left[\mathfrak{T}_{g\,\psi}(gq)\, f\right]\left(\mathrm{Log}_{gq}\, p\right) dq \\
&\overset{(3.14)}{=} \int_{g^{-1}\,\mathcal{N}^{g\psi}_p} \lambda^2_g(q)\, \rho_{g\,\psi}(gq) \left[\mathfrak{T}_{g\,\psi}(gq)\, f\right]\left(\mathrm{Log}_{gq}\, p\right) dq \\
&\overset{(3.11)}{=} \int_{\mathcal{N}^{g\psi}_p} \rho_{g\,\psi}(q') \left[\mathfrak{T}_{g\,\psi}(q')\, f\right]\left(\mathrm{Log}_{q'}\, p\right) dq' \\
&\overset{(3.13)}{=} (g\,\psi * f)(p).
\end{aligned}$$

□

Note that the definition of $\mathcal{N}^{\psi}_p$ in Equation (3.12) is dynamic in the sense that the set of points contributing to the value of the convolution at each point $p \in M$

depends on ψ. In applications such as CNNs where convolutions are computed in succession, this induces additional computational overhead, as the support at each point must be recalculated for every ψ. However, in the special case that Log_p and $\mathfrak{T}_\psi(p)$ are isometries for all $p \in M$ and $\psi \in L^2(M, \mathbb{C})$, the following claim holds.

Claim 3 (Invariant support). *Let M homogeneous space such that $M \cong G/H_0$, for some transformation group G and origin-stabilizing subgroup $H_0 \subset G$. Then if every transformation $h \in H_0$ is an isometry on M and* $\mathrm{Log} : M \to G$ *is defined such that at each point $p \in M$,* $\mathrm{Log}_p \in G$ *is also an isometry, then for all $p \in M$ and $\psi \in L^2(M, \mathbb{C})$*

$$\mathcal{N}_p^\psi = \mathcal{N}_p, \tag{3.15}$$

where $\mathcal{N}_p$ denotes the geodesic ball of radius ε about p.

Proof. Suppose that for each point $p \in M$ and function $\psi \in L^2(M, \mathbb{C})$, both $\mathrm{Log}_p \in G$ and $\mathfrak{T}_\psi(p) \in H_0$ are isometries on M (the latter following from the assumption that all transformations in H_0 are isometries). We now show that $\mathcal{N}_p^\psi = \mathcal{N}_p$ for all $p \in M$ and $\psi \in L^2(M, \mathbb{C})$.

Formally, for any point $p \in M$, $\mathcal{N}_q \subset M$ is the set

$$\mathcal{N}_q = \{q' \in M \mid d\,(q',\, q) < \varepsilon\}, \tag{3.16}$$

where $d\,(\,\cdot\,,\,\cdot\,)$ is the geodesic distance. It follows that for any $\psi \in L^2(M, \mathbb{C})$, the image of $\mathcal{N}_0$ under $\mathrm{Exp}_q \circ \mathfrak{T}_\psi(q)$ is $\mathcal{N}_q$ as

$$\begin{aligned}
\left[\mathrm{Exp}_q \circ \mathfrak{T}_\psi(q)\right] \mathcal{N}_0 &\overset{(3.16)}{=} \left\{ \left[\mathrm{Exp}_q \circ \mathfrak{T}_\psi(q)\right] q' \in M \,\middle|\, d\,(q',\, 0) < \varepsilon \right\} \\
&= \left\{ q' \in M \,\middle|\, d\left(\left[\mathrm{Exp}_q \circ \mathfrak{T}_\psi(q)\right]^{-1} q',\, 0\right) < \varepsilon \right\} \\
&= \left\{ q' \in M \,\middle|\, d\left(q',\, \left[\mathrm{Exp}_q \circ \mathfrak{T}_\psi(q)\right] 0\right) < \varepsilon \right\} \\
&\overset{(3.2)}{=} \{\, q' \in M \mid d\,(q',\, q) < \varepsilon\} \\
&\overset{(3.16)}{=} \mathcal{N}_q,
\end{aligned}$$

where the third equality follows from the fact that isometries preserve distances and the fourth from the fact that $\mathfrak{T}_\psi(q)$ preserves the origin. Thus, for any $p \in M$

$$\begin{aligned} \mathcal{N}_p^\psi &\overset{(3.12)}{=} \left\{ q \in M \,\middle|\, p \in \left[\mathrm{Exp}_q \circ \mathfrak{T}_\psi(q)\right] \mathcal{N}_0 \right\}, \\ &= \{ q \in M \mid p \in \mathcal{N}_q \} \\ &\overset{(3.16)}{=} \{ q \in M \mid d(p, q) < \varepsilon \} \\ &\overset{(3.16)}{=} \mathcal{N}_p, \end{aligned}$$

with the final equality following from the fact that $d(p, q) = d(q, p)$.

□

3.2.5 Optimal filters

Generally speaking, the matching of feature descriptors can be characterized as a feature detection problem. Given an arbitrary homogeneous space $M \cong G/H_0$ and a function $\psi \in L^2(M, \mathbb{R})$, we can use extended convolution to describe the region about an arbitrary point $p \in M$ by designing a filter that will maximize the response at p.

Given a function $\psi \in L^2(M, \mathbb{R})$ and fixing the maps $\mathfrak{T}$ and ρ, extended convolution and can be thought of as a map from the space of real-valued filters on M to the space of real-valued functions on M,

$$\begin{aligned} \mathcal{E}_\psi : L^2(M, \mathbb{R}) &\longrightarrow L^2(M, \mathbb{R}) \\ f &\overset{(3.6)}{\mapsto} \psi * f \end{aligned}$$

From the definition of extended convolution and the linearity of left-shifts – Equation (2.35) – it is clear that $\mathcal{E}_\psi$ is linear in f, as is the map

$$\begin{aligned} \mathcal{E}_\psi^p : L^2(M, \mathbb{R}) &\longrightarrow \mathbb{R} \\ f &\overset{(3.6)}{\mapsto} (\psi * f)(p) \end{aligned}$$

obtained by evaluating the function returned by $\mathcal{E}_\psi$ at a point $p \in M$. Using the fact that the space of filters $L^2(M, \mathbb{R})$ is an inner product space under the inner product defined in Equation (2.17) and applying the Riesz Representation Theorem, there exists a filter $f_\psi^p \in L^2(M, \mathbb{R})$ such that

$$\mathcal{E}_\psi^p(f) \equiv \langle\, f,\, f_\psi^p \,\rangle \tag{3.17}$$

for all $f \in L^2(M, \mathbb{R})$. In particular, up to scale, f_ψ^p is exactly the filter that maximizes the response of the extended convolution at p.

To evaluate the optimal filter f_ψ^p at arbitrary points $x \in M$, we can compute its inner product with a delta function centered at x, δ_x. As shown above, this is equivalent to the evaluation of the extended convolution of ψ with δ_x at the point p:

$$f_\psi^p(x) = \langle\, \delta_x,\, f_\psi^p \,\rangle \overset{(3.17)}{=} \mathcal{E}_\psi^p(\delta_x) = (\psi * \delta_x)(p).$$

In practice, we would like descriptions of the keypoint p to be local. To this end we note that if we restrict ourselves to filters that are supported within a geodesic ball $\mathcal{N}_0 \subset M$ of radius ε centered at $0 \in M$, the filter maximizing the response of the extended convolution at p is still, up to scale, f_ψ^p restricted to $\mathcal{N}_0$, with

$$f_\psi^p(x) \overset{(3.13)}{=} \int_{\mathcal{N}_p^\psi} \rho_\psi(q)\, \left[\mathfrak{T}_\psi(q)\, \delta_x\right]\left(\mathrm{Log}_q\, p\right) dq, \tag{3.18}$$

where $\mathcal{N}_p^\psi$ is defined as in Equation (3.12).

Invariance

From the extended convolution, the optimal filters inherit the desirable property of *invariance* to transformations in G, *i.e.* for any $\psi \in L^2(M, \mathbb{C})$ and $p \in M$,

$$f_\psi^p = f_{g\psi}^{gp}, \tag{3.19}$$

for all $g \in G$.

Claim 4 (Invariance of optimal filters). *If for all $\psi \in L^2(M, \mathbb{C})$ and $g \in G$, the operators $\mathfrak{T}$ and ρ satisfy the conditions in Equation (3.8) for all $p \in M$, then for any $\psi \in L^2(M, \mathbb{C})$, $g \in G$, and $p \in M$, Equation (3.19) holds.*

Proof. Given $\mathfrak{T}$ and ρ satisfying the conditions in Equation (3.8), the proof follows trivially from Claim 2: For any $\psi \in L^2(M, \mathbb{C})$, $g \in G$, and $p \in M$, we have

$$f^p_\psi(x) \overset{(3.18)}{=} (\psi * \delta_x)(p) \overset{\text{Claim 2}}{=} (g\psi * \delta_x)(gp) \overset{(3.18)}{=} f^{gp}_{g\psi}(x),$$

for all $x \in \mathcal{N}_0$. □

3.3 Realization on the canonical domains

We conclude this chapter making the general theory presented in §3.1–3.2 concrete by realizing equivariant extended convolutions on the plane $\mathbb{C}$, the Riemann sphere $\widehat{\mathbb{C}}$, and the disk $\mathbb{D}$.

3.3.1 Extended convolution on $\mathbb{C}$

Recalling that $\mathrm{SE}(2)$ is the group of isometries of the plane, viewing $\mathbb{C}$ as the homogeneous space $\mathrm{SE}(2)/\mathrm{U}(1)$ allows us to define isometry-equivariant extended convolutions and correlations.

On $\mathbb{C}$ we define the generalized logarithm at $z \in \mathbb{C}$ as element of the translational subgroup of both $\mathrm{SE}(2)$ taking z to $0 \in \mathbb{C}$,

$$\mathrm{Log}_z \equiv \begin{bmatrix} 1 & -z \\ 0 & 1 \end{bmatrix}. \tag{3.20}$$

It is easy to see that for $\kappa : \mathrm{SE}(2)/\mathrm{U}(1) \longrightarrow \mathbb{C}$ as defined in Equation (2.4), the definition of Log_z in Equation (3.20) satisfies the condition in Equation (3.2) with

$$\widetilde{\pi}\left(\mathrm{Log}_z^{-1}\right) = z.$$

Note that for any $z, y \in \mathbb{C}$,

$$\operatorname{Log}_z y = y - z,$$

which to the classical definition of the relative position of y with respect to z in the plane.

Here the frame and density operators are maps

$$\mathfrak{T} : L^2(\mathbb{C}, \mathbb{C}) \to L^2(\mathbb{C}, \mathrm{U}(1)) \qquad \text{and} \qquad \rho : L^2(\mathbb{C}, \mathbb{C}) \to L^2(\mathbb{C}, \mathbb{C}).$$

From Equation (3.6), the general form of the extended convolution of a function ψ with a filter f, both in $L^2(\mathbb{C}, \mathbb{C})$, is

$$\begin{aligned} (\psi * f)(y) &\overset{(3.6)}{=} \int_{\mathbb{C}} \rho_\psi(z) \left[\mathfrak{T}_\psi(z)\, f\right]\left(\operatorname{Log}_z y\right) dz \\ &\overset{(3.20)}{=} \int_{\mathbb{C}} \rho_\psi(z) \left[\mathfrak{T}_\psi(z)\, f\right](y - z)\, dz, \end{aligned} \tag{3.21}$$

with dz the standard area measure under the Euclidean metric as in Equation (2.23).

Construction of Operators

Recall that for the extended convolution to be equivariant under the action of transformations $g \in \mathrm{SE}(2)$ in the sense of Equation (3.7), $\mathfrak{T}$ and ρ must satisfy the conditions in Equation (3.8) for $M = \mathbb{C}$, $G = \mathrm{SE}(2)$, and $D_z^g \in \mathrm{U}(1)$.

Claim 5 (Planar frame fields)**.** *If $\mathfrak{T}$ is defined as*

$$\mathfrak{T}_\psi(x) \equiv \begin{bmatrix} \operatorname{sgn} \overline{d \operatorname{Log}_x \psi\big|_0} & 0 \\ 0 & 1 \end{bmatrix}, \tag{3.22}$$

where $d \operatorname{Log}_x \psi\big|_0 \in \mathbb{C}$ is the differential of $\operatorname{Log}_x \psi$ evaluated at the origin and sgn is the complex signum function $z \mapsto z\,|z|^{-1}$, then the condition for the frame operator in Equation (3.8) is satisfied.

Proof. For any $\psi \in L^2(\mathbb{C}, \mathbb{C})$ and $g \in \mathrm{SE}(2)$ it follows from Equation (3.3) that for any point $x \in \mathbb{C}$ the transformation $D_x^g : \mathbb{C}_x \to \mathbb{C}_{gx}$ describing the deformation of the tangent space at x is an element of the origin-preserving subgroup $\mathrm{U}(1)$, and we denote $D_x^g = \left[\begin{smallmatrix} e^{i\theta} & 0 \\ 0 & 1 \end{smallmatrix}\right]$ for some $\theta \in [0, 2\pi)$. Furthermore,

$$(\mathrm{Log}_{gx} \circ g)\,\psi \overset{(3.3)}{=} (D_x^g \circ \mathrm{Log}_x)\psi,$$

for all $x \in \mathbb{C}$, and applying the chain rule and evaluating at the origin using Equation (2.25) gives

$$d\,(\mathrm{Log}_{gx} \circ g)\,\psi\big|_0 = e^{-i\theta}\left[d\,\mathrm{Log}_x \psi\big|_0\right]. \tag{3.23}$$

Then,

$$\begin{aligned}
\mathfrak{T}_{g\psi}(gx) &\overset{(3.22)}{=} \begin{bmatrix} \mathrm{sgn}\,\overline{d\,(\mathrm{Log}_{gx} \circ g)\,\psi\big|_0} & 0 \\ 0 & 1 \end{bmatrix} \\
&\overset{(3.23)}{=} \begin{bmatrix} e^{i\theta}\,\mathrm{sgn}\,\overline{d\,\mathrm{Log}_x\,\psi\big|_0} & 0 \\ 0 & 1 \end{bmatrix} \\
&= \underbrace{\begin{bmatrix} e^{i\theta} & 0 \\ 0 & 1 \end{bmatrix}}_{=\ D_x^g} \underbrace{\begin{bmatrix} \mathrm{sgn}\,\overline{d\,\mathrm{Log}_x\,\psi\big|_0} & 0 \\ 0 & 1 \end{bmatrix}}_{\overset{(3.22)}{=}\ \mathfrak{T}_\psi(x)}
\end{aligned}$$

as desired. □

Since elements of $\mathrm{SE}(2)$ are isometries of the plane, we have $\lambda_g^2(z) = 1$ for all $g \in \mathrm{SE}(2)$ and $z \in \mathbb{C}$. This gives significant flexibility in how we define the density operator, however the definition of $\mathfrak{T}_\psi$ in Equation (3.22) motivates defining ρ as

$$\rho_\psi(x) \equiv \left|d\,\mathrm{Log}_x\,\psi\big|_0\right|. \tag{3.24}$$

This way, at a point where $\mathfrak{T}_\psi(x)$ is ill-defined – those at which $d\,\mathrm{Log}_x \psi\big|_0$ vanishes – $\rho_\psi(x)$ also vanishes and the point contributes nothing to the extended convolution.

We also note that for any $x \in \mathbb{C}$, the differential of the generalized exponential $\mathrm{Exp}_x = \mathrm{Log}_x^{-1}$ is constant with

$$d\,\mathrm{Exp}_x = 1,$$

and that in the plane, the differential of a function is equivalent to the gradient, since the Euclidean metric is the identity. Thus $\mathfrak{T}_\psi$ and ρ_ψ as defined in Equations (3.22) and (3.24) can be equivalently expressed as

$$\mathfrak{T}_\psi(x) = \begin{bmatrix} \operatorname{sgn} \overline{\nabla\psi|_x} & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{and} \qquad \rho_\psi(x) = \big|\nabla\psi|_x\big|, \tag{3.25}$$

where $\nabla\psi$ denotes the gradient of ψ.

Optimal filters

Assuming filters are compactly-supported on an ε-disk $\mathcal{N}_0$ about the origin, then for any function $\psi \in L^2(\mathbb{C}, \mathbb{R})$, at each point $y \in \mathbb{C}$ extended convolution is locally-supported on $\mathcal{N}_y^\psi = \mathcal{N}_y$ since Log_z and elements of $\mathrm{U}(1)$ are isometries of the plane. Here, the filter maximizing the response of the planar extended convolution at a given point $y \in \mathbb{C}$ is given by

$$f_\psi^y(x) \overset{(3.18)}{=} \int_{\mathcal{N}_y} \rho_\psi(z) \left[\mathfrak{T}_\psi(z)\, \delta_x\right](y - z)\, dz. \tag{3.26}$$

It follows from Claims 4 and 5 that f_ψ^y is invariant under transformations in $\mathrm{SE}(2)$ in the sense of Equation (3.19) if $\mathfrak{T}$ and ρ are defined as in Equations (3.22) and (3.24) (or equivalently Equation (3.25))

3.3.2 Extended convolution on $\widehat{\mathbb{C}}$

Viewing $\widehat{\mathbb{C}}$ as the homogeneous space $\mathrm{SL}(2, \mathbb{C})/\mathrm{L}$, we define conformally-equivariant extended convolutions under the action of $\mathrm{SL}(2, \mathbb{C})$.

Note that the choice of generalized logarithm in Equation (3.20) is ill-defined at $\infty \in \widehat{\mathbb{C}}$. Instead, we define the generalized logarithm at $z \in \widehat{\mathbb{C}}$ as a rotation $\mathrm{Log}_z \in \mathrm{SU}(2)$ taking z to the origin,

$$\mathrm{Log}_z \equiv \frac{1}{|c|\sqrt{1 + |z|^2}} \begin{bmatrix} c & -cz \\ \bar{c}\bar{z} & \bar{c} \end{bmatrix}, \tag{3.27}$$

where $c \in \widehat{\mathbb{C}}$ is arbitrary. This definition of the generalized logarithm is well-defined at $z = \infty$ with $\mathrm{Log}_\infty = \frac{1}{|c|}\left[\begin{smallmatrix} 0 & -c \\ \bar{c} & 0 \end{smallmatrix}\right]$, and satisfies the condition in Equation (3.2):

$$\widetilde{\pi}\left(\mathrm{Log}_z^{-1}\right) \overset{(3.27)}{=} \widetilde{\pi}\left(\frac{1}{|c|\sqrt{1+|z|^2}}\begin{bmatrix} \bar{c} & cz \\ -\bar{c}\bar{z} & c \end{bmatrix}\right) \overset{(2.6)}{=} z,$$

with the second equality following from the fact that the conjugate transpose is equivalent to the inverse of an element in $\mathrm{SU}(2)$. While any choice of c in the definition of the generalized logarithm gives a rotation taking z to the origin, here we set $c = \sqrt{\bar{z}}$ which ensures that the great circle going through the origin and z is mapped to the real line, enabling the use of the fast Spherical Harmonic Transform [DH94, KR08] in Chapter 7.

The frame and density operators are maps

$$\mathfrak{T} : L^2(\widehat{\mathbb{C}}, \mathbb{C}) \longrightarrow L^2(\widehat{\mathbb{C}}, \mathrm{L}) \qquad \text{and} \qquad \rho : L^2(\widehat{\mathbb{C}}, \mathbb{C}) \longrightarrow L^2(\widehat{\mathbb{C}}, \mathbb{C}),$$

and the extended convolution of a function ψ with a filter f, both in $L^2(\widehat{\mathbb{C}}, \mathbb{C})$, is

$$(\psi * f)(y) \overset{(3.6)}{=} \int_{\widehat{\mathbb{C}}} \rho_\psi(z)\left[\mathfrak{T}_\psi(z)\, f\right]\left(\mathrm{Log}_z\, y\right) dz, \tag{3.28}$$

with dz the area measure under the round metric as in Equation (2.27).

Construction of operators

For extended convolution on $\widehat{\mathbb{C}}$ to be equivariant under the action of Möbius transformations $g \in \mathrm{SL}(2, \mathbb{C})$ in the sense of Equation (3.7), $\mathfrak{T}$ and ρ must satisfy the conditions in Equation (3.8) for $M = \widehat{\mathbb{C}}$, $G = \mathrm{SL}(2, \mathbb{C})$, and $D_z^g \in \mathrm{L}$.

Claim 6 (Conformal frame fields). *If $\mathfrak{T}$ is defined as*

$$\mathfrak{T}_\psi(x) \equiv \begin{bmatrix} \left[d\,\mathrm{Log}_x\psi\big|_0\right]^{-\frac{1}{2}} & 0 \\ \frac{1}{2}\left[\nabla d\,\mathrm{Log}_x\psi\big|_0\right]\left[d\,\mathrm{Log}_x\psi\big|_0\right]^{-\frac{3}{2}} & \left[d\,\mathrm{Log}_x\psi\big|_0\right]^{\frac{1}{2}} \end{bmatrix} \tag{3.29}$$

where $d\,\mathrm{Log}_x\psi\big|_0$, $\nabla d\,\mathrm{Log}_x\psi\big|_0 \in \mathbb{C}$ are the differential and Hessian of $\mathrm{Log}_x\psi$ evaluated at the origin, then the condition for the frame operator in Equation (3.8) is satisfied.

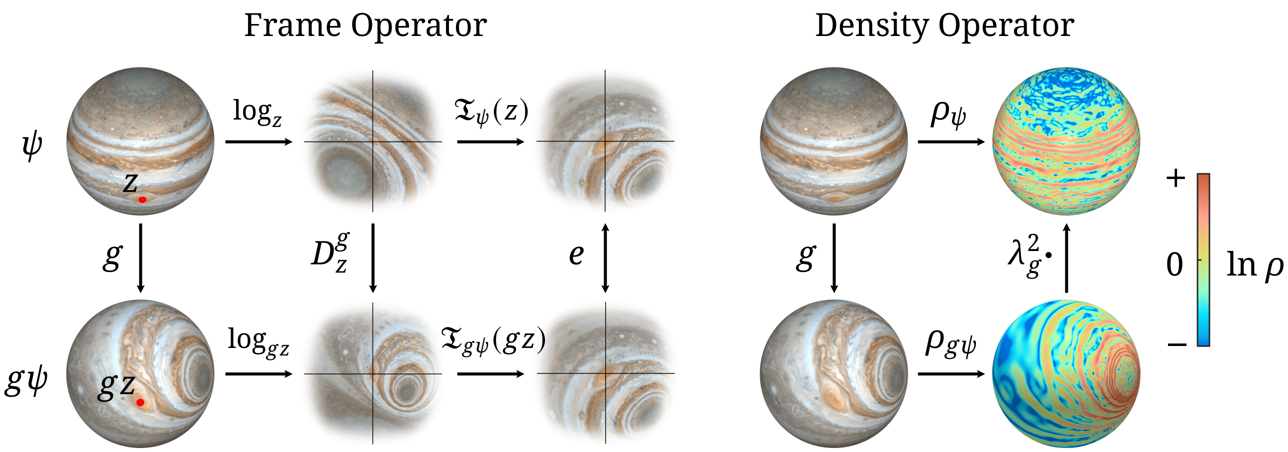


Figure 3-1. Conformally-equivariant convolutions on $\widehat{\mathbb{C}}$ require two ingredients: a *frame operator* and a *density operator*. Filters assign weights based on the relative positions of points and the frame operator $\mathfrak{T}$ corrects for the deformation of the local "tangent space" under a Möbius transformation $g \in \mathrm{SL}(2,\mathbb{C})$. Similarly, the density operator ρ adjusts for the change in the area measure used for integration, proportional to the conformal scale factor λ_g^2.

Proof. For any $\psi \in L^2(\widehat{\mathbb{C}},\mathbb{C})$ and $g \in \mathrm{SL}(2,\mathbb{C})$ it follows that for any point $x \in \widehat{\mathbb{C}}$, $D_x^g : \widehat{\mathbb{C}}_x \to \widehat{\mathbb{C}}_{gx} \in \mathrm{L}$ and we denote $D_x^g = \left[\begin{smallmatrix} a & 0 \\ n & a^{-1}\end{smallmatrix}\right]$. As in the planar case, applying the chain rule and evaluating at the origin using Equation (2.24) gives

$$d\,(\mathrm{Log}_{gx} \circ g)\,\psi\big|_0 = a^{-2}\left[d\,\mathrm{Log}_x \psi\big|_0\right] \tag{3.30}$$

and

$$\nabla d\,(\mathrm{Log}_{gx} \circ g)\,\psi\big|_0 = a^{-4}\left[\nabla d\,\mathrm{Log}_x \psi\big|_0\right] + 2na^{-3}\left[d\,\mathrm{Log}_x \psi\big|_0\right]. \tag{3.31}$$

Then, the upper diagonal element of $\mathfrak{T}_{g\psi}(gx)$ is given by

$$\begin{aligned} \left[\mathfrak{T}_{g\,\psi}(gx)\right]_{11} &\overset{(3.29)}{=} \left[d\,(\mathrm{Log}_{gx} \circ g)\,\psi\big|_0\right]^{-\frac{1}{2}} \\ &\overset{(3.30)}{=} a\left[d\,\mathrm{Log}_x \psi\big|_0\right]^{-\frac{1}{2}} \\ &\overset{(3.29)}{=} a\left[\mathfrak{T}_\psi(x)\right]_{11}. \end{aligned}$$

A similar argument shows that the lower diagonal element satisfies $\left[\mathfrak{T}_{g\,\psi}(gx)\right]_{22} =$

$a^{-1}\left[\mathfrak{T}_{\psi}(x)\right]_{22}$. For the nonzero off-diagonal element we have

$$\begin{aligned}
\left[\mathfrak{T}_{g\psi}(gx)\right]_{21} &\overset{(3.29)}{=} \frac{1}{2}\left[\nabla d\,(\mathrm{Log}_{gx}\circ g)\,\psi\big|_0\right]\left[d\,(\mathrm{Log}_{gx}\circ g)\,\psi\big|_0\right]^{-\frac{3}{2}} \\
&\overset{(3.31)}{=} n\left[d\,\mathrm{Log}_x\psi\big|_0\right]^{-\frac{1}{2}} + a^{-1}\,\frac{1}{2}\left[\nabla d\,\mathrm{Log}_x\psi\big|_0\right]\left[d\mathrm{Log}_x\psi\big|_0\right]^{-\frac{3}{2}} \\
&\overset{(3.29)}{=} n\left[\mathfrak{T}_{\psi}(x)\right]_{11} + a^{-1}\left[\mathfrak{T}_{\psi}(x)\right]_{21},
\end{aligned}$$

from which it follows that

$$D_x^g \circ \mathfrak{T}_{\psi}(x) = \mathfrak{T}_{g\psi}(gx)$$

as desired. □

Claim 7 (Conformal densities). *If ρ is defined as*

$$\rho_{\psi}(x) \equiv \left|d\,\mathrm{Log}_x\psi\big|_0\right|^2 \tag{3.32}$$

then the condition for the density operator in Equation (3.8) is satisfied.

Proof. From Equations (3.27) and (2.24), the differential of the generalized exponential $\mathrm{Exp}_x = \mathrm{Log}_x^{-1}$ at the origin is given by

$$d\,\mathrm{Exp}_x\big|_0 = \frac{|c|^2(1+|x|^2)}{c^2},$$

from which it follows that

$$\left|d\,\mathrm{Exp}_x\big|_0\right| = (1+|x|^2) \tag{3.33}$$

for any choice of $c \in \mathbb{C}$. Then, for any $g \in \mathrm{SL}(2,\mathbb{C})$ applying the chain rule to the definition of D_x^g in Equation (3.3) gives

$$\begin{aligned}
\left|d\,D_x^g\big|_0\right|^2 &\overset{(3.3)}{=} \left|d\,\mathrm{Exp}_x\big|_0\right|^2 \left|d\,g\big|_x\right|^2 \left|d\,\mathrm{Log}_{gx}\big|_{gx}\right|^2 \\
&= \left|d\,\mathrm{Exp}_x\big|_0\right|^2 \left|d\,g\big|_x\right|^2 \left|d\,\mathrm{Exp}_{gx}\big|_0\right|^{-2} \\
&\overset{(3.33)}{=} (1+|x|^2)^2\left(\frac{1}{|cx+d|^4}\right)\left(\frac{1}{(1+|gx|^2)^2}\right) \\
&= \frac{(1+|x|^2)^2}{(1+|gx|^2)^2|cx+d|^4} \\
&\overset{(2.29)}{=} \lambda_g^2(x),
\end{aligned} \tag{3.34}$$

where the second equality follows from the fact that Log_x is an isometry of $\widehat{\mathbb{C}}$ with $\mathrm{Log}_x^{-1} = \mathrm{Exp}_x$. It follows that for any $\psi \in L^2(\widehat{\mathbb{C}}, \mathbb{C})$ and $g \in \mathrm{SL}(2, \mathbb{C})$,

$$\begin{aligned}
\rho_{g\psi}(gx) &\overset{(3.32)}{=} \left| d\,(\mathrm{Log}_{gx} \circ g)\, \psi\big|_0 \right|^2 \\
&\overset{(3.3)}{=} \left| d\,(D_x^g \circ \mathrm{Log}_x)\, \psi\big|_0 \right|^2 \\
&= \left| d\,\mathrm{Log}_x \psi \circ \left[D_x^g\right]^{-1}\big|_0 \right|^2 \\
&= \left| d\, D_x^g\big|_0 \right|^{-2} \left| d\,\mathrm{Log}_x \psi\big|_0 \right|^2 \\
&\overset{(3.34)}{=} \lambda_g^{-2}(x) \left| d\,\mathrm{Log}_x \psi\big|_0 \right|^2 \\
&\overset{(3.32)}{=} \lambda_g^{-2}(x)\, \rho_\psi(x),
\end{aligned}$$

as desired. □

3.3.3 Extended convolution on $\mathbb{D}$

Last, we realize isometry-equivariant extended convolutions on the hyperbolic disk $\mathbb{D} \cong \mathrm{SU}(1,1)/\mathrm{U}(1)$. While we do not consider this formulation in the applications discussed in Part II, it is presented here for the sake of completeness.

The generalized logarithm on $\mathbb{D}$ can be defined in a similar manner as was done for $\widehat{\mathbb{C}}$ in §3.3.2. Specifically, the generalized logarithm at a point $z \in \mathbb{D}$ is the *hyperbolic* rotation $\mathrm{Log}_z \in \mathrm{SU}(1,1)$ taking z to the origin

$$\mathrm{Log}_z \equiv \frac{1}{|c|\sqrt{1 - |z|^2}} \begin{bmatrix} c & -cz \\ -\overline{c}\overline{z} & \overline{c} \end{bmatrix}, \tag{3.35}$$

where $c \in \mathbb{C}$ is arbitrary. It's easy to see that the generalized logarithm is well-defined for all $z \in \mathbb{D}$ and that it satisfies the condition in Equation (3.2):

$$\widetilde{\pi}\left(\mathrm{Log}_z^{-1}\right) \overset{(3.35)}{=} \widetilde{\pi}\left(\frac{1}{|c|\sqrt{1 - |z|^2}} \begin{bmatrix} c & cz \\ \overline{c}\overline{z} & \overline{c} \end{bmatrix} \right) \overset{(2.11)}{=} z.$$

Recalling that $\mathrm{U}(1)$ is the origin-preserving subgroup of $\mathrm{SU}(1,1)$, the frame and density operators are maps

$$\mathfrak{T} : L^2(\mathbb{D}, \mathbb{C}) \to L^2(\mathbb{D}, \mathrm{U}(1)) \qquad \text{and} \qquad \rho : L^2(\mathbb{D}, \mathbb{C}) \to L^2(\mathbb{D}, \mathbb{C}),$$

and the extended convolution of a function ψ with a filter f, both in $L^2(\mathbb{D}, \mathbb{C})$, is

$$(\psi * f)(y) \overset{(3.6)}{=} \int_{\mathbb{D}} \rho_\psi(z) \left[\mathfrak{T}_\psi(z)\, f\right] \left(\mathrm{Log}_z\, y\right) dz, \tag{3.36}$$

with dz the area measure under the hyperbolic metric as in Equation (2.33).

Construction of operators

For extended convolution to be equivariant under SU(1,1) – the group of isometries of the disk – the frame and density operators must satisfy conditions in Equation (3.8) for $M = \mathbb{D}$, $G = \mathrm{SU}(1,1)$, and $D_z^g \in \mathrm{U}(1)$. Namely, defining $\mathfrak{T}$ and ρ in a similar manner as was done for extended convolution in the plane

$$\mathfrak{T}_\psi(x) \equiv \begin{bmatrix} \left[\mathrm{sgn}\, \overline{d\,\mathrm{Log}_x\, \psi\big|_0}\right]^{\frac{1}{2}} & 0 \\ 0 & \left[\mathrm{sgn}\, \overline{d\,\mathrm{Log}_x\, \psi\big|_0}\right]^{-\frac{1}{2}} \end{bmatrix}$$

$$\rho_\psi(x) \equiv \left|d\,\mathrm{Log}_x\, \psi\big|_0\right|,$$

ensures the conditions are satisfied. The proof follows the same argument in that of Claim 5.

3.4 Convolution vs. correlation

In applications, convolutions are often replaced by *correlations*, which are perhaps more intuitive as they are, in essence, moving dot products. In the plane, the correlation of a function $\psi \in L^2(\mathbb{C}, \mathbb{C})$ with a filter $f \in L^2(\mathbb{C}, \mathbb{C})$ is expressed as the function

$$(\psi \star f)(z) = \int_{\mathbb{C}} \psi(y)\, \overline{f(y - z)}\, dy. \tag{3.37}$$

Its interpretation is straight-forward: correlation slides the filter across the plane and at each point z, the value of ψ at each neighboring point y is weighted by the (conjugated) value of the filter at the relative position of y as seen from z, which is given by $y - z$. This is exactly equivalent to evaluating the inner product of ψ with the filter f recentered at z.

Recalling that for any points $p, q \in M$, the generalized logarithm of q with respect to p gives us a description of the "position" of q as seen from p, we extend correlation to allow the filter to adaptively transform in a similar manner as was done for convolution in §3.2. Namely, given maps $\mathfrak{T}$ and ρ as in Equation (3.5), we define the *extended correlation* of a function $\psi \in L^2(M, \mathbb{C})$ with a filter $f \in L^2(M, \mathbb{C})$ to be the function in $L^2(M, \mathbb{C})$ with

$$(\psi \star f)(p) = \int_M \rho_\psi(q)\, \overline{\left[\mathfrak{T}_\psi(p)\, f\right]\left(\mathrm{Log}_p\, q\right)}\, dq. \tag{3.38}$$

Furthermore, it can be shown that if $\mathfrak{T}$ and ρ satisfy the conditions in Equation (3.8), then the extended correlation is equivariant under transformations in G in the same sense as the extended convolution with

$$g\,(\psi * f) = (g\,\psi * f).$$

So, why are we focusing on convolution? After all, the extended correlation has the same equivariance properties and is more intuitive – why bother with convolution?

To answer this question, we can look at how extended correlation and extended convolution each distribute the weights determined by the filter. Specifically, extended correlation is a *gathering* operation – at each point $p \in M$ the filter weights are distributed with respect to the single coordinate frame $\mathfrak{T}_\psi(p)$ at p. In other words, the operation is equivalent to taking the dot product of a function ψ with the filter that has *first* been transformed by $\mathfrak{T}_\psi(p)$, *then* recentered at p. In contrast, extended convolution is a *scattering* operation. Filter weights are distributed by having each neighbor q describe the position of the point p in their own coordinate frames $\mathfrak{T}_{\psi}(q)$.

In applications, the problem with extended correlation is that gathering operations with frame fields are inherently fragile. On the plane and the Riemann sphere, the frame fields – Equations (3.22) and (3.29) – are defined relative to the differential and Hessian of $\text{Log}_x\,\psi$. While this gives us a repeatable recipe for equivariant convolutions (and correlations), we can see that the frames are ill-defined whenever the differential vanishes. In a gathering operation such as extended correlation wherein the response at p is depends entirely on $\mathfrak{T}_{\psi}(p)$, the operation is ill-defined wherever the frame is ill-defined. Furthermore, the dependence of the response on a single coordinate frame makes extended correlation highly sensitive to disturbances of the frame field. If the frame at a given point is perturbed, either due to noise or other artifacts in the calculation, then the response of the extended correlation will change significantly, as the contribution of each of the neighboring values will be affected.

In contrast, scattering operations like extended convolution are naturally robust. If the frame at a given point is perturbed or ill-defined, it has no effect on the response of the convolution at the point since the assignment of filter weights depends on the frames of its neighbors. Furthermore, if the frames at several neighboring points are influenced by noise or other nuisance factors, only their contributions to the response will be affected. Then, if the majority of the neighboring frames remain relatively stable, the effect of this perturbation will be minimized and the value of the convolution will also remain stable.

Chapter 4

Extended Convolution on Surfaces

In applications, we will be interested in analyzing functions on the surfaces of 3D shapes. The challenge performing spatial aggregations on surfaces is that classical notions of convolution and correlation in Euclidean spaces cannot simply be transposed onto curved domains. Unlike images, points on a surface have no canonical orientation, without which simple operations fundamental to the spatial propagation of information, such as the alignment of filters with the local signal, cannot be computed in a repeatable manner. A key strength of extended convolution is that it entirely sidesteps this problem – no canonical frame is needed as we define our own.

In what follows we define a notion of extended convolution on arbitrary 2D surfaces (2D oriented Riemannian manifolds), which then generalize to a convolution operator on surface vector fields we call *field convolution*. While any homogeneous space can be viewed as a Riemannian manifold, an arbitrary 2D surface is not necessarily a homogeneous space. This lack of additional structure will force us to retreat from designing convolutions equivariant to arbitrary diffeomorphisms, and here we restrict our focus to isometries. However, as we will demonstrate through experiments in the latter half of this thesis, the resulting framework is highly descriptive and robust, and is well-suited to feature description and CNNs on surfaces.

4.1 Extended convolution

An arbitrary 2D surface M is not necessarily a homogeneous space, and as such, we cannot define a notion of the generalized logarithm as an element of a transformation group as discussed in §3.1.1. Given points p and q in M, we revert to expressing the "position" of q in the frame at p as the point in T_pM given by the Riemannian logarithm of q with respect to p, denoted $\log_p q$ as in Equation (2.13).

On surfaces, the frame and density operators are defined in a similar manner as on homogeneous spaces (§3.2.1). Formally, the frame operator $\mathfrak{T}$ maps a function $\psi \in L^2(M, \mathbb{C})$ to a frame field $\mathfrak{T}_\psi$, associating to each point $p \in M$ an orthonormal map $\mathfrak{T}_\psi(p)$ from the plane to the tangent space at p. That is, if $\mathcal{F}M$ is the fiber bundle with $\mathcal{F}M_p$ the group of orthonormal transformations from $\mathbb{C}$ to T_pM, then $\mathfrak{T}_\psi \in \Gamma(\mathcal{F}M)$ is a section of this bundle, and we write

$$\mathfrak{T} : L^2(M, \mathbb{C}) \to \Gamma(\mathcal{F}M) \qquad \text{and} \qquad \rho : L^2(M, \mathbb{C}) \to L^2(M, \mathbb{C}) \quad . \tag{4.1}$$

Assuming the basis $\{\mathbf{e}_1, \mathbf{e}_2\}_p$ assigned to T_pM is orthonormal for all $p \in M$, the action of the orthonormal map $\mathfrak{T}_\psi(p)$ taking $\mathbb{C}$ to T_pM is equivalent to multiplication by a unit complex number $e^{i\phi_p} \in \mathrm{U}(1)$, for some $\phi_p \in [0, 2\pi)$. In what follows we will abuse notation and consider the frame operator to be a map

$$\mathfrak{T} : L^2(M, \mathbb{C}) \to L^2(M, \mathrm{U}(1)) \tag{4.2}$$

though formally the frame field belongs to $\Gamma(\mathcal{F}M)$.

Functions ψ on a surface are convolved with planar filters $f \in L^2(\mathbb{C}, \mathbb{C})$. Since the logarithm is only well-defined locally, at each point p filters are supported on $\log_p(\mathcal{N}_p) \subset \mathbb{C}$ – the image of the geodesic ε-ball about p, $\mathcal{N}_p \subset M$, under the logarithm at p. Specifically, the extended convolution of a function $\psi \in L^2(M, \mathbb{C})$ with a filter $f \in L^2(\mathbb{C}, \mathbb{C})$ is the function in $L^2(M, \mathbb{C})$ with

$$(\psi * f)(p) = \int_{\mathcal{N}_p} \rho_\psi(q) \left[\mathfrak{T}_\psi(q)\, f\right](\log_q p)\, dq. \tag{4.3}$$

4.1.1 Isometry-equivariance

The role of the frame field is to correct for the rotation of the tangent space induced by isometries $\boldsymbol{\gamma} : M \longrightarrow M'$, between surfaces M and M'. In fact, if the frame and density operators satisfy a similar property as in Claim 1, then the extended convolution commutes with local isometries of the surface.

Claim 8 (Isometry-equivariant extended convolution on surfaces)**.** *Consider surfaces M and M' and any two points $p \in M$ and $p' \in M'$. Let $\mathcal{N}_p \subset M$ and $\mathcal{N}'_{p'} \subset M'$ be ε–balls about the points and suppose that $\mathcal{N}_p$ and $\mathcal{N}'_{p'}$ are isometric. That is, there exists a map $\boldsymbol{\gamma} : M \longrightarrow M'$ taking p to p' and satisfying $\forall q_1, q_2 \in \mathcal{N}_p$,*

$$d\left(q_0,\, q_1\right) = d\left(q'_0,\, q'_1\right), \quad q'_i = \boldsymbol{\gamma}(q_i) \in \mathcal{N}'_{p'}, \tag{4.4}$$

where $d\left(\,\cdot\,,\,\cdot\,\right)$ is the geodesic distance. Then, if for all $\psi \in L^2(M, \mathbb{C})$, $\mathfrak{T}$ and ρ satisfy

$$\left[\, d\boldsymbol{\gamma}|_p \,\right] \circ \mathfrak{T}_\psi(q) = \mathfrak{T}_{\boldsymbol{\gamma}\psi}(\,\boldsymbol{\gamma}(q)\,) \qquad \text{and} \qquad \rho_\psi(q) = \rho_{\boldsymbol{\gamma}\psi}(\,\boldsymbol{\gamma}(q)\,) \tag{4.5}$$

for all $q \in \mathcal{N}_p$, then for any filter f, the extended convolution commutes with $\boldsymbol{\gamma}$ at p such that

$$(\psi * f)(p) = (\boldsymbol{\gamma}\psi * f)(\,\boldsymbol{\gamma}(p)\,). \tag{4.6}$$

Proof. Suppose $\mathfrak{T}$ and ρ satisfy the conditions in Equation (4.5) for some locally isometric diffeomorphism $\boldsymbol{\gamma} : M \longrightarrow M'$ at $p \in M$. Recall that for $q \in \mathcal{N}_p$, the action of the differential $d\boldsymbol{\gamma}|_q : T_qM \longrightarrow T_{\boldsymbol{\gamma}(q)}M'$ can be expressed as a rotation by an angle γ_q as in Equation (2.20), so the condition for the frame operator in Equation (4.5) can be equivalently expressed as

$$e^{i\gamma_q} \circ \mathfrak{T}_\psi(q) = \mathfrak{T}_{\boldsymbol{\gamma}\psi}(\,\boldsymbol{\gamma}(q)\,). \tag{4.7}$$

It follows that for any filter f and $q \in \mathcal{N}_p$, we can relate the expression of the filter over T_qM to the expression of the filter over $T_{\boldsymbol{\gamma}(q)}M'$:

$$\begin{aligned}
\left[\mathfrak{T}_\psi(q)\, f\right](\,\log_q\, p\,) &= f\left(\left[\mathfrak{T}_\psi(q)\right]^{-1}\log_q\, p\right)\\
&\overset{(4.7)}{=} f\left(\left[\mathfrak{T}_{\boldsymbol{\gamma}\psi}(\,\boldsymbol{\gamma}(q)\,)\right]^{-1}e^{i\gamma_q}\,\log_q\, p\right)\\
&\overset{(2.21)}{=} f\left(\left[\mathfrak{T}_{\boldsymbol{\gamma}\psi}(\,\boldsymbol{\gamma}(q)\,)\right]^{-1}\,\log_{\boldsymbol{\gamma}(q)}\,\boldsymbol{\gamma}(p)\right)\\
&= \left[\mathfrak{T}_{\boldsymbol{\gamma}\psi}(\,\boldsymbol{\gamma}(q)\,)\, f\right](\,\log_{\boldsymbol{\gamma}(q)}\,\boldsymbol{\gamma}(p)\,)
\end{aligned} \tag{4.8}$$

Using this relationship we have

$$\begin{aligned}
(\psi * f)(p) &\overset{(4.3)}{=} \int_{\mathcal{N}_p} \rho_\psi(q)\left[\mathfrak{T}_\psi(q)\, f\right]\left(\log_q p\right)dq\\
&\overset{(4.8)}{=} \int_{\mathcal{N}_p} \rho_\psi(q)\left[\mathfrak{T}_{\boldsymbol{\gamma}\psi}(\,\boldsymbol{\gamma}(q)\,)\, f\right]\left(\log_{\boldsymbol{\gamma}(q)}\,\boldsymbol{\gamma}(p)\,\right)dq\\
&\overset{(4.5)}{=} \int_{\mathcal{N}_p} \rho_{\boldsymbol{\gamma}\psi}(\,\boldsymbol{\gamma}(q)\,)\left[\mathfrak{T}_{\boldsymbol{\gamma}\psi}(\,\boldsymbol{\gamma}(q)\,)\, f\right]\left(\,\log_{\boldsymbol{\gamma}(q)}\,\boldsymbol{\gamma}(p)\,\right)dq\\
&= \int_{\mathcal{N}'_{\boldsymbol{\gamma}(p)}} \rho_{\boldsymbol{\gamma}\psi}(q')\left[\mathfrak{T}_{\boldsymbol{\gamma}\psi}(q')\, f\right]\left(\log_{q'}\,\boldsymbol{\gamma}(p)\,\right)dq'\\
&\overset{(4.3)}{=} (\boldsymbol{\gamma}\psi * f)(\,\boldsymbol{\gamma}(p)\,),
\end{aligned}$$

where the second to last equality follows from the change of variables,

$$\begin{aligned}
q &\mapsto \boldsymbol{\gamma}(q) = q'\\
\mathcal{N}_p &\mapsto \boldsymbol{\gamma}\left(\mathcal{N}_p\right) = \mathcal{N}'_{\boldsymbol{\gamma}(p)},
\end{aligned} \tag{4.9}$$

with $dq = d\boldsymbol{\gamma}(q)$ since $\boldsymbol{\gamma}$ is an isometry. □

Construction of operators

Constructing operators $\mathfrak{T}$ and ρ that satisfy the conditions in Equation (4.5) is straightforward and mirrors the corresponding approach for extended convolution in the plane as described in §3.3.1. Given a function $\psi \in L^2(M, \mathbb{C})$, at any point $p \in M$ its gradient $\nabla\psi|_p$ is an element of T_pM and is pushed forwarded under a diffeomorphism $\boldsymbol{\gamma} : M \to M'$ by the differential $d\boldsymbol{\gamma}|_q : T_qM \to T_{\boldsymbol{\gamma}(q)}M'$. It follows that we can use the direction and magnitude of $\nabla\psi|_p$ to define the values of $\mathfrak{T}_\psi$ and ρ_ψ at each point $p \in M$ that satisfy the conditions in Equation (4.5). This construction is formalized in the following claim:

Claim 9 (Construction of operators on surfaces). *If $\mathfrak{T}$ and ρ are defined as*

$$\mathfrak{T}_\psi(p) \equiv \operatorname{sgn} \overline{\nabla\,\psi|_p} \qquad \textit{and} \qquad \rho_\psi(p) \equiv \left|\nabla\,\psi|_p\right|, \tag{4.10}$$

then for any function $\psi \in L^2(M,\mathbb{C})$, filter $f \in L^2(\mathbb{C},\mathbb{C})$, and diffeomorphism $\boldsymbol{\gamma} : M \to M'$, the conditions in Equation (4.5) are satisfied whenever $\boldsymbol{\gamma}$ is a local isometry.

Proof. Consider any function $\psi \in L^2(M,\mathbb{C})$, point $p \in M$, and diffeomorphism $\boldsymbol{\gamma} : M \to M'$ such that the restriction of $\boldsymbol{\gamma}$ to a local neighborhood about p is an isometry. Then, applying the chain rule to the gradient of $\boldsymbol{\gamma}\psi \in L^2(M',\mathbb{C})$ at $\boldsymbol{\gamma}(p) \in M'$ gives

$$\begin{aligned} \nabla\boldsymbol{\gamma}\psi|_{\boldsymbol{\gamma}(p)} &= \left[\, d\boldsymbol{\gamma}^{-1}|_{\boldsymbol{\gamma}(p)} \,\right] \, \nabla\psi|_p \\ &= \left[\, d\boldsymbol{\gamma}|_p \,\right]^{-1} \, \nabla\psi|_p \\ &\overset{(2.20)}{=} e^{-i\gamma_p} \, \nabla\psi|_p \,. \end{aligned} \tag{4.11}$$

It follows that

$$\begin{aligned} \mathfrak{T}_{\boldsymbol{\gamma}\psi}(\,\boldsymbol{\gamma}(p)\,) &\overset{(4.10)}{=} \operatorname{sgn} \overline{\nabla\boldsymbol{\gamma}\psi|_{\boldsymbol{\gamma}(p)}} \\ &\overset{(4.11)}{=} e^{i\gamma_p} \circ \operatorname{sgn} \overline{\nabla\psi|_p} \\ &\overset{(4.10)}{=} e^{i\gamma_p} \circ \mathfrak{T}_\psi(p) \\ &\overset{(2.20)}{=} \left[\, d\boldsymbol{\gamma}|_p \,\right] \circ \mathfrak{T}_\psi(p), \end{aligned}$$

and

$$\begin{aligned} \rho_{\boldsymbol{\gamma}\psi}(\,\boldsymbol{\gamma}(p)\,) &\overset{(4.10)}{=} \left|\nabla\boldsymbol{\gamma}\psi|_{\boldsymbol{\gamma}(p)}\right| \\ &\overset{(2.20)}{=} \left|e^{-i\gamma_p} \, \nabla\psi|_p\right| \\ &= \left|\nabla\psi|_p\right| \\ &\overset{(4.10)}{=} \rho_\psi(p), \end{aligned}$$

as desired. □

4.1.2 Optimal filters

Given an arbitrary surface M and function $\psi \in L^2(M, \mathbb{R})$, we can describe the region about a point $p \in M$ by computing the filter maximizing the response of the extended convolution in exactly the same manner as described in §3.2.5 for extended convolution on homogeneous spaces. That is, given a function $\psi \in L^2(M, \mathbb{C})$ and fixing the $\mathfrak{T}$ and ρ, at any keypoint $p \in M$, we can define a map from the space of filters to the space of extended convolution responses p.

$$\begin{aligned} \mathcal{E}_\psi^p : L^2(\mathbb{C}, \mathbb{R}) &\to \mathbb{R} \\ f &\overset{(4.3)}{\mapsto} (\psi * f)(p) \end{aligned} \tag{4.12}$$

Applying the Riesz Representation Theorem, it follows that there exist a filter f_p^ψ such that the map in Equation (4.12) can be written in terms of the inner product on $L^2(\mathbb{C}, \mathbb{R})$ with

$$\mathcal{E}_\psi^p(f) \equiv \langle f, f_\psi^p \rangle, \tag{4.13}$$

for all $f \in L^2(\mathbb{C}, \mathbb{R})$. It is clear that up to scale, f_ψ^p maximizes the response the extended convolution p and can be evaluated by computing the extended convolution of ψ with a delta function such that

$$f_\psi^p(x) \overset{(4.3)}{=} \int_{\mathcal{N}_p} \rho_\psi(q) \left[\mathfrak{T}_\psi(q)\, \delta_x\right] \left(\log_q\, p\right) dq. \tag{4.14}$$

It follows directly from Claim 8 that if $\mathfrak{T}$ and ρ satisfy the conditions in Equation (4.5), then for any diffeomorphism $\boldsymbol{\gamma} : M \to M'$ between surfaces M and M',

$$f_\psi^p = f_{\boldsymbol{\gamma}\psi}^{\boldsymbol{\gamma}(p)}$$

whenever $\boldsymbol{\gamma}$ is locally isometric at p.

4.2 Field convolution

Up to this point, we have defined extended convolution as an operator on scalar functions. Here, we generalize extended convolution by combining it with parallel transport, resulting in a convolutional operator on surface vector fields we call field convolution.

Denoting $\Gamma(TM)$ as the space of vector fields on M, the frame and density operators take a vector field $X \in \Gamma(TM)$ and return a frame field taking values in $\mathrm{U}(1)$ and *vector-valued* "density" field,

$$\begin{array}{c} \mathfrak{T} : \Gamma(TM) \to L^2(M, \mathrm{U}(1)) \\ X \mapsto \mathfrak{T}_X \end{array} \qquad \text{and} \qquad \begin{array}{c} \rho : \Gamma(TM) \to \Gamma(TM) \\ X \mapsto \rho_X \end{array} . \tag{4.15}$$

Vector fields $X \in \Gamma(TM)$ are convolved with planar filters $f \in L^2(\mathbb{C}, \mathbb{C})$ supported on the the image of the geodesic ε-ball about each point $p \in M$ – $\mathcal{N}_p$ – under the logarithm map. Formally, the *field convolution* of a vector field X with a filter f is the vector field in T_pM with

$$(X * f)(p) = \int_{\mathcal{N}_p} \mathcal{P}_{p \leftarrow q}(\, \rho_X(q)\,) \left[\mathfrak{T}_X(q)\, f\right](\log_q\, p)\; dq, \tag{4.16}$$

where $\mathcal{P}_{p \leftarrow q} : T_qM \to T_pM$ is the transport operator as defined in Equation (2.14).

For specific choices of $\mathfrak{T}$ and ρ, field convolution commutes with isometries in a manner analogous to extended convolution. In fact, the choices of operators are somewhat canonical – specifically, we define $\mathfrak{T}$ and ρ such that

$$\mathfrak{T}_X(q) \equiv \operatorname{sgn} \overline{X(q)} \qquad \text{and} \qquad \rho_X(q) \equiv X(q). \tag{4.17}$$

Given points $p, q \in M$, we can express the evaluation of a vector field $X \in \Gamma(TM)$ at p as

$$X(p) \equiv \varrho_p\, e^{i\phi_p} \in T_pM, \tag{4.18}$$

and following §2.2.1, the logarithm and transport operators can be written in polar form as

$$\log_p q \overset{(2.13)}{\equiv} r_{pq}\, e^{i\theta_{pq}} \qquad \text{and} \qquad \mathcal{P}_{p\leftarrow q}(\mathbf{v}) \overset{(2.14)}{\equiv} e^{i\varphi_{pq}}\, \mathbf{v},$$

for all $\mathbf{v} \in T_pM$. Then, for $\mathfrak{T}$ and ρ defined as in Equation (4.17), the field convolution and of a vector field X with a filter f can be expressed concretely as

$$(X * f)(p) = \int_{\mathcal{N}_p} \varrho_q\, e^{i(\phi_p + \varphi_{pq})}\, f\left(r_{qp}\, e^{i\left(\theta_{qp} - \phi_q\right)}\right)\, dq. \tag{4.19}$$

Claim 10 (Commutativity of field convolutions)**.** *Consider surfaces M and M' and any two points $p \in M$ and $p' \in M'$. Let $\mathcal{N}_p \subset M$ and $\mathcal{N}'_{p'} \subset M'$ be ε–balls about the points and suppose that there exists a diffeomorphism $\boldsymbol{\gamma} : M \longrightarrow M'$ taking p to p' with $\boldsymbol{\gamma}(\mathcal{N}_p) = \mathcal{N}'_{p'}$, such that its restriction $\boldsymbol{\gamma} : \mathcal{N}_p \longrightarrow \mathcal{N}'_{p'}$ is an isometry. Then, for any vector field $X \in \Gamma(TM)$ and filter $f \in L^2(\mathbb{C}, \mathbb{C})$, field convolution commutes with $\boldsymbol{\gamma}$ at p such that*

$$\left[\, d\boldsymbol{\gamma}|_p \,\right] \left[(X * f)(p)\right] = \left(\left[\, d\boldsymbol{\gamma}|_p \,\right] X * f\right)(\, \boldsymbol{\gamma}(p)\,). \tag{4.20}$$

Proof. Consider any vector field $X \in \Gamma(TM)$, point $p \in M$, and diffeomorphism $\boldsymbol{\gamma} : M \longrightarrow M'$ such that the restriction of $\boldsymbol{\gamma}$ to a local neighborhood $\mathcal{N}_p$ about p is an isometry. Let $X' \in \Gamma(TM')$ be the push-forward of X under $d\,\boldsymbol{\gamma}$ such that

$$X'(\boldsymbol{\gamma}(p)) = \left[\, d\boldsymbol{\gamma}|_p \,\right] X(p) = \varrho'_{\boldsymbol{\gamma}(p)}\, e^{i\phi'_{\boldsymbol{\gamma}(p)}}.$$

For any $q \in \mathcal{N}_p$, denoting $\log_q p = r_{qp}\, e^{i\theta_{qp}}$, and the angle resulting from the parallel transport along the shortest geodesic from q to p as φ_{pq}, it follows from Equations (2.20-2.21) that

$$\begin{aligned} {\varrho'}_{\boldsymbol{\gamma}(q)} &= \varrho_q \quad \text{and} \quad & {\phi'}_{\boldsymbol{\gamma}(q)} &= \phi_q + \gamma_q, \\ r_{\boldsymbol{\gamma}(q)\boldsymbol{\gamma}(p)} &= r_{qp} \quad \text{and} \quad & \theta_{\boldsymbol{\gamma}(q)\boldsymbol{\gamma}(p)} &= \theta_{qp} + \gamma_q, \end{aligned}$$

$$\varphi_{\boldsymbol{\gamma}(p)\boldsymbol{\gamma}(q)} = \varphi_{pq} + \gamma_p - \gamma_q$$

where γ_q is the angle of rotation corresponding to the action of the differential $d\boldsymbol{\gamma}|_q$, taking vectors in T_qM to $T_{\boldsymbol{\gamma}(q)}M'$. (Recall that as $\boldsymbol{\gamma}$ is an isometry, the action of $[\ d\boldsymbol{\gamma}|_q]$ on T_qM is equivalent to a rotation $e^{i\gamma_q} \in \mathrm{U}(1)$ when expressed in the orthonormal basis for T_qM). Then, we can relate the transport of $X(q)$ from T_qM to T_pM to that of $X'(\boldsymbol{\gamma}(q))$ from $T_{\boldsymbol{\gamma}(q)}M'$ to $T_{\boldsymbol{\gamma}(p)}M'$,

$$\varrho'_{\boldsymbol{\gamma}(q)}\, e^{i\left(\phi'_{\boldsymbol{\gamma}(q)}+\varphi_{\boldsymbol{\gamma}(p)\boldsymbol{\gamma}(q)}\right)} = \varrho_q\, e^{i\left(\phi_q+\varphi_{pq}+\gamma_p\right)}, \tag{4.21}$$

in addition to the filter argument in the expression of field convolution in Equations (4.19),

$$r_{\boldsymbol{\gamma}(q)\boldsymbol{\gamma}(p)}\, e^{i\left(\theta_{\boldsymbol{\gamma}(q)\boldsymbol{\gamma}(p)}-\phi'_{\boldsymbol{\gamma}(q)}\right)} = r_{qp}\, e^{i\left(\theta_{qp}-\phi_q\right)}. \tag{4.22}$$

It follows that

$$\begin{aligned}
\left[\ d\boldsymbol{\gamma}|_p\ \right]\left[(X * f)(p)\right] &\overset{(4.19)}{=} \int_{\mathcal{N}_p} \varrho_q\, e^{i\left(\phi_p+\varphi_{pq}+\gamma_p\right)}\, f\left(r_{qp}\, e^{i\left(\theta_{qp}-\phi_q\right)}\right)\ dq \\
&\overset{(4.21)}{=} \int_{\mathcal{N}_p} \varrho'_{\boldsymbol{\gamma}(q)}\, e^{i\left(\phi'_{\boldsymbol{\gamma}(q)}+\varphi_{\boldsymbol{\gamma}(q)\boldsymbol{\gamma}(p)}\right)}\, f\left(r_{qp}\, e^{i\left(\theta_{qp}-\phi_q\right)}\right)\ dq \\
&\overset{(4.22)}{=} \int_{\mathcal{N}_p} \varrho'_{\boldsymbol{\gamma}(q)}\, e^{i\left(\phi'_{\boldsymbol{\gamma}(q)}+\varphi_{\boldsymbol{\gamma}(q)\boldsymbol{\gamma}(p)}\right)}\, f\left(r_{\boldsymbol{\gamma}(q)\boldsymbol{\gamma}(p)}\, e^{i\left(\theta_{\boldsymbol{\gamma}(q)\boldsymbol{\gamma}(p)}-\phi'_{\boldsymbol{\gamma}(q)}\right)}\right)\ dq \\
&\overset{(4.9)}{=} \int_{\mathcal{N}'_{\boldsymbol{\gamma}(p)}} \varrho'_{q'}\, e^{i\left(\phi'_{q'}+\varphi_{q'\boldsymbol{\gamma}(p)}\right)}\, f\left(r_{q'\boldsymbol{\gamma}(p)}\, e^{i\left(\theta_{q'\boldsymbol{\gamma}(p)}-\phi'_{q'}\right)}\right)\ dq' \\
&\overset{(4.19)}{=} \left(\left[\ d\boldsymbol{\gamma}|_p\ \right]X * f\right)(\ \boldsymbol{\gamma}(p)\ ),
\end{aligned}$$

where the second to last equality follows from the same change of variables as in Equation (4.9). □

Part II

Applications

Chapter 5

ECHO Descriptors

5.1 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 neighborhoods 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 [TSDS10a, 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 neighborhood about a keypoint. Typically, the construction of these descriptors consists of first defining a rotationally equivariant frame at the keypoint and then describing the neighboring 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].

Here we present a novel, highly descriptive and robust framework for rotation- and isometry-invariant image and surface feature descriptors based on the optimal filters maximizing the response of the extended convolution at a given point (§3.2.5 and §4.1.2). Intuitively, 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. Viewing the optimal filters – Equations (3.18) and (4.14) – as recipes for constructing a characterization of the local region in a histogram by accumulating the density values into the bins indexed by the position of the keypoint in the neighbors frame, we call the resulting representations *Extended Convolution Histogram of Orientation* (**ECHO**) descriptors.

5.2 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 neighborhood is then encoded relative to the frame corresponding to the assigned scale and orientation to achieve invariance.

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. Later descriptors built on top of the SIFT framework, such as GLOH [MS05], sought to increase descriptiveness at the cost of computational complexity with a specific focus on both improving repeatability and robustness in the assignment of orientations.

Other methods inspired by SIFT, including SURF [BTVG06, BETVG08] and DAISY [TLF09], produce descriptors of comparable quality while reducing computational cost; particular care is taken to devise strategies that minimize the complexity of the orientation assignment process without making large sacrifices in robustness. SURF itself has become one of the most popular and distinguished descriptors, due in part to its effectiveness in real-time applications. More recent successful 2D descriptors, BRISK [LCS11], KAZE [AS11, ABD12] and ORB [RRKB11], also use local frames to achieve either full or partial similarity 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 [TSDS10a] RoPS [GSB+13], and USC [TSDS10b] 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 centered 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, SSS19b, CPK19]. However, most of these methods learn local descriptors from existing handcrafted techniques [KZK17, DBI18], rather than input data [SSS19b]. 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].

5.3 Method overview

Given a point of interest and surrounding neighborhood, our proposed descriptor corresponds to the filter that maximizes the response of the extended convolution at the keypoint. On images, rotation-invariant descriptors are computed with respect to the formulation of extended convolution in the plane (§3.21); On surfaces, we compute isometry-invariant descriptors using the definition of extended convolution making use the Riemannian logarithm (§4.1). In both cases, the optimal filters belong to $L^2(\mathbb{C}, \mathbb{R})$ and are rasterized in two-dimensional arrays. Using the frame and density operators as defined in Equations (3.22) and (3.24) (resp. Equation (4.10)), the construction is straightforward: Given an image (resp. an intrinsic signal on the surface), we use the gradients to attach a weighted frame to every point in the neighborhood. Then, each point casts a vote into the bin centered at the position of the keypoint as seen from the frame assigned to the point, weighted by the gradient magnitude.

5.4 ECHO image descriptors

Recall from §3.3.1 that given a function $\psi \in L^2(\mathbb{C}, \mathbb{R})$, the SE(2)-invariant optimal filter maximizing response of the planar extended convolution at a point $y \in \mathbb{C}$,

$$\begin{gathered} f_\psi^y(x) \overset{(3.26)}{=} \int_{\mathcal{N}_y} \rho_\psi(z) \left[\mathfrak{T}_\psi(z)\, \delta_x\right](y - z)\, dz, \\ \text{with} \\ \mathfrak{T}_\psi(x) \overset{(3.25)}{\equiv} \begin{bmatrix} \operatorname{sgn} \overline{\nabla \psi|_p} & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{and} \qquad \rho_\psi(x) \overset{(3.25)}{\equiv} \left| \nabla \psi|_p \right|, \end{gathered} \tag{5.1}$$

can be viewed as a characterization of ψ on the ε-disk $\mathcal{N}_y$ about y. Since translation-invariance is relatively trivial, we will refer to the optimal filters as being rotation-invariant, even though they are technically invariant under rotations *and* translations.

In what follows it will be helpful to define a coordinate function

$$\begin{aligned} C_{\mathfrak{T}_\psi}^y : \mathcal{N}_y &\to \mathbb{C} \\ z &\mapsto \left[\mathfrak{T}_\psi(z)\right]^{-1}(y - z) \end{aligned} \tag{5.2}$$

taking neighboring points to the position of the keypoint as seen in their own frames, allowing the optimal filter to be re-expressed as

$$f_\psi^p(x) = \int_{\mathcal{N}_y} \rho_\psi(z)\, \delta_x\left(C_{\mathfrak{T}_\psi}^y(z) \right) dz. \tag{5.3}$$

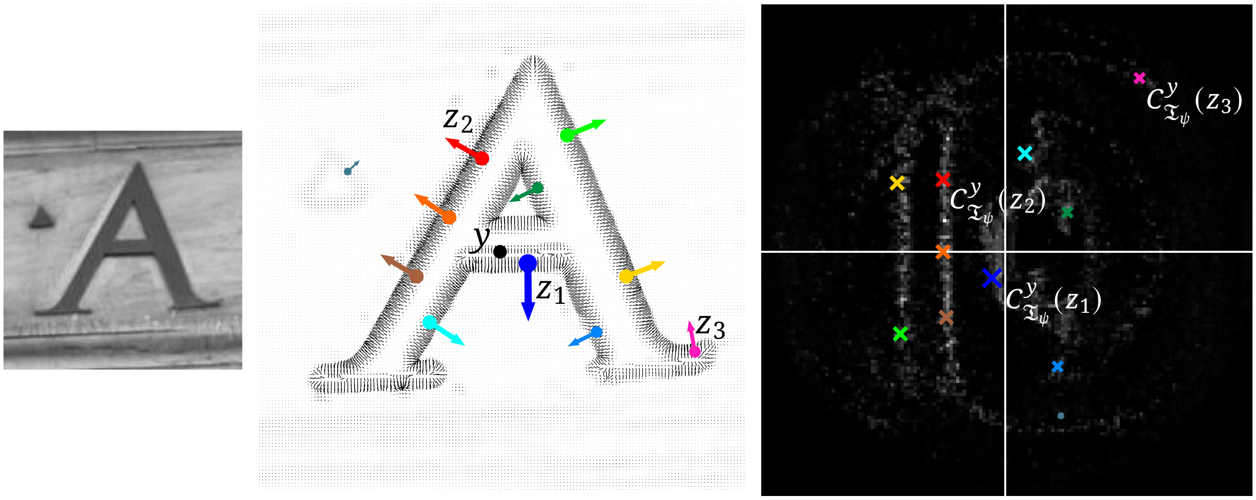

z_2
y
z_1
z_3
$C^y_{\mathfrak{T}_\psi}(z_3)$
$C^y_{\mathfrak{T}_\psi}(z_2)$
$C^y_{\mathfrak{T}_\psi}(z_1)$

Figure 5-1. Visualization of the construction of the optimal filter defined in Equation (5.1): A crop from an input image – taken to be ψ – is shown on the left. The gradients – whose direction and magnitude determine the values of $\mathfrak{T}_\psi$ and ρ_ψ at each point, the keypoint y, and neighboring points z_i are shown in the middle. The derived filter is shown on the right.

5.4.1 Construction

The construction of the ECHO image descriptor follows from the discretization of the optimal filter in Equation (5.1). Figure 5-1 shows an example of constructing the optimal filter (right) for an 'A' pattern (left) constructed with respect to the frame field determined by the gradients (middle). For each point z in the vicinity of the center point y, the gradient determines both the position of the bin and the weight of the vote with which z contributes to the filter. For example, since the gradient at z_1 is interpreted as the x-axis of a frame centered at z_1, the position of y relative to this frame will have negative coordinates. The gradient at z_1 has a large magnitude, so the point z_1 contributes a large vote to bin $C^y_{\mathfrak{T}_\psi}(z_1)$. The keypoint y has positive coordinates relative to the frame at z_3 but since the gradient is small, it contributes a lesser vote to bin $C^y_{\mathfrak{T}_\psi}(z_3)$.

Iterating over all points in the neighborhood of the pattern's center, we obtain the filter shown on the right. While the filter does not visually resemble the initial pattern, several properties of the pattern can be identified. For example, since the gradients along the outer left and right sides of the ‘A’ tend to be outward facing, points on these edges cast votes into bins with negative x-coordinates, corresponding to the two vertical seams on the left side of the filter. Similarly, the gradients on the inner edges point inwards, producing the small wing-like structures on the right side of the filter.

Algorithm 1: ECHO Image Descriptor

```
Input:
    keypoint $y \in \mathbb{C}$,
    frame field $\mathfrak{T}_\psi : \mathbb{C} \longrightarrow \mathrm{U}(1)$,
    density $\rho_\psi : \mathbb{C} \longrightarrow \mathbb{R}$,
    support radius $\varepsilon \in \mathbb{R}_{>0}$,
    descriptor radius $n \in \mathbb{Z}_{>0}$,
    Gaussian deviation $\sigma \in \mathbb{R}_{>0}$
Output:
    ECHO descriptor $\mathbf{f}^y_\psi \in \mathbb{R}^{(2n+1)\times(2n+1)}$

$\alpha = \varepsilon \,/\, n$                                        ▹ Image to descriptor scale
$\mathbf{f}^y_\psi = \mathbf{0}^{(2n+1)\times(2n+1)}$                            ▹ Initialize descriptor
$\forall\, z \in \mathbb{C}$   such that   $|y - z| \leq \varepsilon$
    $y_z = \alpha \cdot C^y_{\mathfrak{T}_\psi}(z)$                         ▹ Position of $y$ in the frame of $z$
    $\mathcal{B}_z = \{x \in \mathbb{C} \mid |x| \leq n \text{ and } |x - y_z| \leq 2\sigma\}$
    $\forall\, x = x_1 + i x_2 \in \mathcal{B}_z$                        ▹ Splat density into vicinity of $y_z$
        $\mathbf{f}^y_\psi(x_1, x_2) \mathrel{+}= \rho_\psi(x) \cdot k_{x,\sigma}(y_z)$
```

Taking ψ to be an image and given a keypoint $y \in \mathbb{C}$, our goal is to compute a rasterization of the of the optimal filter in Equation (5.1) on a $(2n+1) \times (2n+1)$ grid, denoted as $\mathbf{f}^y_\psi \in \mathbb{R}^{(2n+1)\times(2n+1)}$. Pseudocode for the computation of the discrete descriptor is shown in **Algorithm 1**: The image gradients are used to define the frame field $\mathfrak{T}_\psi$ and density ρ_ψ as in Equation (5.1). The *support radius*, ε, determines the size of the descriptor in pixel coordinates. That is, the descriptor encodes the circular region of radius ε centered at y in the image. Similar to SIFT, the ECHO descriptor is computed by binning votes on a grid. The *descriptor radius* is given by the input parameter n and the grid resolution is $2n+1$. The lattice is centered at the origin and corresponds to a histogram consisting of unit-width bins whose centers are lattice coordinates. The total number of bins in the histogram is $(2n+1)^2$.

For each neighboring point z, we compute y_z – the position of y in the frame of z, scaled to the descriptor resolution. To discretize the integral, we replace the delta function with a compactly supported kernel that approximates a Gaussian with deviation σ:

$$k_{a,\sigma}(b) = \begin{cases} \exp\left(-\frac{|a-b|^2}{\sigma^2}\right) & |a-b| \leq 2\sigma \\ 0 & \text{otherwise} \end{cases} . \tag{5.4}$$

and compute $\mathcal{B}_z$ – the set of grid points that fall within a disk of radius n of the origin and whose kernel function supports y_z; for each grid point $x \in \mathcal{B}_z$, we increment the value of the descriptor at x by $\rho_\psi(z)$, weighted by the value of the kernel function centered at x, evaluated at the position of y in the coordinate frame of the neighbor, $k_{x,\sigma}(y_z)$.

5.4.2 Evaluation

We compare the ECHO image descriptor against SIFT in the context of feature matching on a challenging, large-scale dataset. We choose to compare against SIFT for several reasons. Foremost, SIFT has stood the test of time. Despite its introduction over two decades ago, SIFT is arguably the premier detection and description pipeline and remains widely used across a number of fields, including robotics, vision, and medical imaging. Competing pipelines have generally emphasized computational efficiency and have yet to definitively outperform SIFT in terms of discriminative power and robustness [KPS17, TS18].

The advent of deep learning in imaging and vision has coincided with the introduction of a number of contemporaneous learned descriptors which have been shown to significantly outperform SIFT and other traditional methods in certain applications [MMRM17, HLS18, LSZ+18, ZR19]. However, the performance of learned descriptors is often domain-dependent and “deterministic" descriptors such as SIFT can provide either comparable or superior performance in specialized domains that learned descriptors are not specifically designed to handle [ZFR19]. More generally, “classical" methods for image alignment and 3D reconstruction, *e.g.* SIFT + RANSAC, may still outperform state-of-the-art learned approaches with the proper settings [SHSP17, JMM+20].

The scope of our contribution is limited to local image descriptors – we do not consider the related problem of feature detection. The SIFT pipeline integrates both feature detection and description in the sense that keypoints are chosen based on the distinctive potential of the surrounding area. As we seek to compare against the SIFT descriptor directly, we perform two sets of experiments. In the first, we replace the SIFT descriptor with ECHO within the SIFT pipeline to compare practical effectiveness. The goal of the second experiment is to more directly evaluate our contribution with respect to the design of rotationally invariant descriptors. Specifically, we seek an answer to the following question: By having all points in the local region encode the keypoint relative to their own frames, do we produce a more robust and discriminating descriptor than one constructed relative to the keypoint’s frame?

Comparison regime

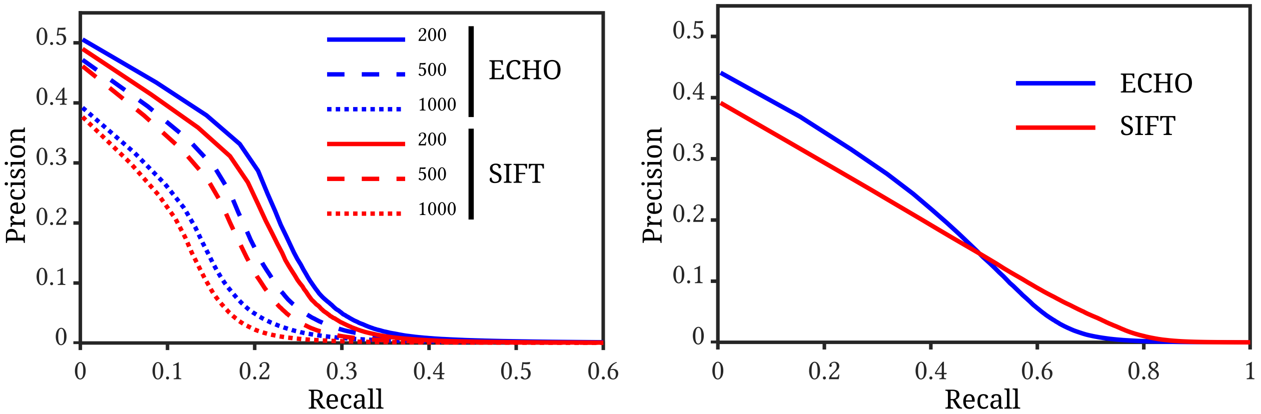
In both sets of experiments, we evaluate ECHO and SIFT in the context of descriptor matching using the publicly available photo-tourism dataset associated with the 2020 CVPR Image Matching Workshop [JMM+20]. The dataset consists of collections of images of international landmarks captured in a wide range of conditions using different devices. As such, we use the dataset to simultaneously evaluate descriptiveness and robustness. The dataset also includes 3D ground-truth information in the form of the camera poses and depth maps corresponding to each image. In all of our experiments, we use the implementation of SIFT in the OpenCV library [Bra00] with the default parameters.

Due to the large size of the dataset, we restrict our evaluations to the image pools corresponding to six landmarks: `reichstag`, `pantheon_exterior`, `sacre_coeur`, `taj_mahal`, `temple_nara_japan`, and `westminster_abbey`, which we believe reflect the diversity of the dataset as a whole. Experiments are performed by evaluating the performance of the descriptors in matching a set of *scene* images to a smaller set of *models*.

For each landmark, five *model* images are chosen and removed from the pool. These images are picked such that their subjects overlap but differ significantly in terms of viewpoint and image quality. The *scenes* are those images in the remainder of the pool that best match the models.

Specifically, SIFT keypoints are computed for all model in each pool. Keypoints without a valid depth measure are discarded. For each landmark, images in the pool are assigned a score based on the number of keypoints that are determined to correspond to at least one keypoint from the five models originally drawn from the pool.

Keypoints are considered to be in correspondence if the distance between their associated 3D points is less than a threshold value τ. For each of the five models, all pixels with valid depth are projected into 3D using the ground-truth depth maps and camera poses. These points are used to compute a rough triangulation corresponding to the surface of the landmark. As in [GBS+16], we define the threshold



(a) Keypoints and scale determined by the SIFT feature detector

(b) Randomly selected keypoints and scale estimated from ground truth

Figure 5-2. The mean precision-recall curves for the ECHO and SIFT descriptors. On the left, the keypoints and corresponding scales are computed in the SIFT pipeline. The three different curves correspond to the average over all scenes using the first 200, 500, and 1000 keypoints in each. On the right, keypoints are selected at random and scale is estimated from the ground truth. The curves are averaged over all scenes using 1000 keypoints in each.

value relative to the area of the image, A,

$$\tau = 0.005 \cdot \sqrt{A \,/\, \pi}\,. \tag{5.5}$$

The top 15 images with the highest score from each pool are chosen as the *scenes*. The scaling factor in the value of τ was determined empirically; it provides a good balance between keypoint distinctiveness and ensuring each scene contributes approximately 1000 keypoints to the total.

Comparisons within the SIFT pipeline

In our first experiment, we perform comparisons with keypoints selected using the SIFT keypoint detector to gauge ECHO's practical effectiveness. For each model, we compute SIFT keypoints and sort them in descending order by "contrast" [Low04]. Of these, we retain the first 1000 distinct keypoints having a valid depth measure, preventing models with relatively large numbers of SIFT keypoints from having an outsize influence in our comparisons.

For each scene, we compute SIFT keypoints and discard those without valid depth. Those that remain are sorted by contrast and only the first 1000 distinct keypoints that match at least one keypoint from the five corresponding models are retained.

Next, ECHO and SIFT descriptors are computed at each keypoint for both the models and the scenes. Both descriptors are computed at the location in the Gaussian pyramid assigned to the keypoint in the SIFT pipeline. The support radius, ϵ, of the SIFT descriptor is determined by the scale associated with the keypoint in addition to the number of bins used in the histogram. The ECHO descriptor uses more bins and we find that it generally exhibits better performance using a support radius 2.5 times larger that of the corresponding SIFT descriptor.

Comparisons using randomized keypoints

Our second set of experiments are performed in the same manner using the same collection of models and scenes. The only difference is that the keypoints are selected at random so as to avoid the influence of the SIFT feature detection algorithm on the results. Specifically, for each model, 1000 keypoints are randomly chosen out of the collection of points that have a valid depth measure. Then, for each scene, we randomly select keypoints with valid depth and keep only those that correspond to at least one keypoint from the five associated images in the models. This process is iterated until 1000 such points are obtained.

We use the ground-truth 3D information to provide an idealized estimation of the scale. That is, for a keypoint, the associated 3D point is first translated by 2τ in a direction perpendicular to the camera's view direction and then projected into the image plane. For both descriptors, the distance between the 2D keypoint and the projected offset defines the support radius.

Matching performance

In both sets of experiments, we evaluate the matching performance of the SIFT and ECHO descriptors by computing precision-recall curves for all keypoints in the scenes, an approach that has been demonstrated to be well-suited to this task [KS04, MS05]. Given a scene keypoint, y and corresponding descriptor $\mathbf{f}^y$, all keypoints from across all models are sorted based on the descriptor distance, giving $\{z_1, \ldots, z_N\}$ with

$$\|\mathbf{f}^y - \mathbf{f}^{z_i}\| \leq \|\mathbf{f}^y - \mathbf{f}^{z_{i+1}}\|.$$

Some keypoints may be assigned multiple descriptors in the SIFT pipeline depending on the number of peaks in the local orientation histogram. In such cases we use the minimal distance over all of the keypoint's descriptors.

Scene and model keypoints are considered to match if they correspond to the same landmark and the distance between their 3D positions is less than the threshold τ defined in Equation (5.5). We define $\mathcal{M}_p$ to be the set of all model keypoints that are valid matches with p. Following [SMKF04], the the precision $\mathcal{P}_p$ and recall $\mathcal{R}_p$ assigned to p are defined as functions of the top r model keypoints:

$$\mathcal{P}_p(r) = \frac{|\mathcal{M}_p \cap \{z_i\}_{i \leq r}|}{r} \quad \text{and} \quad \mathcal{R}_p(r) = \frac{|\mathcal{M}_p \cap \{z_i\}_{i \leq r}|}{|\mathcal{M}_p|}. \tag{5.6}$$

5.4.3 Results and discussion

We aggregate the results by computing the mean precision and recall across all keypoints in the scenes. For the first set of experiments, we compute three curves for each descriptor corresponding to the top 200, 500, and 1000 keypoints in each scene as ranked by contrast. The resulting precision-recall curves are shown in Figure 5-2a. For the second set, we compute a single mean curve for each descriptor using all 1000 keypoints in each scene; these are shown in Figure 5-2b.

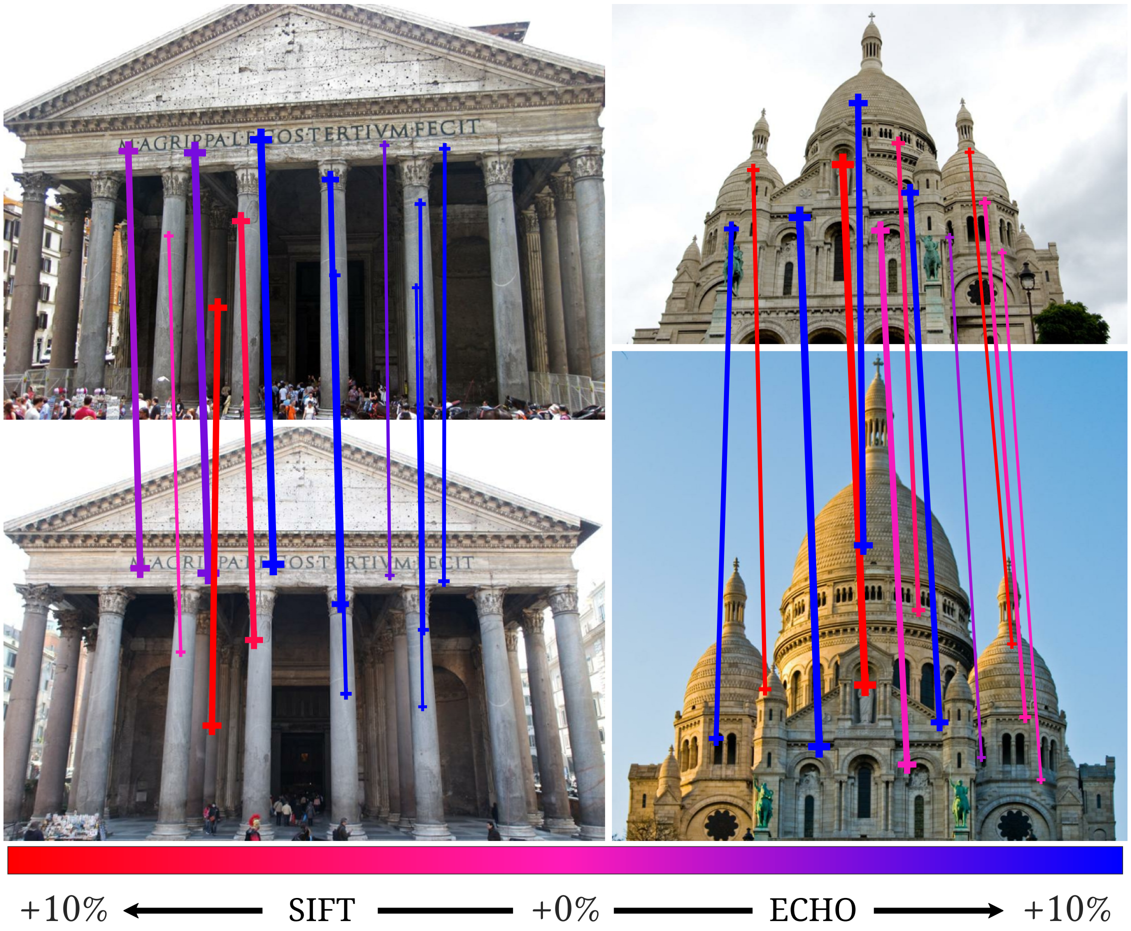


Figure 5-3. Relative performance of SIFT and ECHO in matching randomly selected keypoints in two pairs of scene (top) and model (bottom) images: Pairs of corresponding scene and model keypoints are grouped together and are visualized as vertical lines between the two images. Lines are colored to show the difference in the percentage of valid matches found by each descriptor and the thickness gives the number of corresponding pairs in each group.

Overall we see that ECHO performs better than SIFT in our evaluations, though the difference is more pronounced when keypoint detection and scale estimation are decoupled from the SIFT pipeline as in our second set of experiments. In the former case, the precision of each descriptor decreases as the number of scene keypoints increases. This is not surprising as each successive keypoint added is of lower quality in terms of potential distinctiveness. Figure 5-3 shows a comparison of the valid matches found using the SIFT and ECHO descriptors between two pairs of scene (top) and model (bottom) images in the randomized keypoint paradigm. We find that ECHO tends to find slightly more valid matches than SIFT in less challenging scenarios, as in the case on the left where the scene and model image differ mainly in terms of a small change in the 3D position of the cameras. However, both descriptors perform similarly in more challenging scenarios as shown on the right.

We do not argue that the results presented here show that the ECHO descriptor is superior. Rather, they demonstrate that the ECHO descriptor is distinctive, repeatable, and robust in its own right and has the potential to be an effective tool in challenging image matching scenarios. However, it is important to note that effective implementations of the ECHO descriptor may come at an increased cost. In our experiments, we find that ECHO performs best with a descriptor radius of 7, which translates to a descriptor size of 225 elements, roughly twice the number of elements in the standard implementation of SIFT.

The run-time of our proof-of-concept implementation of ECHO does not compare favorably to the highly optimized implementation of SIFT in OpenCV. (SIFT runs up to a factor of ten times faster.) However, both approaches have the same complexity, requiring similar local voting operations to compute the descriptor, and we believe that ECHO can be optimized in the future to be more competitive.

5.5 ECHO surface descriptors

Given a function ψ on a surface M, recall from §4.1 that the isometry-invariant optimal filter maximizing the response of the extended convolution on M at a point $p \in M$ is given by

$$f_{\psi}^{p}(x) \overset{(4.14)}{=} \int_{\mathcal{N}_p} \rho_{\psi}(q) \left[\mathfrak{T}_{\psi}(q)\, \delta_x\right]\left(\log_q p\right) dq,$$

$$\text{with} \tag{5.7}$$

$$\mathfrak{T}_{\psi}(p) \overset{(4.10)}{\equiv} \operatorname{sgn} \overline{\nabla \psi|_p} \qquad \text{and} \qquad \rho_{\psi}(p) \overset{(4.10)}{\equiv} \left|\nabla \psi|_p\right|.$$

Equation (5.7) gives a recipe for constructing a description of the local surface in the geodesic ε-ball $\mathcal{N}_p$ about p in the same manner as was done on images in §5.4. Similarly, it is helpful to define the coordinate function

$$\begin{aligned} C^p_{\mathfrak{T}_\psi} : \mathcal{N}_p &\to \mathbb{C} \\ q &\mapsto \left[\mathfrak{T}_\psi(p)\right]^{-1} \log_q p \end{aligned} \tag{5.8}$$

taking neighboring points on the surface to the position of the keypoint as seen in their own frames. This allows for the the optimal filter to be re-expressed as

$$f^p_\psi(x) = \int_{\mathcal{N}_p} \rho_\psi(q)\, \delta_x\big(C^p_{\mathfrak{T}_\psi}(q)\big)\, dq. \tag{5.9}$$

Choosing ψ

In constructing planar descriptors, we take ψ to be an image. On surfaces, we assume that all we are given is the surface itself, M. Since extended convolution (and by extension the optimal filters) are defined relative to a signal on the domain, we need to recover a function $\psi \in L^2(M, \mathbb{R})$ such that the derived descriptor is both stable and descriptive.

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, *i.e.* it commutes with isometries, ensuring that the proposed descriptor is invariant under isometric deformations of the surface.

Expressing the keypoint in the frames of its neighbors

The major computational step in the construction of our proposed surface descriptor is evaluating the (Riemannian) logarithm, giving an expression of the keypoint as a point in the tangent spaces of its neighbors. State-of-the-art algorithms for computing the logarithm map parameterize the region about a given point through approaches based either on heat diffusion [SSC19a, 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 $d_g^p(q) \equiv d_g(p,q)$ to be the geodesic distance from p and using the symmetry of geodesic distances, the logarithm at q and geodesic distance from p are related by

$$\log_q\, p = -d_g^p(q) \cdot \nabla d_g^p\big|_q = -d_g^p(q) \cdot \left(\frac{\nabla d_g^p\big|_q}{\big|\nabla d_g^p\big|_q\big|} \right), \tag{5.10}$$

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

More generally, letting $d : M \times M \to \mathbb{R}_{\geq 0}$ denote any distance function on M, and setting $d^p(q) \equiv d(p,q)$, Equation (5.10) can be used to compute the coordinate function $C_\psi^p(q)$ giving the *position of p in the coordinate frame of q*:

$$C_{\mathfrak{T}_\psi}^p(q) = \left[\mathfrak{T}_\psi(q)\right]^{-1} \log_q p = -d^p(q) \cdot \left[\mathfrak{T}_\psi(q)\right]^{-1} \cdot \left(\frac{\nabla d^p\big|_q}{\big|\nabla d^p\big|_q\big|} \right) \tag{5.11}$$

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.

Algorithm 2: ECHO Surface Descriptor

```
Input:
    triangle mesh $M = (\mathcal{V}, \mathcal{T})$,
    keypoint $p \in \mathcal{V}$,
    frame field $\mathfrak{X}_\psi : \mathcal{T} \to \mathrm{U}(1)$,
    density $\rho_\psi : \mathcal{V} \to \mathbb{R}$,
    distance map $d^p : \mathcal{V} \to \mathbb{R}_{\geq 0}$,
    support radius $\varepsilon \in \mathbb{R}_{>0}$,
    descriptor radius $n \in \mathbb{Z}_{>0}$,
    Gaussian deviation $\sigma \in \mathbb{R}_{>0}$,
    quadrature degree $m \in \mathbb{Z}_{>0}$
Output:
    ECHO descriptor $\mathbf{f}_p^\psi \in \mathbb{R}^{(2n+1)\times(2n+1)}$

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

$\mathbf{f}_p^\psi = \mathbf{0}^{(2n+1)\times(2n+1)}$                          ▹ Initialize descriptor

$\forall\, t = (v_0, v_1, v_2) \in \mathcal{T}$ such that $\exists\, i$ with $d^p(v_i) \leq \varepsilon$
    $\mathcal{Q}_{\mathbf{t}}^m = \text{QuadratureSamples}(m,\ t) \subset t \times \mathbb{R}$
    $\forall\, (q,\ w_q) \in \mathcal{Q}_t^m$                         ▹ Integrate over the triangle
        $p_q = \alpha \cdot C^p_{\mathfrak{X}_\psi}(q)$                      ▹ Position of $p$ in the frame of $q$
        $\mathcal{B}_q = \{x \in \mathbb{Z}^2 \ \mid\ |x| \leq n \text{ and } |x - p_q| \leq 2\sigma\}$
        $\forall\, x = x_1 + i x_2 \in \mathcal{B}_q$                    ▹ Splat signal into vicinity of $p_q$
            $\mathbf{f}_\psi^p(x_1, x_2) \mathrel{+}= \rho_\psi(q) \cdot w_q \cdot k_{x,\sigma}(p_q)$
```

5.5.1 Construction

We address the computation of the ECHO descriptor with respect to a discrete representation of surfaces as triangle meshes, $M = (\mathcal{V}, \mathcal{T})$. As is standard, a signal $\psi : \mathcal{V} \to \mathbb{R}$ is represented by its values at the vertices and is extended by linear interpolation to the interior of triangles. Vector and frame fields are represented using a constant value per triangle.

Given a vertex $p \in \mathcal{V}$, our goal is to compute the discretized descriptor $\mathbf{f}_p^\psi \in \mathbb{R}^{(2n+1)\times(2n+1)}$, sampled on a $(2n+1)\times(2n+1)$ grid. To do this, we need to discretize the coordinate function $C^p_{\mathfrak{T}_\psi}$, giving the position of the keypoint in the frames of its neighbors, and the density ρ_ψ, and we need to estimate the integral in Equation (5.9).

To define a discretized coordinate function and density, we need to represent $C^p_{\mathfrak{T}_\psi}$ and ρ_ψ 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 $\mathcal{T}_q = \{t \in \mathcal{T} \,\big|\, t \ni q\}$ denote the subset of triangles incident on $q \in \mathcal{V}$, we define the coordinates of the position of vertex p in the tangent space of vertex q (relative to the prescribed fame field) as:

$$C^p_{\mathfrak{T}_\psi}(q) = -d^p(q) \cdot \frac{\sum\limits_{t\in\mathcal{T}_q} |t| \cdot \left[\mathfrak{T}_\psi(t)\right]^{-1} \cdot \left(\frac{\nabla d^p\big|_t}{\left|\nabla d^p\big|_t\right|}\right)}{\left|\sum\limits_{t\in\mathcal{T}_q} |t| \cdot \left[\mathfrak{T}_\psi(t)\right]^{-1} \cdot \left(\frac{\nabla d^p\big|_t}{\left|\nabla d^p\big|_t\right|}\right)\right|}, \tag{5.12}$$

where $|t|$ is the area of triangle t. Similarly, we define the vertex-based signal $\rho_\psi : \mathcal{V} \to \mathbb{R}_{\geq 0}$ by taking the area-weighted average of the magnitudes of the gradients adjacent to a vertex q:

$$\rho_\psi(q) = \frac{\sum\limits_{t\in\mathcal{T}_q} |t| \cdot |\nabla\psi|_t|}{\sum\limits_{t\in\mathcal{T}_q} |t|}. \tag{5.13}$$

In discretizing the integral, we replace the delta function with the same compactly supported kernel approximating a Gaussian with deviation σ as in Equation (5.4). Then, we approximate the integral over each triangle using m-th degree Gaussian quadrature samples [Cow73]. Specifically, setting $Q_t^m \subset t \times \mathbb{R}$ to be the (finite) set of quadrature points and quadrature weights of degree m, we approximate the integral over the local neighborhood 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:

$$\begin{aligned}
\mathbf{f}_\psi^p(x_1, x_2) &= \int_{q\in\mathcal{N}_p} \rho_\psi(q) \cdot k_{x,\sigma}\big(C_{\mathfrak{I}_\psi}^p(q)\big)\, dq \\
&= \sum_{t\in\mathcal{T}} \int_{q\in t\cap\mathcal{N}_p} \rho_\psi(q) \cdot k_{x,\sigma}\big(C_{\mathfrak{I}_\psi}^p(q)\big)\, dq \\
&\approx \sum_{t\in\mathcal{T}} \sum_{\{(q,w_q)\in Q_t^m \,|\, q\in\mathcal{N}_p\}} \rho_\psi(q) \cdot\ w_q \cdot k_{x,\sigma}\big(C_{\mathfrak{I}_\psi}^p(q)\big).
\end{aligned}$$

Pseudocode for the computation of the discrete descriptor is shown in **Algorithm 2**: we iterate over all triangles $t \in \mathcal{T}$ which have at least one vertex within a distance of ε of the keypoint; for each triangle t, we compute the set of m-th degree quadrature samples Q_t^m; for each quadrature point q, we compute p_q, the position of p in the coordinate frame of q, scaled to the descriptor resolution; we also compute $\mathcal{B}_q$ – the set of grid points that fall within a disk of radius n of the origin and whose kernel function supports p_q; for each grid point $x \in \mathcal{B}_q$, we increment the value of the descriptor at x by the value of the density at the quadrature point, $\rho_\psi(q)$, weighted by the quadrature weight, w_q, and the value of the kernel function

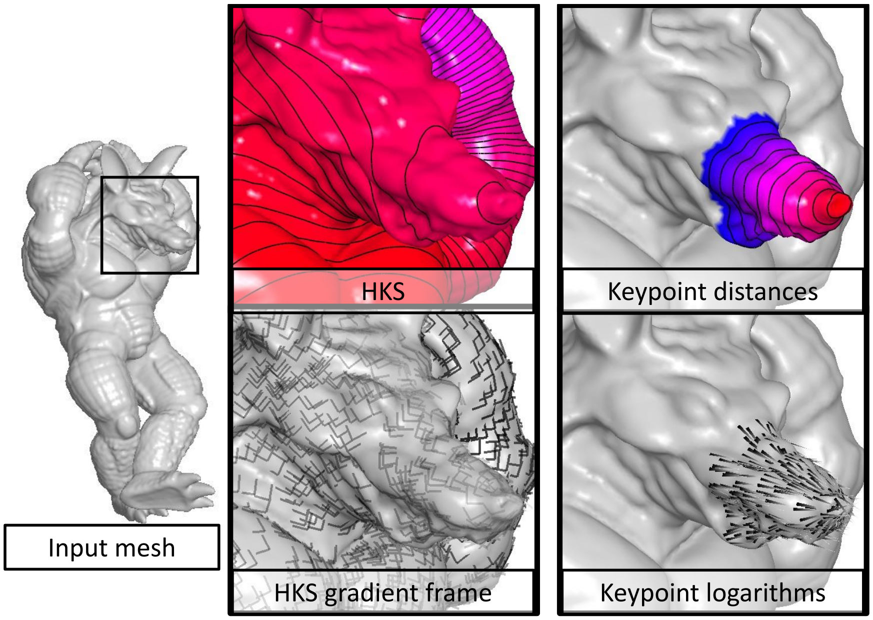


Figure 5-4. Visualizations of the input geometry (left), the Heat Kernel Signature and derived frame field (middle), and the geodesic distances from the keypoint and the keypoint's logarithm in its neighbors' tangent frames (right). For the visualization of the signals, red corresponds to lower values 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.

centered at x, evaluated at the position of p in the coordinate frame of quadrature point, $k_{x,\sigma}(p_q)$.

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 d^p it is desirable to choose a method that allows the calculation to be truncated to exclude points whose distance from p exceeds ε. Assuming that the set of points in M within a distance of ε of the keypoint 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 neighborhood. An example of the computed geodesic distances and derived logarithms is shown on the right in Figure 5-4.

5.5.2 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 [TSDS10a, STDS14], RoPS [GSB+13], USC [TSDS10b], 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.

The size of the support radius ε 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

$$\varepsilon = 0.08\sqrt{A / \pi}\,, \tag{5.14}$$

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 5-4 (middle).

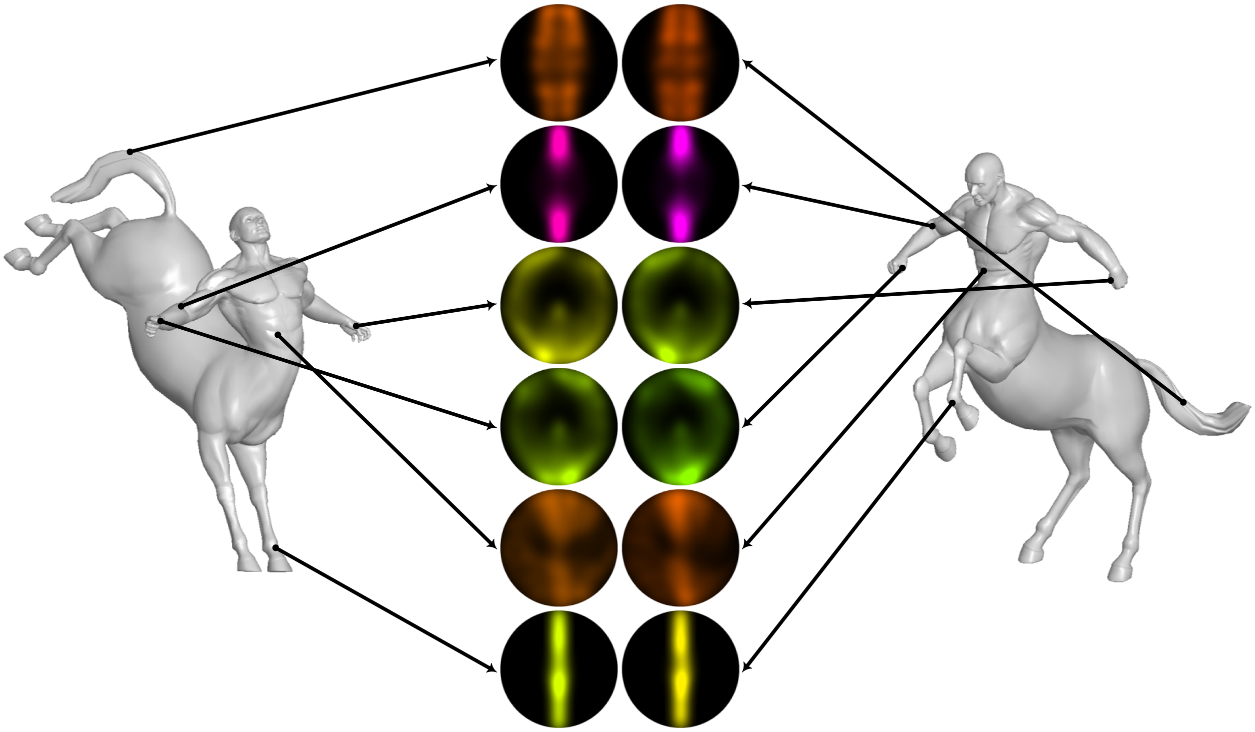


Figure 5-5. Visualizations of biharmonic ECHO descriptors computed at corresponding points on two meshes from the `centaur` class of the TOSCA dataset [BBK08]. ECHO's intrinsic construction and use of frame fields enables rich and distinctive characterizations that remain consistent in the presence of 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.

5.5.2.1 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 5-5.

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 (5.14). 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, in exactly the same manner as was done in evaluation of the image descriptions. Given a scene keypoint, p and corresponding descriptor $\mathbf{f}^p$, all keypoints from across all models are sorted based on the descriptor distance. Then letting $\mathcal{M}_p$ be the set of all model keypoints that are valid matches with p, precision and recall are defined as functions of the top r model keypoints as in Equation (5.6).

We note that 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 p is scaled by a factor of E_p / E where E_p is the average length of the edges incident on 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. Specifically, we use OpenFlipper's [MK10] incremental mesh decimation module with the decimation priority determined by the distance to the original mesh.

5.5.2.2 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 triangulations, occluded geometry and various other sources of interference [DSL+19]. The dataset contains 76 pre-defined pairs of meshes partitioned into four increasingly

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

Table 5-I. Mean descriptor run time over all model keypoints in the feature matching experiments. The ‘null’ meshes contain between 5,000 and 53,000 vertices, with an average of approximately 30,000 vertices.

challenging test sets: (1) articulations and rigid deformations, (2) near-isometric deformations, (3) non-isometric deformations and (4) topological and geometric changes. The authors provide 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. Specifically, each pair 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 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, q^*, and the model vertex q with the smallest descriptor distance,

$$d_g(q*, q)\sqrt{\pi / A}, \tag{5.15}$$

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

5.5.2.3 Complexity

The mean run time for each descriptor over all model keypoints in the feature matching experiments is shown in Table 5-I. 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.

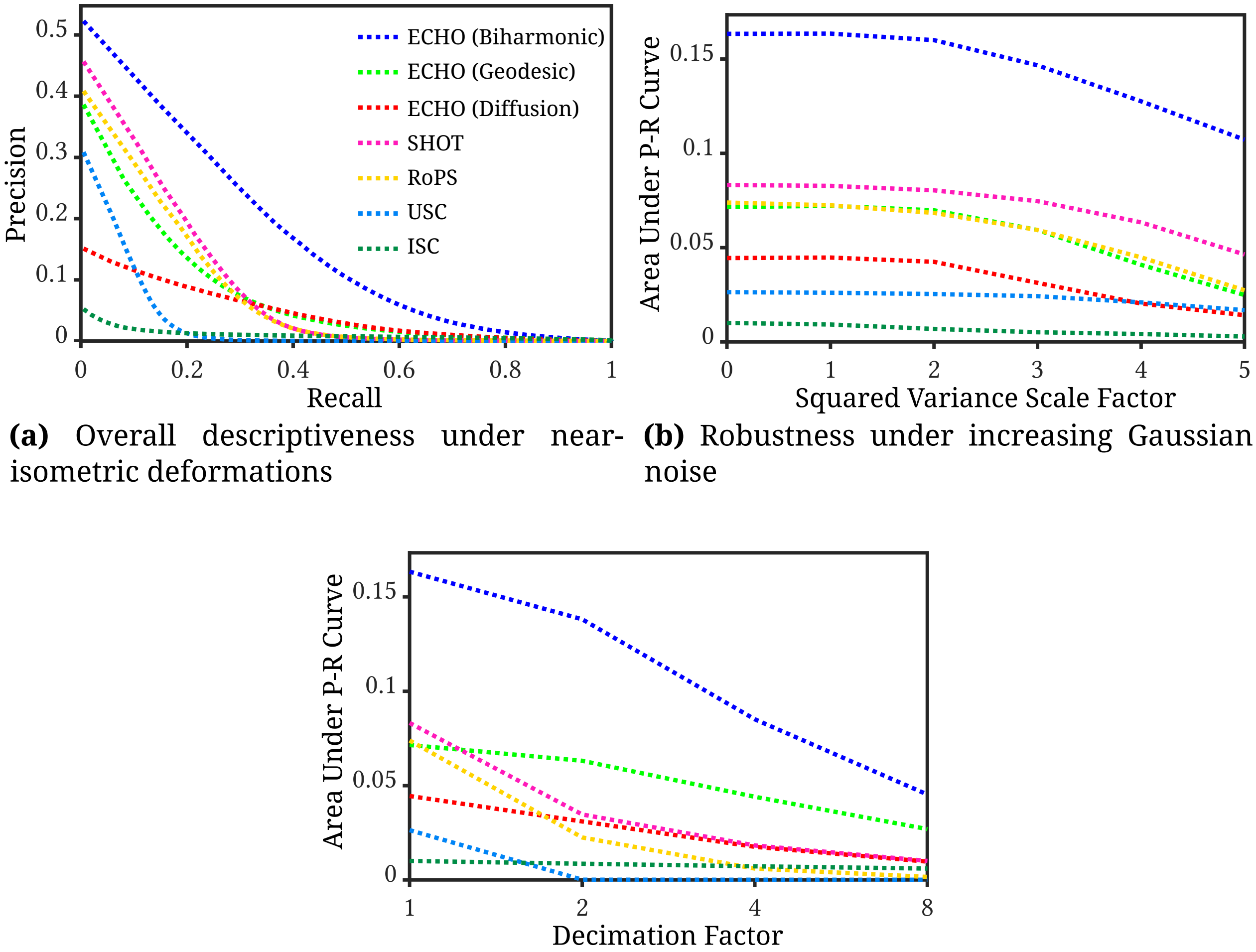
5.5.3 Results and discussion

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

5.5.3.1 TOSCA

The results of the feature matching experiments using the TOSCA dataset are shown in Figure 5-6. 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 5-6a.

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.



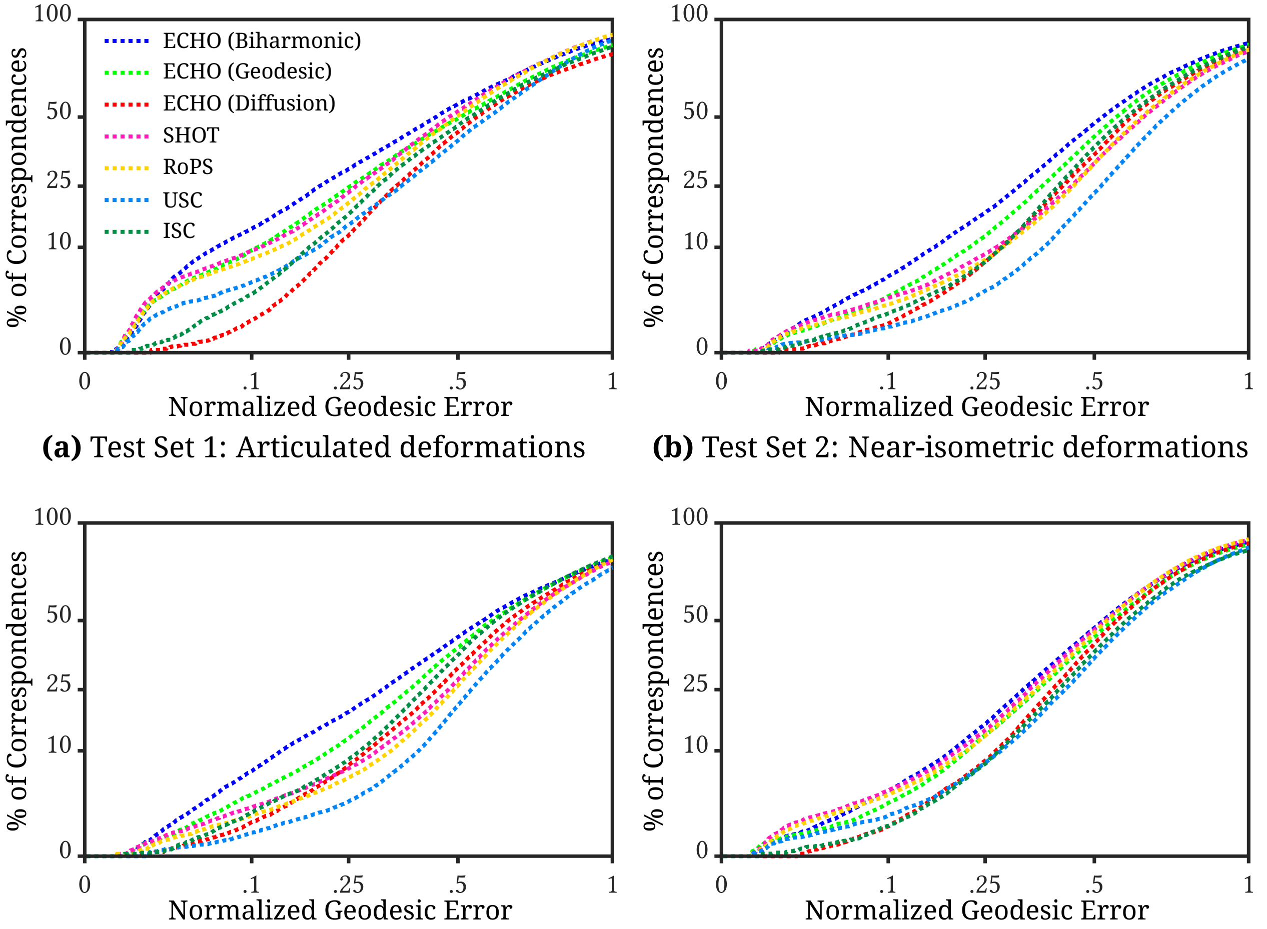
(a) Overall descriptiveness under near-isometric deformations

(b) Robustness under increasing Gaussian noise

(c) Robustness under decreasing mesh resolution

Figure 5-6. 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.

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 5-6b. 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.



(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 changes

Figure 5-7. 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

Descriptor matching performance relative to changes in mesh resolution is shown in Figure 5-6c, 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.

5.5.3.2 SHREC ’19 Shape Correspondence Benchmark

The results of the sparse correspondence experiments using the SHREC 2019 Shape Correspondence Benchmark dataset are shown in Figure 5-7. For each test set, the

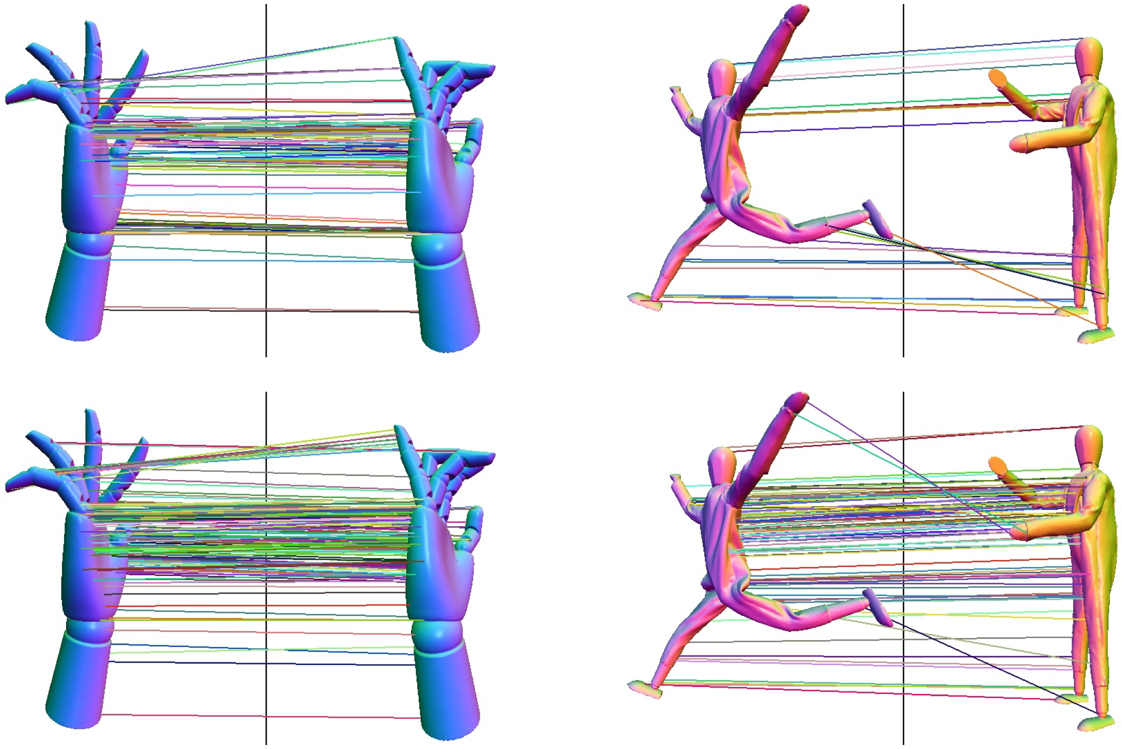


Figure 5-8. Visualizations of sparse correspondences with a normalized geodesic error $\leq$ 0.08 found by the SHOT (top row) and biharmonic ECHO (bottom row) descriptors between two example mesh pairs from the first (left column) and third (right column) test sets in the SHREC 2019 Shape Correspondence 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.

.

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, which correspond to articulated, near-isometric, and non-isometric shape deformations, are shown in Figures 5-7a, 5-7b, and 5-7c, 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 5-8.

The error curves for test set 4, which considers topological and geometric changes, are shown in Figure 5-7d. 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 changes [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 of a variety of challenging nuisance factors commonly found in real data.

5.5.3.3 Discussion

The local shape descriptor we compute at the point 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^p and ψ, and from the latter derive and an intrinsic frame field $\mathfrak{X}_\psi$ and density ρ_ψ. Here d^p should depend on the point p being described. We define a histogram characterizing the point p by aggregating information from neighboring points q. Each point q contributes a vote with weight $\rho_\psi(q)$ into the bin describing the position of p relative to q. Expressing the position in polar coordinates, the radius is given by $d^p(q)$ and the angle is defined to be the direction of the tangent vector $\nabla d^p\big|_q$ in the frame of $\mathfrak{X}_\psi$.

Here we consider several choices of d^p including geodesic, diffusion, and biharmonic distances from p, with ψ taken to be the Heat Kernel Signature. 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 ψ 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^p, ρ_ψ, and $\mathfrak{X}_\psi$ so long as the latter two maps satisfy the equivariance conditions in Equation (4.5). In particular, we require that $\rho_\psi = 0$ whenever $\mathfrak{X}_\psi$ vanishes so that the descriptor remains well-defined even when the angular component of the polar coordinates of p relative to 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, DSL$^+$19] or regularize [VLB$^+$17] the registration process.

5.6 Conclusion

In this chapter we proposed a novel family of local image and surface descriptors, which correspond to rasterizations of the optimal filter maximizing the response of the extended convolution. We evaluated the performance of our proposed descriptors against that of the premier image and surface descriptors. The ECHO and SIFT image descriptors achieve comparable performance on a challenging, large-scale image dataset. 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.

Chapter 6

Field Convolutions for Surface CNNs

6.1 Introduction

The advent of deep learning in imaging, vision, and graphics has coincided with the development of numerous techniques for the analysis and processing of curved surfaces based on convolutional neural networks (**CNNs**). The challenge in reproducing the success of CNNs on surfaces is that classical notions of convolution and correlation in Euclidean spaces cannot simply be transposed onto curved domains. Unlike images, points on a surface have no canonical orientation, without which simple operations fundamental to the spatial propagation of information, such as moving dot products, cannot be computed in a repeatable manner.

Geometric deep learning is a young field, and many successful methods can be broadly categorized in relation to two emerging paradigms characterized by specific approaches to convolution: *diffusive* propagation and *equivariant* propagation. Diffusive approaches closely intertwine convolution operations with heat diffusion on manifolds wherein filters represented by anisotropic heat kernels or Gaussians are used to propagate scalar features [MBBV15, BMR+16, BMRB16, MBM+17, LLHL20, SACO20]. In contrast, equivariant convolutions distribute tensor features that transform with local coordinate systems [PO18, TSK+18, SDL18, CWKW19a, dHWCW20, WEH20, YLP+20].

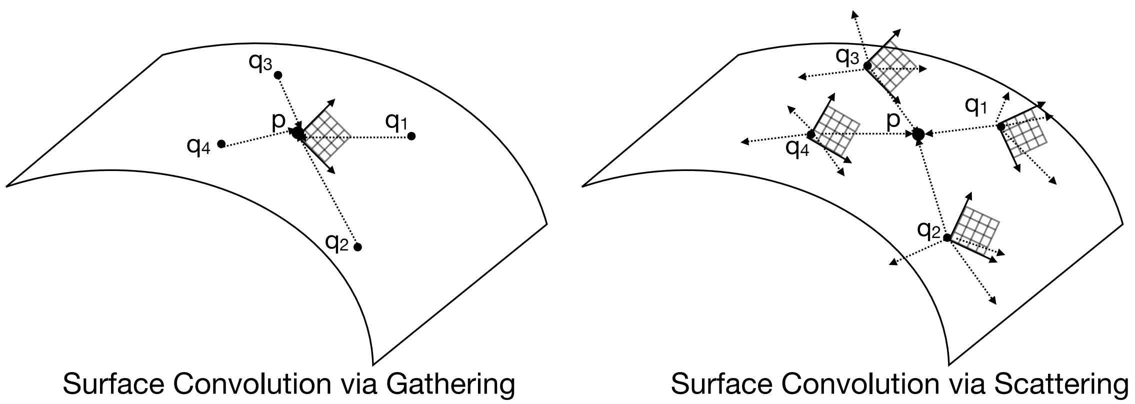

q3
p
q1
q4
q2
Surface Convolution via Gathering
q3
q1
p
q4
q2
Surface Convolution via Scattering

Figure 6-1. Prior approaches define patch-based convolution operators as *gathering* operations (left), which are sensitive to noise or disruptions in the local coordinate system. Field convolution is a *scattering* operation and is robust under perturbations as it does not rely on a single coordinate system to aggregate features.

Critically, virtually all state-of-the-art approaches sacrifice filter descriptiveness to define a notion of convolution that does not depend on the choice of local coordinate frames. Gaussian filters can facilitate efficient evaluations in the spectral domain but are individually undiscriminating. Equivariant approaches have the potential to provide expressive notions of convolution on surfaces due to the encoding of geometric information in the transport of tangent vector features. However, equivariance of the response is almost universally achieved by placing constraints on the filters themselves [PO18, PRPO19, CWKW19b, dHWCW20, HJZS20, WEH20], limiting descriptiveness and necessitating complex architectures to support the algebraic relationships between kernels. Furthermore, these regimes formulate spatial propagation as a *gathering* operation, analogous to correlations on Euclidean domains; features are weighted based on their position relative to a coordinate frame defined at a single point (Figure 6-1, left), making them sensitive to inconsistencies or disruptions in local parameterizations.

Field convolution (§4.2) generalizes extended convolution on surfaces to an operator on vector fields well-suited for CNNs on surfaces. Features are combined by having each neighbor q parameterize p within its own coordinate frame (Figure 6-1, right). This formulation combines intrinsic spatial weighting with parallel transport *while placing no constraints on the filters themselves*, providing a definition of convolution that commutes with the action of isometries (Claim 10) and has increased descriptive potential. In addition, as a *scattering* operation, it is less sensitive to noise and other nuisance factors as it does not rely on a single coordinate system about each point to aggregate features.

Field convolutions are flexible and straight-forward to incorporate into surface learning frameworks. Parallel transport captures additional geometric information and and their highly discriminating nature has cascading effects throughout the learning pipeline, allowing us to achieve state-of-the-art results on standard benchmarks in applications including shape classification, segmentation, correspondence, and sparse matching.
All code and evaluations are publicly available at github.com/twmitchel/FieldConv.

6.2 Related work

The field of geometric deep learning has grown extensively since its inception. Here, we only review the techniques most closely related to ours – those designed specifically for the analysis of 3D shapes. Generally speaking, these methods exist on a spectrum between extrinsic and intrinsic techniques, with the former performing signal processing using the embedding of the surface in 3D and the latter only using the Riemannian structure.

Point-based methods offer a purely extrinsic framework for applying deep learning to 3D shapes by representing them in terms of point clouds. A majority of these approaches can trace their lineage to the influential PointNet [QSMG17] and PointNet++ architectures [QYSG17] and recent approaches such as DGCNN [WSL+19], PCNN [AML18], KPCNN [TQD+19], TFN [TSK+18], QEC [ZBL+20] and SPHNet [PRPO19] have sought to extend the framework by incorporating connectivity information, dynamic filter parameterizations, and equivariance to rigid transformations. Convolution is typically expressed by applying radially isotropic filters over local 3D neighborhoods and aggregating the results with the maximum or summation operations. This approach offers a simple foundation for extremely flexible and noise-robust networks, though at the expense of descriptive potential. More generally, these methods tend to struggle in the presence of non-rigid isometric deformations, making them less effective in scenarios like deformable shape matching [DSO20, GFK+18, SACO20].

Representational approaches sit between extrinsic and intrinsic techniques. These methods exploit the data's underlying connectivity to form convolutional operators, often making use of well-developed techniques for graph-based learning on irregular structures [DBV16, YSGG17, VBV18, FELWM18, LMBB18, GCBZ19, CWKW19b, LT20]. In particular, convolutions are performed using filters defined relative to the explicit graph structure as functions on edges or vertices, often with only immediate local support such as the surrounding one-ring. A particularly notable example is MeshCNN [HHF+19], which specifically leverages the ubiquitous representation of surfaces as triangle meshes to construct a similarity-invariant convolution operator propagating edge-based features. While this enables graph-based convolutions to better handle non-rigid deformations compared to point-based approaches, it also makes them sensitive to changes in connectivity.

One approach to defining *intrinsic* convolution has been to parametrize the surface over a simple domain such as the the sphere [HSBH+19], torus [MGA+17], or plane [SBR16] where standard CNNs can be applied. However, such parameterizations depend on the genus and often exhibit significant distortion.

A second class of approaches has been to define intrinsic convolution over the Riemannian manifold, and can generally be classified in relation to two emerging paradigms: *diffusive* convolutions and *equivariant* convolutions. In the former, convolution operations are closely related to heat diffusion on surfaces wherein heat (e.g. Gaussian) kernels are used to propagate scalar features. While early diffusive approaches including GCNN [MBBV15], ADD [BMRB16], ACNN [BMR+16] and MoNet [MBM+17] perform convolutions over local patches, recent state-of-the-art networks ACSCNN [LLHL20] and DiffusionNet [SACO20] represent convolution in the spectral domain. Despite their success in a variety of scenarios, most notably in dense shape correspondence [GFK+18, DSO20, LLHL20, SACO20], these methods face an intractable problem: radially symmetric filters are individually undiscriminating and diffusive frameworks are not naturally suited to handle the orientation ambiguity problem introduced by the use of more descriptive, anisotropic kernels. To compensate, these methods supplement convolutions with basic orientation-aware operations on tangent vector features [SACO20] in addition to employing various strategies that are either fragile, such as aligning kernels along the directions of principal curvature [BMR+16, MBM+17], or discarding information by pooling over samplings of orientations or by specifying directions of maximum activation[MBBV15, LLHL20].

Recently, several techniques have been introduced for *equivariant* surface convolutions such as MDGCNN [PO18], GCN [CWKW19a, dHWCW20] and HSN [WEH20]. In contrast to diffusive approaches, equivariant convolutions are designed specifically to address the rotation ambiguity problem by propagating tangent vector features that transform with local coordinate systems. To make the convolution independent of the choice of local coordinate frame, most existing methods strongly constrain the class of filters that can be used [PO18, PRPO19, CWKW19b, dHWCW20, HJZS20, WEH20]. An exception to this is PFCNN [YLP+20] which also discards information by pooling over multiple kernel orientations. Often, these parameterizations are so restrictive that they necessitate complex network architectures to be effective: even the state-of-the-art HSN [WEH20] is formulated as a multi-stream U-Net with various pooling operations. Furthermore, in moving from radially isotropic to anisotropic filters, prior equivariant regimes universally formulate spatial propagation as a *gathering* operation wherein all features in the local surface are weighted based on their position in a *single* coordinate system. While this approach may seem natural as it is analogous to correlation on Euclidean domains, a feature's dependence on a single local parameterization increases sensitivity to noise.

6.3 Method overview

Field convolutions facilitate the construction of highly discriminating yet simple networks, without the need for pooling, normalization, or specialized architecture. We first discuss the discretization of field convolution as defined in Equation (4.19), relative to surfaces represented as triangle meshes. The principal module in applications is the field convolution ResNet (**FCResNet**) block, consisting of two successive field convolutions with a residual connection between the input and output layers [HZRS16]. FCResNet blocks are self-contained and flexible, and can easily be incorporated into isometry-invariant surface learning regimes. In addition, we leverage the connection between field convolutions and ECHO descriptors to construct a novel final layer specifically designed for labeling tasks with isometry-invariant surface networks, which we refer to as an ECHO block. This block takes vector field channels as input, mapping them to scalar ECHO descriptors which are then fed through an MLP to make predictions, essentially converting the problem to one of image classification in the final layer of the network.

6.4 Discretization

Given a surface M, recall from §4.2 that we can express the evaluation of a vector field $X \in \Gamma(TM)$ at p as the complex number

$$X(p) \equiv \varrho_p \, e^{i\phi_p} \in T_pM,$$

and that the logarithm and transport operators can be written in polar form as

$$\log_p q \overset{(2.13)}{\equiv} r_{pq} \, e^{i\theta_{pq}} \qquad \text{and} \qquad \mathcal{P}_{p\leftarrow q}(\mathbf{v}) \overset{(2.14)}{\equiv} e^{i\varphi_{pq}} \, \mathbf{v},$$

for all $q \in M$ and $\mathbf{v} \in T_pM$. Then, the field convolution of a vector field X with a compactly-supported planar filter $f \in L^2(\mathbb{C}, \mathbb{C})$ can be expressed concretely as the vector field in $\Gamma(TM)$ with

$$(X * f)(p) \overset{(4.19)}{=} \int_{\mathcal{N}_p} \varrho_q \, e^{i(\phi_p + \varphi_{pq})} \, f\left(r_{qp} \, e^{i\left(\theta_{qp} - \phi_q\right)}\right) \, dq.$$

where $\mathcal{N}_p \subset M$ denotes the geodesic ε-ball about p.

In practice, we discretize a surface M by a triangle mesh with vertices $\mathcal{V}$. To every $p \in \mathcal{V}$, we associate the collection of vertices $\mathcal{N}_p \subset \mathcal{V}$ belonging to the geodesic ε-ball about p. At each point, real-valued filters $f \in L^2(\mathbb{C}, \mathbb{R})$ are supported on $\log_p(\mathcal{N}_p) \subset \mathbb{C}$, and parameterized as sums of angular frequencies with band-limit B. That is, for any $z = re^{i\theta} \in \mathbb{C}$ with $|r| \leq \varepsilon$, the evaluation of f at z is expressed as

$$f(z) = \sum_{m=-B}^{B} f_m(r) \cdot e^{im\theta} \tag{6.1}$$

where $f_m(r) \in \mathbb{C}$ is the m-th Fourier coefficient of f, restricted to radius r and $f_{-m}(r) = \overline{f_m(r)}$ because f is real-valued. We discretize the function $f_m(r)$ using linear interpolation, setting $f_m(r) = \mathbf{r}^\top(r) \, \mathbf{f}_m$, where $\mathbf{r}(r) \in \mathbb{R}^N$ is the vector of linear interpolation weights (with $\mathbf{r}_i(r) \neq 0$ only if $i \in \{\lfloor rN/\epsilon \rfloor, \lceil rN/\epsilon \rceil\}$) and $\mathbf{f}_m \in \mathbb{C}^N$ is the vector of Fourier coefficients at the discrete radii.

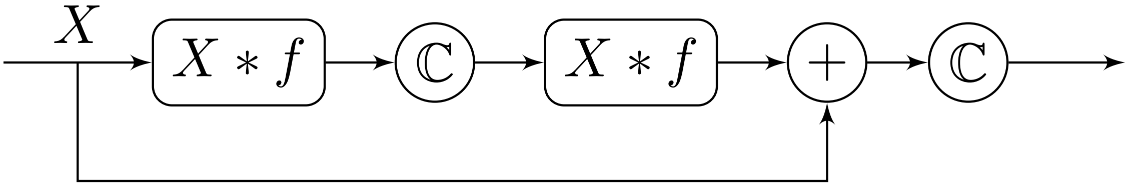


Figure 6-2. The FCResNet block. Here $\mathbb{C}$ denotes the complex ReLU in Equation (6.6).

Then, letting $\{w_p\} \subset \mathbb{R}_{>0}$ denote the area weights associated with vertices $p \in \mathcal{V}$ and letting $X \in \mathbb{C}^{|V|}$ be a discrete vector field, the evaluation of the field convolution $X * f$ at a vertex $p \in \mathcal{V}$ is given by

$$\big(X * f\big)(p) = \sum_{\substack{q \in \mathcal{N}_p \\ |m| \le B}} w_q \, \varrho_q \, e^{i(\phi_q + \varphi_{pq})} \, f_m(r_{qp}) \, e^{im(\theta_{qp} - \phi_q)}. \tag{6.2}$$

The values of w_q, φ_{pq}, r_{qp}, and θ_{qp}, corresponding to the weight, transport change of angle, geodesic distance, and logarithm for each $p \in \mathcal{V}$ and $q \in \mathcal{N}_p$ can be precomputed to speed up training. Similar to [WEH20], we apply rotational offsets $e^{i\mathrm{sgn}(m)\beta_{|m|}}$ to the coefficients corresponding to each frequency, providing additional learned degrees of freedom.

6.5 Surface CNNs with field convolutions

The goal of this section is to introduce the fundamental building blocks for incorporating field convolutions into isometry-invariant surface learning paradigms.

FCResNet Blocks

The atomic unit for field convolutions in surface CNN frameworks is the FCResNet block, which consists of two field convolutions each followed by a non-linearity and a residual connection between the input and output streams (Figure 6-2). They are entirely self-contained, and map vector field features to vector field features without relying on any supporting or complementary convolution operations that are a common fixture in other equivariant approaches [PO18, WEH20]. As such, they represent a flexible and descriptive layer that can be easily employed in isometry-invariant learning pipelines.

Learned gradients

In practice, inputs to surface CNNs are often scalar features, such as the raw 3D positions of points. To lift such features to a vector field, we use a learnable operation analogous to a weighted gradient calculation. For any function $\psi \in L^2(\mathcal{V}, \mathbb{R})$ we learn the magnitude and direction of its "gradient" separately, with respect to compactly supported radially isotropic filters $f_1,\ f_2 \in L^2(\mathbb{R}_{\geq 0}, \mathbb{R})$. That is, we learn the vector field $\Phi_{f_1} : \mathcal{V} \to \mathbb{C}$ and scalar field $P_{f_2} : \mathcal{V} \to \mathbb{R}$ with

$$\Phi_{f_1}(p) = e^{i\beta} \sum_{q \in \mathcal{N}_p} w_q \ (\psi(q) - \psi(p)) \ f_1(r_{pq})\, e^{i\theta_{pq}}, \tag{6.3}$$

$$P_{f_2}(p) = \sum_{q \in \mathcal{N}_p} w_q\, \psi(q)\, f_2(r_{pq}) \tag{6.4}$$

with w_q, r_{pq}, θ_{pq} defined as in Equation (6.2) and β a learnable rotational offset. Using these, we define the "gradient" of ψ with respect to f_1 and f_2 as the vector field

$$P^2_{f_2}(p) \frac{\Phi_{f_1}(p)}{\left\|\Phi_{f_1}(p)\right\|} \tag{6.5}$$

ECHO Blocks

A secondary contribution of this work is the concept of an ECHO block for label-prediction tasks, which leverages the connection between vector fields and the ECHO surface descriptor. Given a scalar signal and a frame field, ECHO descriptors provide an intrinsic, isometry-invariant characterization of the local surface about a feature point. A vector field can be used to compute ECHO descriptors at every point $p \in M$, using the magnitude and direction of each vector to define the values of the density and frame field at p.

The idea behind ECHO blocks is to convert feature vector fields to pointwise descriptors, turning the task of vector field classification into one of image classification in the final layer of the network. These blocks consist of two steps: 1) A field convolution layer is used to map the input feature channels to D output feature channels (with D the desired number of descriptors). These are then used to compute pointwise ECHO descriptors, resulting in H isometry-invariant scalar features per channel, where H is the number of samples used to represent the ECHO descriptor. 2) The $D \times H$ values are linearized and fed to a three-layer MLP. Here, we construct ECHO descriptors by splatting the contributions of individual vertices into the histogram, rather than integrating over individual triangles in the manner described in §5.5.1. The computation relies only on the logarithm map and the integration weights associated with each vertex, so no additional pre-processing is required.

Linearities and Non-Linearities

Since we represent tangent vector features as complex numbers, we apply linearities in the form of multiplication by complex matrices in the same manner as is done for real-valued features. However, our linearities do not include translational offsets to preserve commutativity with the action of isometries.

For similar reasons, non-linearities are applied only to the radial components of features as is done in [WEH20]. Namely, given a feature vector field $X \in \mathbb{C}^{|V|}$ we apply pointwise ReLUs with a learned offset b such that

$$\mathrm{ReLU}_b\big(X(p)\big) = \mathrm{ReLU}\left(\varrho_p + b\right)\, e^{i\phi_p}. \tag{6.6}$$

FCNet: A Generic Surface CNN for Vector Fields

In our experiments we use a simple, generalizable architecture we call an **FCNet**, which is simply a series of FCResNet blocks. For labeling tasks, we append an ECHO block to the end of the network to make predictions. For FCNets consisting of three or more layers, we add additional residual connections after every two FCResNet blocks as we find this significantly accelerates training. In all experiments, we take the raw 3D positions of points as inputs and use a learnable gradient layer to map them to vector fields which are then fed to the network. We could also use the intrinsic Heat Kernel Signature [SOG09] as input, thereby obtaining a fully isometry-invariant pipeline. However, as demonstrated by Sharp *et al.* [SACO20], the 3D coordinates work as well in practice and are easier to compute.

Despite this elementary construction, we show that FCNets achieve state-of-the-art results in a variety of fundamental geometry processing tasks.

6.6 Evaluation

We compare our method against leading surface learning paradigms on four benchmarks corresponding to fundamental tasks in geometry processing: classification, segmentation, correspondence, and feature matching.

6.6.1 Implementation

Our framework is implemented using PyTorch Geometric [FL19]. We employ the same, simple FCNet architecture discussed in Section 6.5 in all of our experiments, varying the number of FCResNet blocks based on task complexity. For label-prediction tasks on large datasets, we append an ECHO block to the end of the network to make predictions. Otherwise we use the magnitudes of the output feature vectors.

Method	Accuracy
FC (ours)	**99.2%**
DiffusionNet [SACO20]	98.9%
MeshWalker [LT20]	97.1%
HSN [WEH20]	96.1%
MeshCNN [HHF+19]	91.0%
GWCNN [ESKBC17]	90.3%

Table 6-I. Classification accuracy on the SHREC '11 Dataset [LGB+11] .

As input, we take the 3D coordinates, which are lifted to 16 tangent vector features in the initial gradient layer, followed by either 32 or 48 features in the FCResNet stream. We use the ADAM optimizer [KB15] to a cross-entropy loss with an initial learning rate of 0.01 and a batch size of 1. We randomly rotate all inputs to ensure there are no consistencies in the spatial embedding of shapes.

Our pre-processing regime parallels [WEH20], omitting the operations necessary to support their multi-scale and pooling operations. All shapes are normalized to have unit surface area and we use the Vector Heat Method [SSC19b] to compute the geodesic ε-ball $\mathcal{N}_p \subset \mathcal{V}$ corresponding to each vertex $p \in \mathcal{V}$, in addition to the logarithm and parallel transport associated with each edge $(p, q) \in \{p\} \times \mathcal{N}_p$. Area weights are assigned in the standard way, using one third of the vertex's one-ring area, and are normalized by the sum of the weights within the geodesic ϵ-ball. While we process shapes as triangle meshes in our experiments, we note that recent work by Sharp *et al.* [SC20] has made possible efficient computations of logarithmic parameterizations and vector transport on point clouds, with which our method can be extended to analyze point cloud shape data.

6.6.2 Classification

First, we use an FCNet with two FCResNet blocks to classify meshes in the SHREC '11 dataset [LGB+11], containing 30 shape categories. Filters are supported on geodesic neighborhoods of radius $\epsilon = 0.2$ and are parameterized using $N = 6$ radial samples

Method	# Features	Accuracy
FC (ours)	3	**92.9%**
MeshWalker [LT20]	NA	92.7%
MeshCNN [HHF+19]	5	92.3%
DiffusionNet [SACO20]	16	91.5%
HSN [WEH20]	3	91.1%
SNGC [HSBH+19]	3	91.0%
PointNet++ [QSMG17]	3	90.8%

Table 6-II. Segmentation accuracy on the composite dataset of [MGA+17].

with band-limit $B = 2$. Due to the small scale of the task, we omit the ECHO block in the final layer and instead use a global mean pool over the feature magnitudes to give a prediction. As in prior works [HHF+19, WEH20, SACO20], we train on 10 samples per class and report results over three random samplings of the training data. Our FCNet converges quickly, and we train on just 30 epochs – far fewer than the 100 or more used in previous work.

Results are shown in Table 6-I. Due to the wide adoption of the dataset, we only list the results of methods achieving a classification accuracy of 90% or higher. Our simple FCNet achieves the highest reported accuracy, reaching a classification rate of 100% on two of the three random samplings of the training data. Like HSN and DiffusionNet who also report high classification accuracy, our FCNet uses relatively few parameters compared to other networks and is agnostic to both mesh connectivity and isometric deformations – all providing a significant advantage on the SHREC '11 dataset which has a small number of training samples and consists of poor-quality meshes with in-class deformations mainly limited to rigid articulations. The superior performance of our FCNet is likely due to the descriptiveness of field convolutions, as HSN uses specially parameterized filters.

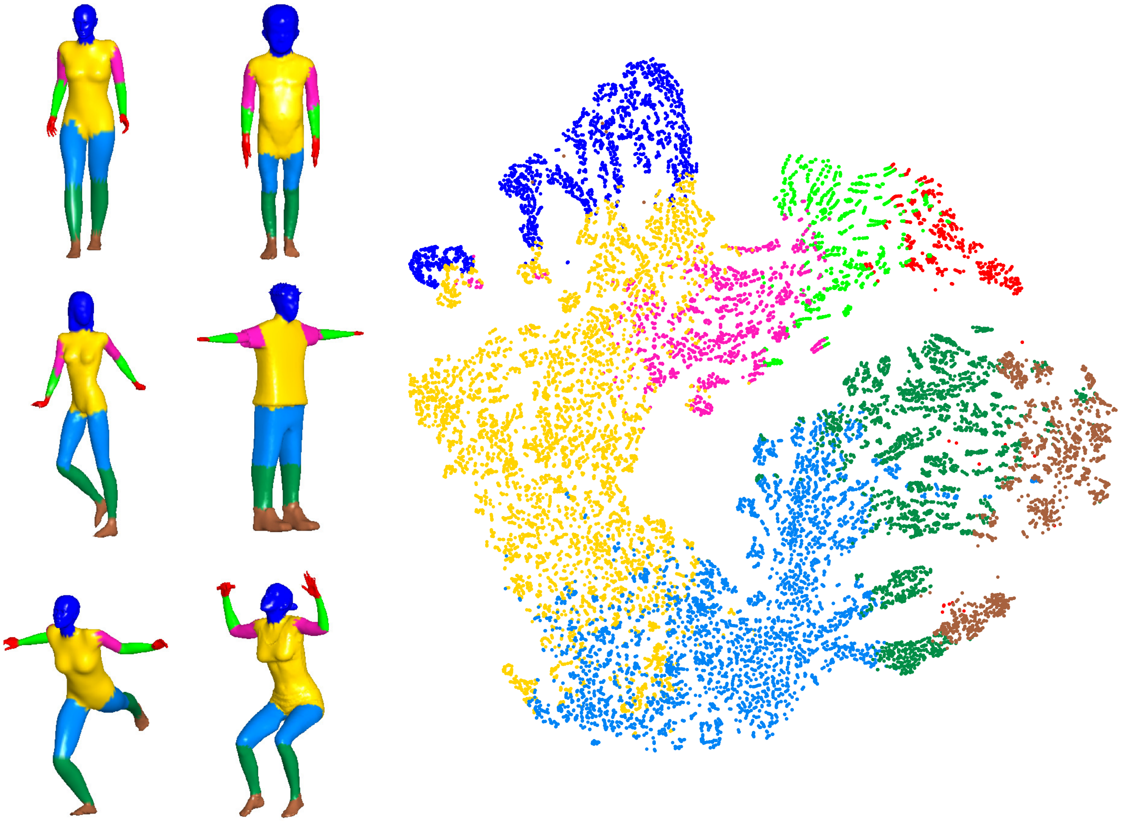


Figure 6-3. We visualize the descriptors computed in our FCNet's ECHO block in the segmentation task. Left: Models from the test split of the composite dataset [MGA+17], color-coded by ground-truth labels. Right: 2D projections of the descriptors using t-SNE [VdMH08].

6.6.3 Segmentation

Next, we apply our field convolution framework to the task of human body segmentation, using the dataset proposed by [MGA+17], which consists of a composite of various human shape datasets [Ado16, ASK+05, GBP07, VBMP08, BRLB14]. The varied nature of the collection of models in terms of human subjects, acquisition method, and connectivity serve to test both descriptiveness and robustness to variety of nuisance factors.

We use an FCNet with four successive FCResNet blocks ($N = 6$, $B = 2$, $\epsilon = 0.2$) followed by an ECHO block, trained to predict a body part annotation for each point on the mesh. The ECHO block computes $D = 32$ descriptors with $H = 33$ samples (corresponding to three samples per geodesic radius) for a total of 1056 scalar feature channels. The three-layer MLP first maps these features to 256 channels, then 124, and finally to the desired number of output channels. Due to the large number of vertices per model, we downsample each mesh to 1024 vertices using farthest point sampling, an approach also used by [HHF⁺19, WEH20]. Our network converges quickly and we train for only 15 epochs with a label smoothing regularization [SVI⁺16] factor of 0.2.

Results in the form of the percentage of correctly classified vertices across all test shapes are shown in Table 6-II. As in the classification experiments, we only list the results of methods that achieve a segmentation accuracy of 90% or higher on the dataset. Again, our basic network achieves state-of-the-art results, outperforming all other methods with a minimal number of input features. The improvement due to field convolutions is especially evident relative to other techniques that employ surface convolutions, such as HSN [WEH20] and DiffusionNet [SACO20] approaches.

To understand the features learned by our network, we use t-SNE [VdMH08, PVG⁺11] to visualize the descriptors computed in the ECHO block for all models in the test dataset, color-coded using the ground-truth labels (Figure 6-3). We observe a distinct clustering of points, not only corresponding to similarly labeled regions but also reflecting the connectivity between adjacent regions on the meshes. This suggests that our FCNet is able to learn at least some measure of intrinsic similarities between shapes, despite starting with the extrinsic 3D coordinates as input.

6.6.4 Correspondence

Here we use an FCNet to find pointwise correspondences between similar shapes. Over the last half-decade, the FAUST dataset [BRLB14] has become the *de facto* standard for evaluating network performance in correspondence tasks and many re-

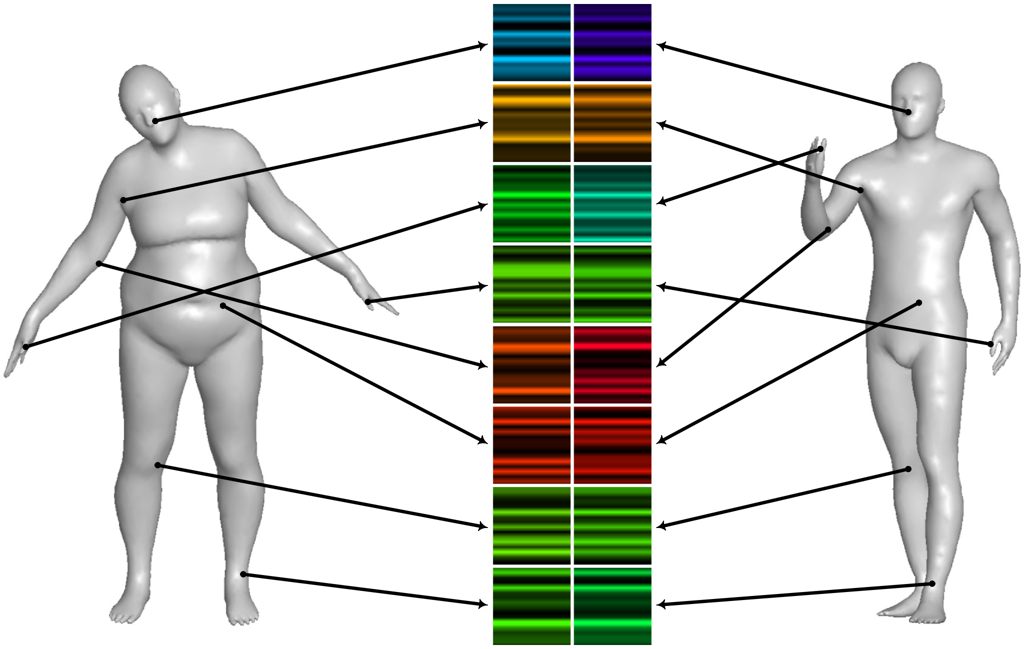


Figure 6-4. FCNet features at corresponding points on models in the remeshed FAUST dataset [DSO20]. Features are drawn using the HSV scale – hue encodes the absolute magnitude and value encodes the relative magnitude with saturation fixed at one.

cent approaches have achieved near-perfect accuracy on the dataset [FELWM18, dHWCW20, LLHL20]. However, shapes in the dataset share the same connectivity, and there has been some question as to whether these methods have primarily learned the mesh graph structure, rather than deformation-invariant characterizations of the shape themselves [SACO20]. To this point, we perform evaluations on a fully remeshed version of the dataset [DSO20], a more challenging task better representative of real-world applications. As in prior work, we train on the first 80 models out of the 100 in the dataset, and use the remainder for testing.

We train an FCNet to predict the indices of corresponding vertices on a template shape. Due to the degree of precision required by this task, we use a deeper network, consisting of eight FCResNet blocks ($N = 3$, $B = 1$, $\epsilon = 0.05$) followed by an ECHO block ($D = 12$, $H = 13$) with a 124–64–32 MLP, and additional residual connections after every two FCResNet blocks. To make predictions, we add two linear layers

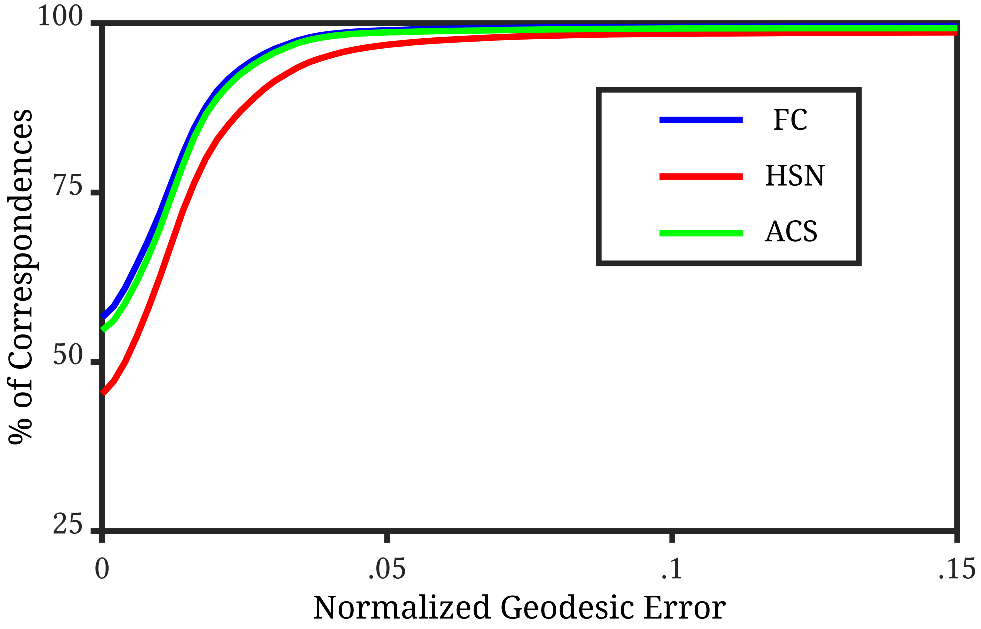


Figure 6-5. Percentage of correspondences for a given geodesic error on the remeshed FAUST dataset using field convolutions (FC), HSN, and ACSCNN (ACS).

after the ECHO block, taking the 32 features first to 256 channels, and then to the number of vertices on the template shape, with a $p = 0.5$ dropout layer in-between. Visualizations of some of the 32 channel features in our FCNet at the bottleneck before the dense final layers are shown in Figure 6-4.

Prior methods have typically used high-dimensional SHOT [TSDS10b] descriptors as inputs for this task, which we feel to be unnecessary due to the expressiveness of the field convolution framework. As such, we train HSN and ACSCNN [LLHL20] with raw 3D coordinates inputs for comparison – two recent methods which have reported state-of-the-art results in similar classification tasks. The results are shown in Figure 6-5, giving the percentage of total correspondences as a function of the normalized geodesic error. Our FCNet achieves the best performance, followed by ACSCNN.

Recent spectral-based networks, ACSCNN and DiffusionNet [SACO20], have significantly outperformed comparable equivariant networks in correspondence related tasks. This is likely for two reasons: 1) In contrast to the local patch-based

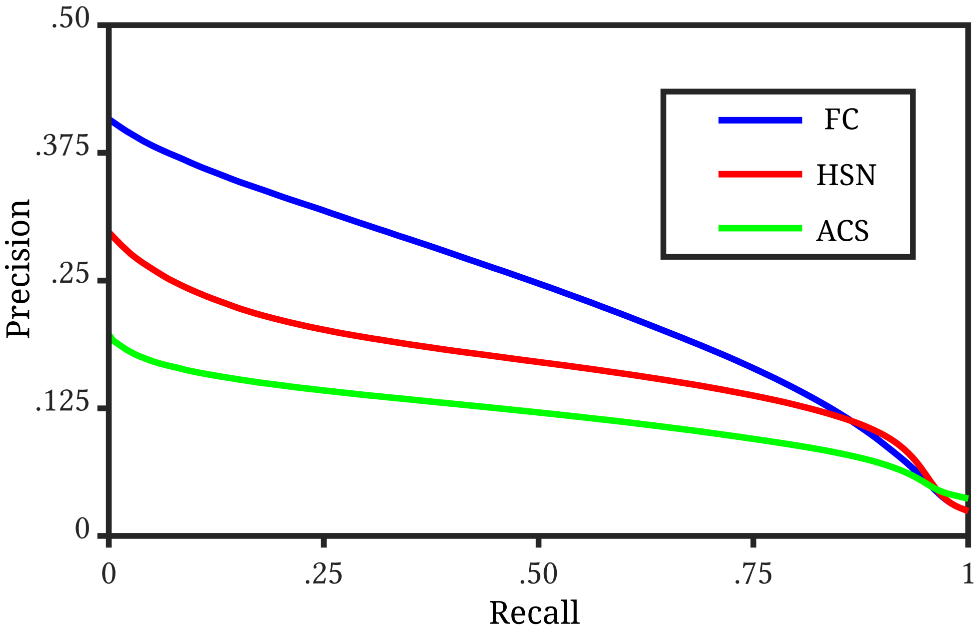


Figure 6-6. Results of feature matching evaluations on the SHREC 2019 Isometric and Non-Isometric Shape Correspondence dataset [DSL+19] in the form of the mean precision-recall curves.

convolution operators used in equivariant networks, spectral-based convolutions are formulated in a Laplace-Beltrami basis, providing an inherently global characterization of shape less sensitive to point-wise noise or local mislabeling; 2) The ability to essentially band-limit convolutions by working in basis of low-frequency eigenfunctions allows spectral-based networks to easily scale to high resolutions whereas equivariant networks must decrease both filter support and the number of parameters to process the same meshes. This makes the performance of our FC-Net particularly notable as it suggests that the network is able to overcome the relative limitations of equivariant frameworks in dense correspondence tasks specifically due to the robust construction of field convolutions as a scattering operation – keeping them stable despite smaller supports – and due to their descriptiveness, the latter of which does not diminish significantly even with fewer parameters.

6.6.5 Feature matching

Last, we train an FCNet to compute point-wise surface feature descriptors on shapes from the SHREC 2019 Isometric and Non-Isometric Shape Correspondence dataset [DSL+19]. The dataset consists of 50 meshes constructed from 3D scans of a jacketed humanoid figurine and a bare and gloved articulated wooden hand with 76 predefined pairs of meshes. We consider this dataset to be extremely challenging with significant non-isometric deformations and topological changes between pairs; as real-world scans the meshes also contain noise, varying triangulations, occluded geometry and various other sources of interference.

To ensure an even distribution of meshes in both the training and testing data, we group all pairs into three categories based on scan source (humanoid, hand, and gloved hand) and randomly select 20% of the pairs in each category to form the test split. Each pair in the SHREC 2019 Correspondence Dataset [DSL+19] consists of a *model* mesh V_M and a *scene* mesh V_S, with the dense ground-truth correspondence mapping the latter to the former. We randomly generate correspondences $C_{SM} = \{(s_i, m_i)\} \subset V_S \times V_M$ and non-correspondences $N_{SM} = (V_S \times V_M) \setminus C_{SM}$ by selecting 2048 points on both the model and the scene mesh using farthest point sampling, mapping the sampled scene points to the model mesh using the ground truth correspondence, and associating each mapped scene point to the geodesically nearest sampled point on the model.

We learn compact, 16-dimensional descriptors at each point using a twin network [LB13, MBBV15, SHG+20], where each mesh in a pair is processed by the same network and a twin loss function is minimized, weighting the descriptor distances between corresponding and non-corresponding points. In training, the objective of the network is to make the outputs for corresponding and non-corresponding pairs as similar and dissimilar as possible, respectively [LB13, MBBV15]. Specifically, for each pair of meshes in each epoch, we randomly subsample 512 pairs of corresponding and non-corresponding points, $P_{SM} = C_{SM}^{512} \cup N_{SM}^{512}$ and minimize the twin loss [SHG+20]

$$\begin{aligned} L\left(P_{SM}\right) = \sum_{(s,\,m)\in P_{SM}} & \alpha_{s,m} \|F_S(s) - F_M(m)\ \|^2 + \\ & \left(1 - \alpha_{s,m}\right) \max\left(0,\ 5 - \|F_S(s) - F_M(m)\ \|^2\right), \end{aligned} \tag{6.7}$$

where $\alpha_{s,m} = 1$ if $(s,\ m) \in C_{SM}$ or is set to a random variable between 0 and 0.2 otherwise. We compute precision-recall in the same manner as was done in the evaluation of the ECHO descriptors as in Equation (5.6), considering the set of sampled model points that are valid matches with a given point p to consist of those whose ground-truth correspondence lies within a geodesic ball of radius 0.05 about p. While this corresponds to a slightly more relaxed definition of correspondence, we find that all methods perform better maintaining a stricter notion of correspondence during training.

We train an FCNet consisting of eight FCResNet blocks $(N = 6,\ B = 1,\ \epsilon = 0.1)$ on the downsampled 2048-vertex mesh pairs, using the magnitudes of the output features as point-wise descriptors. HSN and ACSCNN are trained on the downsampled and full-resolution meshes, respectively. We report results averaged over three random samplings of the test-train split (Figure 6-6); to ensure fair comparisons, we compute the average precision-recall curves over all test pairs using the same set of correspondences for all methods. Our FCNet achieves the best performance by a significant margin, followed by HSN. The difference is likely explained by the increased descriptiveness of field convolution and its robust formulation as a scattering operation, making it better able to characterize flat, featureless areas (Figure 6-7, palm of the hand) and insensitive to high-frequency perturbations of the surface (Figure 6-7, folds in the figurine jacket), as compared to the gathering-based convolution operations used by HSN which rely on strongly constrained filters. We be-

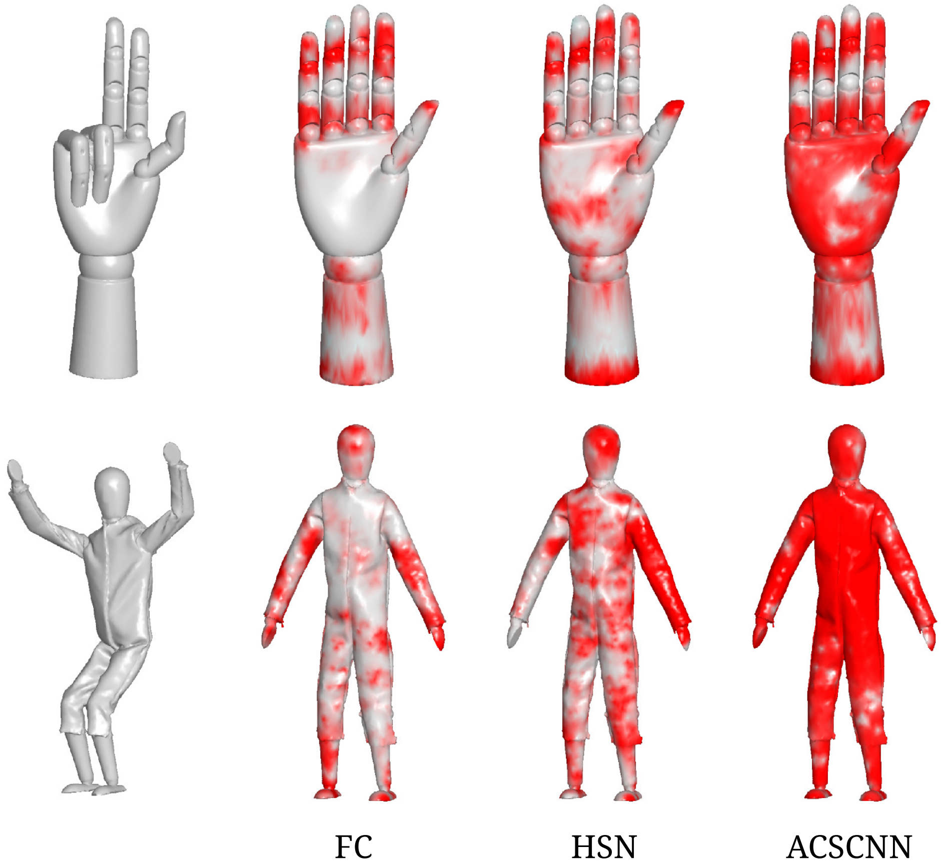


Figure 6-7. Feature-space distances: For each feature from the models on the right, we rank order the features on the model to the left by feature distance. Vertices are then colored from gray to red, showing how deeply one must traverse the rank ordered list before encountering a corresponding feature.

lieve that ACSCNN under-performs relative to the other methods because methods like ACSCNN which depend on the (global) spectral decomposition of the Laplace-Beltrami operator are less stable in the presence of non-isometric deformations, geometric occlusions, and changes in topology between corresponding pairs. While still not giving excellent performance, methods like FCNet and HSN, which use filters with local support, tend to be more robust.

6.6.6 Performance

Field convolutions are among the most efficient equivariant convolution operations, requiring few parameters per convolution operation. On an RTX 2080 GPU and 3.8

GHz CPU, our deepest FCNet trains at approximately 3 min/epoch on the full resolution meshes in the dense correspondence task. Field convolutions use a similar number of parameters as HSN [WEH20] per convolution and with half the memory footprint. HSN's multi-stream convolutions learn a weight matrix for the radial profile and rotational offset corresponding to each stream and the connections between them, resulting in $(N + 1)\, S^2$ total parameters per convolution, with N the number of radial samples and S the number of streams. Similarly, we learn a complex radial profile and rotational offset for each non-negative frequency up to the number of band-limited frequencies with $N(2B+1)+B+1$ total parameters per convolution. In the classification and segmentation experiments, HSN reports results using $N = 6$ radial samples and $M = 2$ streams resulting in 28 total parameters per convolution. In the same experiments, our FCNet achieves state-of-the-art performance with 33 total parameters per convolution, as we use filters with band-limit $B = 2$ and the same number of radial bins. However, HSN stores features for both streams, increasing spatial complexity by a factor of two.

More generally, we see our state-of-the-art results on the segmentation task using the composite dataset [MGA+17] as particularly notable in that other top-performing methods, including MeshCNN [HHF+19] and HSN, use the deepest versions of their network for this task despite the small number of labels involved (eight classes), presumably because of the large size of the training dataset. Our FCNet outperforms these networks with a much shallower architecture and only in the dense correspondence and feature matching tasks – both of which involve learning granular distinctions between large numbers of similar points – do we increase the depth of our network. This suggests that unlike most networks, the depth of an FCNet (or other network built on field convolutions) necessary to achieve good performance is not strongly dependent on the size of the dataset, and scales primarily with task complexity.

6.7 Conclusion

Field convolution offers rich notion surface convolution on vector fields, combining invariant spatial weighting with the parallel transport of features in a scattering operation *while placing no constraints on the filters themselves*. This formulation is highly descriptive, insensitive to a variety of nuisance factors, and straight-forward to implement; with it, we construct simple networks that achieve state-of-the-art results in fundamental geometry-processing tasks.

While the complexity of our method is comparable to existing equivariant approaches, it shares the same drawbacks as filter supports and parameter counts must be limited to process meshes at full resolution. More generally, existing successful surface learning frameworks (including ours) are designed to handle only isometric or nearly-isometric shape deformations and fail to achieve adequate performance in the presence of the kinds of complex deformations, geometric occlusions, and topological changes found in real shape data.

Chapter 7

Möbius-Equivariant Spherical CNNs

7.1 Introduction

Convolutional neural networks (**CNNs**) are effective because convolution responds to a contextualized window on the signal, forcing the learning to be *translation-equivariant*. However, vanilla CNNs assume a fixed orientation for the coordinate frame and lose effectiveness in the presence of deformations that change the frame. This has lead to the development of more general notions of convolution equivariant to transformation groups including rotations [CW16, WGTB17] and dilations [WW19, SSS19a, FSIW20]. Critically, the notion of rotation-equivariance has facilitated the generalization of CNNs to domains without a canonical orientation at each point such as the sphere [CGKW18, CWKW19b, EMD20] and arbitrary surfaces [dHWCW20, WEH20, MKK21]. The resulting networks are *isometry-equivariant* – able to repeatably characterize local features in the presence of distance-preserving transformations – and have excelled in fundamental geometry processing tasks such as shape classification, segmentation, and correspondence.

Despite their success, rotation- and isometry-equivariant CNNs can fail to achieve adequate performance in the presence of the kinds of complex deformations commonly found in real-world image and shape data [MKK21]. Such deformations may potentially be better modeled by higher-dimensional transformation groups. For example, homographies (projective transformations) better approximate changes in camera viewpoints than similarities (rotations and dilations) [HZ03] and, for spherical images, can be represented using conformal transformations [EMSJB14, SS16]. For geometry processing, conformal (angle-preserving) transformations encompass a broader class of deformations than isometries that still preserve the sense of 'shape' [LPRM02, GWC$^+$04, CPS11].

Using Möbius-equivariant extended convolution as defined in §3.3.2 – which we refer to as *Möbius convolution* (**MC**) – we develop the foundations for *Möbius-equivariant* spherical CNNs. To facilitate efficient evaluations, we parameterize filters using log-polar basis functions from which we derive an approximation of the action of the frames, allowing us to compute our convolutions via the Fast Spherical Harmonic Transform [DH94, KR08]. Our framework is flexible, and we demonstrate the utility of our Möbius-equivariant CNNs by achieving promising results on standard benchmarks in both genus-zero shape classification and spherical image segmentation.

7.2 Related work

Existing group-equivariant CNNs can be broadly categorized based on whether convolution is integrated over the group itself or the domain on which it acts. CNNs based on the former approach were first introduced by [CW16, CGW19], where kernels are parameterized in terms of equivariant basis functions *on the group itself* and convolution is performed by lifting features from the domain and searching over all possible transformations of the features or kernels. This approach is highly effective when considering the action of discrete groups on features sampled on a regular lattice, and has since been extended to handle the continuous group of rotations in both two and three dimensions [CGKW18, LW21]. However, this approach isn't readily generalizable – either theoretically or computationally – to higher-dimensional or non-compact groups, where there are more parameters to integrate over, the domains of integration are unbounded, and the representations are infinite-dimensional.

Recent work by [FSIW20] largely sidesteps these problems by considering only the origin-preserving subgroups and integrating with respect to an equivariant Monte Carlo estimator to facilitate evaluations. However, this approach assumes that the group acts linearly on the domain and requires the exponential mapping from the infinitesimal generators to the group to be surjective, precluding its generalization to groups that act projectively.

Equivariant CNNs that integrate over the domain on which the group acts can trace their lineage to earlier work on steerable filters [FA91, SF96, THO99] – kernels are parameterized in terms of equivariant basis functions *on the domain* that rotate or dilate with the local coordinate system [WGTB17, WC19, WW19, SSS19a]. This approach has been extended to volumetric domains [WC19] and point clouds [QSMG17], and has facilitated the development of isometry-equivariant CNNs on domains without canonical coordinate systems such as the sphere [CGKW18, EMD20] and arbitrary 2D surfaces [WEH20, dHWCW20, MKK21]. Unfortunately, finite-dimensional equivariant bases often don't exist for non-commutative and non-compact transformation groups of interest, limiting the practical scope of these approaches.

Recently, *fully-connected* networks have been developed that achieve equivariance to non-compact transformation groups including the Lorentz group [BAO$^+$20], the Poincaré group [VHSF$^+$21], and the symplectic group [FWW21] – all closely related to Möbius transformations. However, we are not aware of existing *convolutional* neural networks that are equivariant to either projective transformations or Möbius transformations, and believe our approach to be the first in both regards.

7.3 Method overview

Möbius convolutions provide a method for Möbius-equivariant spatial aggregation on the sphere. We facilitate an efficient discretization by parameterizing filters using log-polar basis functions from which we derive a linearized approximation of the action of the frames, allowing us to compute Möbius convolutions via the fast spherical harmonic transform [DH94, KR08].

We complete the foundations for Möbius-equivariant CNNs by introducing a conformally-equivariant normalization layer based on filter response normalization [SK20] and we validate equivariance by direct experimental evaluation. The principle module in applications is a simple Möbius convolution ResNet (**MCResNet**) block [HZRS16], which is self-contained and flexible. We demonstrate the utility of our framework by achieving promising results on standard benchmarks in both genus-zero shape classification and spherical image segmentation.

7.4 Discretization

Recall from §3.3.2 that an $\mathrm{SL}(2,\mathbb{C})$-equivariant (*i.e.* Möbius-equivariant) extended convolution of a function ψ with a filter f, both in $L^2(\widehat{\mathbb{C}},\mathbb{C})$, can be defined as:

$$(\psi * f)(y) \overset{(3.28)}{=} \int_{\widehat{\mathbb{C}}} \rho_\psi(z)\,\left[\mathfrak{T}_\psi(z)\,f\right]\left(\mathrm{Log}_z\,y\right)\,dz, \tag{7.1}$$

with the generalized logarithm

$$\mathrm{Log}_z \overset{(3.27)}{\equiv} \frac{1}{|c|\sqrt{1+|z|^2}}\begin{bmatrix} c & -cz \\ \overline{c}\overline{z} & \overline{c} \end{bmatrix} \in \mathrm{SU}(2) \subset \mathrm{SL}(2,\mathbb{C}),$$

and the frame operator $\mathfrak{T} : L^2(\widehat{\mathbb{C}}, \mathbb{C}) \to \mathrm{L}$ and the density operator $\rho : L^2(\widehat{\mathbb{C}}, \mathbb{C}) \to L^2(\widehat{\mathbb{C}}, \mathbb{C})$ given by

$$\begin{aligned} \mathfrak{T}_\psi(x) &\overset{(3.29)}{\equiv} \begin{bmatrix} \left[d\,\mathrm{Log}_x \psi\big|_0\right]^{-\frac{1}{2}} & 0 \\ \frac{1}{2}\left[\nabla d\,\mathrm{Log}_x \psi\big|_0\right]\left[d\,\mathrm{Log}_x \psi\big|_0\right]^{-\frac{3}{2}} & \left[d\,\mathrm{Log}_x \psi\big|_0\right]^{\frac{1}{2}} \end{bmatrix} \in \mathrm{L} \subset \mathrm{SL}(2, \mathbb{C}) \\ \rho_\psi(x) &\overset{(3.32)}{\equiv} \left| d\,\mathrm{Log}_x \psi\big|_0 \right|^2. \end{aligned} \tag{7.2}$$

We call this convolutional operator *Möbius convolution*, and to compute it at the scale necessary to build CNNs, we develop an implementation based on the fast Spherical Harmonic Transform [DH94, KR08]. We give an outline of this process below and leave the details to Appendix I.

Identity convolution with the Spherical Harmonic Transform

To simplify the calculation, we first consider a simpler non-equivariant convolution, we call an *identity convolution*, where we replace the frame and density operators from Equation (7.2) with the trivial frame field $\mathfrak{T}_\psi(z) = e$ (with e the identity) and the density $\rho_\psi(z) = \psi(z)$,

$$(\psi *_e f)(y) = \int_{\widehat{\mathbb{C}}} \psi(z)\, f\left(\mathrm{Log}_z y\right) dz, \tag{7.3}$$

using $*_e$ to distinguish it from the equivariant convolution.

Assuming that ψ and f are band-limited functions, they can be expressed in terms of their spherical harmonic decompositions as

$$\psi = \sum_{l=0}^{B-1} \sum_{m=-l}^{l} \boldsymbol{\psi}_{lm} Y_l^m \quad \text{and} \quad f = \sum_{l'=0}^{B-1} \sum_{m'=-l'}^{l'} \mathbf{f}_{l'm'} Y_{l'}^{m'}. \tag{7.4}$$

with B the band-width. Recalling that Log_z is a rotation, we know that it preserves the frequency content and expand

$$f \circ \mathrm{Log}_z = \sum_{l'=0}^{B-1} \sum_{|m'| \le l'} \sum_{|m''| \le l'} \mathbf{f}_{l'm'} D^{l'}_{-m'm''}(\mathrm{Log}_z) Y_{l'}^{m''}$$

where $D^{l'}_{-m'm''}$ is the Wigner-D function giving the (l', m'')-th spherical harmonic coefficient of the rotation of $Y^{m'}_{l'}$ [CK16]. Furthermore, using the fact that in the $z-y-z$ Euler angle notation our definition of Log_z corresponds to a rotation described by an Euler triplet whose first entry is zero, it follows that the integral vanishes

$$0 = \int_{\widehat{\mathbb{C}}} Y^m_l(z) \cdot D^{l'}_{-m'm''}\left(\mathrm{Log}_z\right)\, dz$$

whenever $m'' \neq m$. Thus, expanding Equation (7.3) we get

$$(\psi *_e f)(y) = \sum_{l'=0}^{B-1} \sum_{m''=-l'}^{l'} \left[\sum_{l=|m''|}^{B-1} \boldsymbol{\psi}_{lm''} \left(\sum_{m'=-l'}^{l'} \mathbf{f}_{l'm'} \Delta^{m'm''}_{ll'} \right) \right] Y^{m''}_{l'} \tag{7.5}$$

where the value of $\Delta^{m'm''}_{ll'}$ is independent of ψ and f,

$$\Delta^{m'm''}_{ll'} = \int_{\widehat{\mathbb{C}}} Y^{m''}_l(z) \cdot D^{l'}_{-m'm''}\left(\mathrm{Log}_z\right)\, dz. \tag{7.6}$$

From this, we can compute identity convolutions efficiently by: 1). Computing the spherical harmonic coefficients of ψ and f using the fast SHT; 2). Summing the two sets of coefficients according to Equation (7.5) to get the coefficients of the convolution; and 3). Applying the fast inverse SHT to reconstruct the convolution.

The complexity of steps 1 and 3 are proportional to those of the fast SHT, which is $O(B^2 \log^2 B)$. However, in our implementation we compute the discrete Legendre transform via sparse matrix multiplication for a total complexity of proportional to $O(B^3 \log B)$, which we find to be more efficient on the GPU. Step 2 has complexity $O(B^4)$, and we show that this computation can be re-used in computing the full (equivariant) convolution.

Spherical log-polar bases

To efficiently compute arbitrary Möbius convolutions as in Equation (7.1), we approximate them as sums of identity convolutions. To do so, we choose a basis for our filters that enables an approximation of the action of $\mathfrak{T}_\psi$.

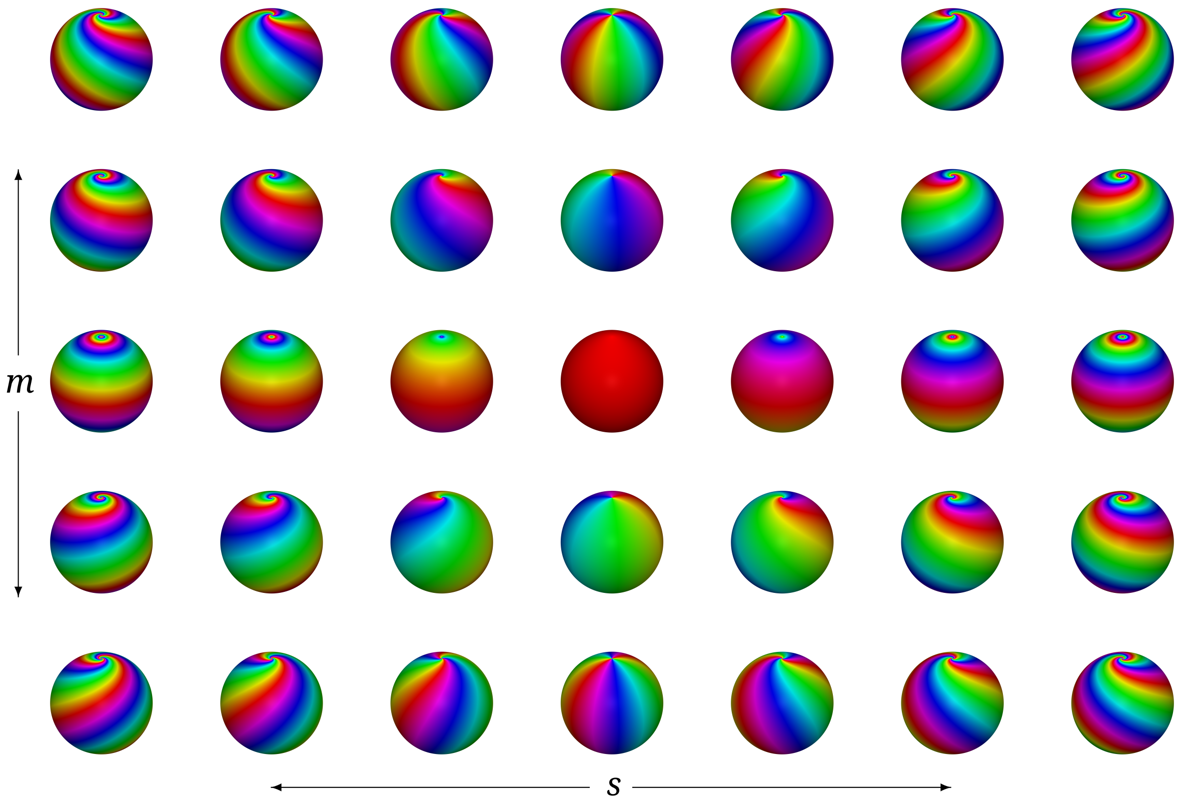


Figure 7-1. Spherical log-polar functions $\mathbf{B}^t_{ms}$ at integer exponents with $t = 0.15$.

To this end we use linear combinations of log-polar (Fourier-Mellin) basis functions [Vil78, VK91],

$$\mathbf{B}^t_{ms}(z) \equiv \frac{|z|^{is}}{|z|^t}\left(\frac{z}{|z|}\right)^m \tag{7.7}$$

with $m \in \mathbb{Z}$ and $s, t \in \mathbb{R}$. These function are localized about $z = 0$ (resp. $z = \infty$) when $t > 0$ (resp. $t < 0$), are discontinuous at $z \in \{0, \infty\}$ when $m \neq 0$ (since the argument of z is not defined) and singular at $z = 0$ (resp. $z = \infty$) when $t > 0$ (resp. $t < 0$). We note that, for the purposes of integration, the singularity can be ignored when $t < 1$. Loosely, this follows from the fact that $\int \frac{1}{x^t} = \frac{1}{1-t}x^{1-t} + c$ which is bounded at $x = 0$ whenever $t \in (0, 1)$. Noting also that $\mathbf{B}^t_{ms}(z) = \mathbf{B}^{-t}_{-m-s}(1/z)$, it follows that the functions are continuous away from $\{0, \infty\}$ and, for the purposes of integration, singularities at $\{0, \infty\}$ can be ignored when $|t| < 1$.

The first several basis functions at integer frequencies $|m| \leq 2$ and $|s| \leq 3$, with

$t = 0.15$ are shown in Figure 7-1. The complex-valued functions are visualized using the HSV scale: hue and value are determined by the arguments and magnitudes of the function values, and saturation is fixed at one.

Using these as basis functions, we consider filters in the span

$$f = \sum_{m=-M}^{M} \sum_{s=-N}^{N} b_{ms} \, \mathbf{B}_{ms}^{t} \tag{7.8}$$

where s is constrained to be an integer and t is positive so as to localize the filter about the origin. In practice we use real-valued filters ($f \in L^2(\widehat{\mathbb{C}}, \mathbb{R})$), so $\overline{b_{-m-n}} = b_{mn}$, giving $(2M+1)(2N+1)$ real parameters per filter.

Approximating the transformation of filters

Unfortunately, it is not the case that the space of band-limited filters spanned by the $\mathbf{B}_{ms}^{t}$ is fixed under the action of the lower-triangular subgroup. This is because, in general, for non-compact, non-commutative groups, a space of functions fixed under the action of the group (i.e. a representation) will be infinite-dimensional.

However, we show in the Appendix I that, given a filter as in Equation (7.8) and given a lower-triangular matrix $L \in \mathrm{L}$, the transformation of the filter f by L can be expanded as

$$L\, f = \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{j=1}^{3} {}_{j}\boldsymbol{\zeta}_{ms}^{t\sigma_j}(L, \mathbf{b})\, \mathbf{B}_{ms}^{\sigma_j}\, ds, \tag{7.9}$$

where $\mathbf{b}$ is the $(2M+1) \times (2N+1)$-dimensional vector of coefficients of f, m is now summed over all integers, s is continuous and integrated over the real line, σ_1, σ_2, and σ_3 (the localization values) are any real values satisfying $t < \sigma_1 < 2$, $t - 1 < \sigma_2 < 0$, and $\sigma_3 = t$, and ${}_{j}\boldsymbol{\zeta}_{ms}^{t\sigma_j}$ are functions taking a lower-triangular matrix and a set of filter coefficients, and returning the coefficient of $\mathbf{B}_{ms}^{\sigma_j}$ in the expansion of the transformed filter. Here, the integral is equivalent to an inverse Mellin transform with frequency variable s, and the bounds on σ_1 and σ_2 are necessary to ensure invertibility [Vil78, VK91].

Obviously, the infinite summation and the integration in Equation (7.9) make evaluation unfeasible. We propose a practical implementation by truncating the summation over the angular frequency m, and replacing the integration over the real line with a discrete approximation using quadrature. The summation is a result of the addition theorem for Bessel functions [Wat95] which appear in the derivation of ${}_j\boldsymbol{\zeta}^{t\sigma_j}_{ms}$; it converges rapidly at low frequencies and can be well-approximated with only several terms [CK16]. The use of quadrature is motivated by the observation that for a fixed transformation L and filter coefficients $\mathbf{b}$ the function ${}_j\boldsymbol{\zeta}^{t\sigma_j}_{ms}$ tends to be smooth and falls off quickly away from $s = 0$. Using the approximation, we get

$$L\, f \approx \sum_{m=-M'}^{M'} \sum_{q=1}^{Q} \sum_{j=1}^{3} w_q \;{}_j\boldsymbol{\zeta}^{t\sigma_j}_{ms_q}(L, \mathbf{b})\, \mathbf{B}^{\sigma_j}_{ms_q} \tag{7.10}$$

where $\{s_q\} \subset \mathbb{R}$ are the quadrature points and $\{w_q\}$ are the associated quadrature weights.

We remark that the principle idea behind the expansion in Equation (7.9) involves exploiting the symmetry of the spherical log-polar basis functions under Möbius transformations taking z to $-z^{-1}$. This allows us to replace the projective action of L with the affine action of the upper-triangular matrices – the group of rotations, translations, and dilations – whose representations are better understood [Vil78, VK91].

Efficient Möbius Convolutions

Plugging the approximation in Equation (7.10) into the definition of Möbius convolution in Equation (7.1) and moving the sums outside the integral gives

$$(\psi * f) \approx \sum_{\substack{-M' \leq m \leq M' \\ 1 \leq q \leq Q \\ 1 \leq j \leq 3}} \left(\rho_\psi \, w_q \; {}_j\boldsymbol{\zeta}^{t\sigma_j}_{m s_q}\left(\mathfrak{T}_\psi, \mathbf{b}\right) \; *_e \; \mathbf{B}^{\sigma_j}_{m s_q} \right). \tag{7.11}$$

Thus, by approximating the pointwise action of the frame operator $\mathfrak{T}_\psi(z)$, we can approximate an arbitrary Möbius convolution as a sum of identity convolutions. The components of ${}_j\boldsymbol{\zeta}^{t\sigma_j}_{m s_q}$ depending only on s can be pre-computed for a fixed set of quadrature points so that, in practice, the complexity of evaluating the function at run time is linear in the coefficients of $\mathbf{b}$ and the sums over m, q, and j.

7.4.1 Complexity

In the approximation of Möbius convolution in Equation (7.11), the right side of the identity convolutions is independent of the filter coefficients, so the innermost bracketed term in Equation (7.5) can be pre-computed for a given band-limit B (for every angular frequency m, quadrature point s_q, and localization index σ_j). Thus, the $O(B^4)$ computational bottle-neck in computing the identity convolution need only be performed once and the total complexity of computing the Möbius convolution is $O(M'QB^3 \log B)$. In applications, we find that setting $M' = M + 1$ and using a $Q = 30$ point trapezoidal quadrature rule in Equation (7.10) allows us to both suitably approximate the transformation of the filter and scale up to $B = 64$.

7.5 Möbius-Equivariant Spherical CNNs

Möbius convolutions provide a flexible framework for Möbius-equivariant spatial aggregations on the sphere. With them, constructing Möbius-equivariant spherical

CNNs is straight-forward and requires no specialized architecture. The atomic units are the same as those found in regular CNNs – a convolutional layer, followed by normalization and a non-linearity.

7.5.1 Convolutional layers

For a Möbius-convolution layer mapping C-channel input features $\psi \in L^2(\widehat{\mathbb{C}}, \mathbb{R}^C)$ to C'-channel output features $\psi' \in L^2(\widehat{\mathbb{C}}, \mathbb{R}^{C'})$, the $c'-$th output feature $\psi'_{c'}$ is computed in the usual manner by summing the convolutions of the input features with the filters in the $c-$th row of the bank. However, the structure of Equation (7.11) allows us to preform the reduction over the input channels *before* computing the convolutions in the sum, such that

$$\psi'_{c'} = \sum_{\substack{-M' \leq m \leq M' \\ 1 \leq q \leq Q \\ 1 \leq j \leq 3}} \left(\sum_{c=1}^{C} \rho_{\psi_c} \, w_q \; {}_j\boldsymbol{\zeta}^{t\sigma_j}_{ms_q}\left(\mathfrak{I}_{\psi_c}, \mathbf{b}^{cc'}\right) *_e \mathbf{B}^{\sigma_j}_{ms_q} \right), \tag{7.12}$$

where $\mathbf{b}^{cc'}$ denotes the $(2M+1)(2N+1)$ parameters for the (c, c')-th filter in the bank. Thus, for each convolutional layer mapping C input features to C' output features, we only need to compute C' Möbius convolutions instead of $C \times C'$.

This advantage is not without caveat. A naive implementation of the inner sum over the input channels produces large intermediate tensors at high resolutions ($B \geq 64$), which can quickly fill GPU memory. Our layers are implemented in PyTorch [PGM+19], where we fuse this operation to reduce its overhead.

7.5.2 Normalization and Non-linearities

Standard normalization techniques don't commute with Möbius transformations, since the mean and standard deviation of spherical signals are not invariant to dilation. Instead, we introduce a conformally-equivariant normalization layer based on filter response normalization [SK20], replacing the square mean with the Dirichlet energy which *is* invariant under Möbius transformations. The normalization is applied on a per-channel basis independent of the batch size via the mapping

$$\psi_c \mapsto \frac{\alpha_c \, \psi_c}{\sqrt{\int_{\widehat{\mathbb{C}}} \rho_\psi(z) \, dz + \epsilon_c}} + \beta_c, \tag{7.13}$$

where ρ_ψ is defined as in Equation (7.2) and $\alpha_c, \ \beta_c \in \mathbb{R}$ and $\epsilon_c \in \mathbb{R}_{>0}$ are learnable per-channel parameters.

Following normalization we apply thresholded activations as non-linearities, which have been shown to better compliment filter response normalization than other activation layers [SK20]. Here, we replace the ReLU with the Mish activation [Mis19] which we find improves training speed and performance. Specifically, non-linearities are applied pointwise as,

$$\psi_c \mapsto \mathrm{Mish}\big(\psi_c - \gamma_c\big) + \gamma_c, \tag{7.14}$$

where $\gamma_c \in \mathbb{R}$ is a learnable per-channel threshold value. We note that the thresholded activation is not fundamental to our framework, and can be replaced with other activation layers if desired.

7.6 Evaluation

We validate our claim of Möbius-equivariance empirically in an ablation study and demonstrate the utility of Möbius-equivariant CNNs by achieving strong results in both geometry and spherical-image processing tasks. In the former paradigm, we apply our framework to the task of genus-zero shape classification by conformally mapping surfaces to the sphere; in the latter, we consider the task of omnidirectional image segmentation.

Our principle module in applications is an MCResNet block [HZRS16], consisting

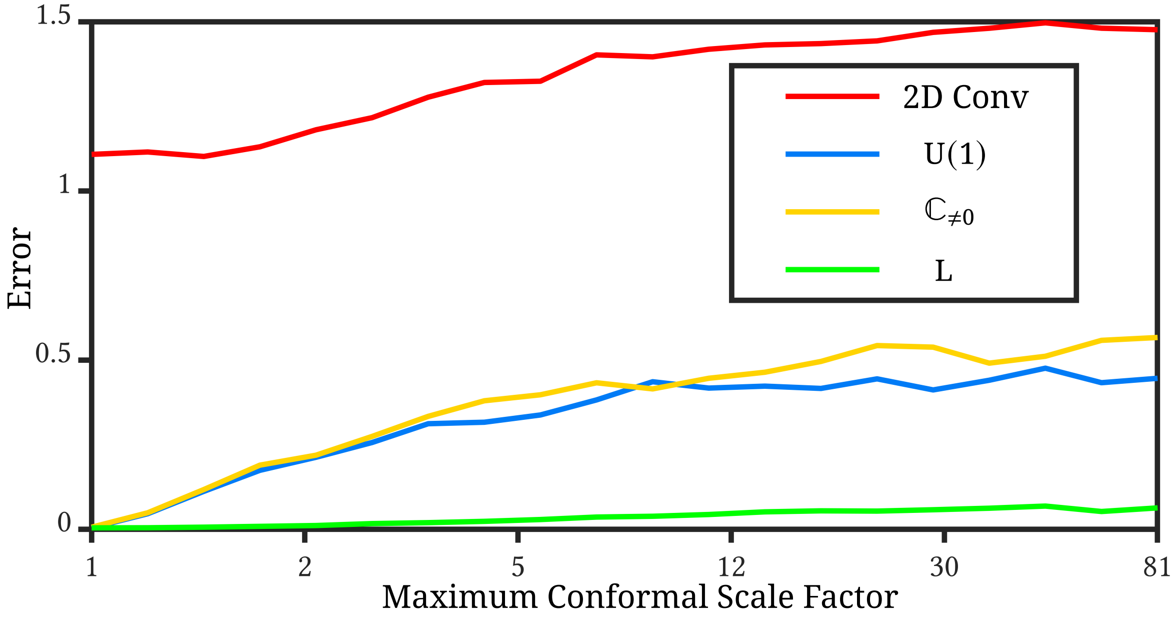


Figure 7-2. The equivariance error plotted as a function of the maximum conformal scale factor. Notably, moving from U(1) (rotations) to $\mathbb{C}_{\neq 0}$ (rotations and dilations) does not provide a benefit – one must consider the full lower-triangular subgroup $\mathrm{L} \subset \mathrm{SL}(2, \mathbb{C})$

of two Möbius convolutions, each followed by the normalization layer and pointwise non-linearity described in Equations (7.13-7.14), with a residual connection between the input and output streams. We use a band-limited space of filters with $M = N = 1$ and set $t = 0.15$, $\sigma_1 = 0.35$, $\sigma_2 = -0.15$. We fit our networks using SGD with Nesterov momentum [SMDH13], training for 60 epochs with an initial learning rate of 10^{-2}, decaying to 10^{-4} on a cosine annealing schedule [LH17]. Our framework is implemented in PyTorch [PGM+19].

7.6.1 Equivariance

We empirically validate the equivariance of our framework by quantifying the degree to which our layers commute with Möbius transformations of increasing area distortion. We consider a 32-channel, $B = 64$ band-limited MCResNet block with the equivariant residual connection removed to avoid bias. We control the area distortion of a Möbius transformation g by composing a series of random rotations and inversions so that the maximal scale factor over $\widehat{\mathbb{C}}$ equals a fixed value [BCK18]. Denoting $\mathcal{R}$ as the mapping induced by passing features through the MCResNet layer, we follow [dHWCW20, WW19, SSS19a] and define the equivariance error for a fixed maximum scale factor $\alpha \in \mathbb{R}_{\geq 0}$ as

$$\text{Error} = \frac{\mathrm{E}\,(\,\mathcal{R}(\,g\psi\,) - g\,\mathcal{R}(\,\psi\,))^2}{\mathrm{Var}\; g\,\mathcal{R}(\,\psi\,)} \quad \text{with} \quad \max_{z \in \widehat{\mathbb{C}}} \lambda_g^2(z) = \alpha, \tag{7.15}$$

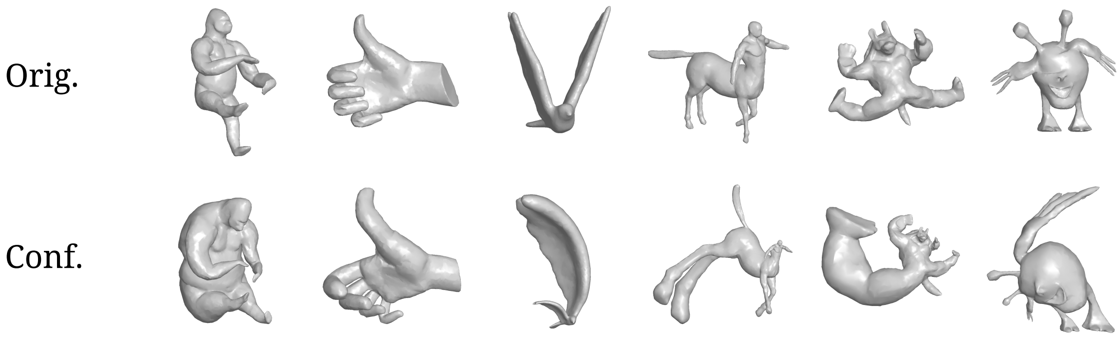
where E and Var denote the mean and variance computed over 100 randomly initialized models, Möbius transformations, and features.

As a baseline, we compare our proposed approach against three other paradigms. In the first, we replace Möbius convolution with a standard 5 × 5 convolution layer taking ρ_ψ as input; in the second, we restrict the transformation field to rotations so that $\mathfrak{T}_\psi(z) \in \mathrm{U}(1)$; in the third, we loosen the restriction to include dilations with $\mathfrak{T}_\psi(z) \in \mathbb{C}_{\neq 0}$. We note that the second and third paradigms are isometry-equivariant, and that the latter is also equivariant to the conformal transformations of the (non-compactified) plane.

The results are shown in Figure 7-2, where the equivariance error in Equation (7.15) is plotted as a function of the maximum scale factor. The green curve is our proposed method with $\mathfrak{T}_\psi(z) \in \mathrm{L}$. Using our method, the error stays very low, indicating that Möbius convolution is approximately equivariant even in the presence of significant changes in scale ($\lambda_g^2(z) \geq 12$). Notably, we see no improvement moving from $\mathfrak{T}_\psi(z) \in \mathrm{U}(1)$ to $\mathfrak{T}_\psi(z) \in \mathbb{C}_{\neq 0}$, suggesting that rotations and dilations alone fail to well-characterize the local deformations induced by Möbius transformations.

7.6.2 Conformal Shape Classification

Next, we use Möbius convolutions to classify genus-zero shapes. The SHREC '11 dataset [LGB+11] has become a popular choice for evaluating network performance in shape classification tasks and several recent approaches have achieved near-



Method	Original Acc.	Conformal Acc.
MC (ours)	99.1%	86.5%
DiffusionNet [SACO20]	99.5%	64.9%
FC [MKK21]	99.2%	40.7%

Table 7-I. Genus-zero shape classification. Several conformally deformed meshes from the SHREC '11 dataset [LGB+11] are shown above.

perfect accuracy on the dataset [WNEH21, MKK21, MLR+20, SACO20]. However, shapes within each of the 30 categories in the dataset differ only by (approximately) isometric deformations. To better highlight the strengths of our approach, we extend the dataset to include deformations given by random conformal transformations with several examples shown above Table 7-I.

To apply our framework, we conformally map each mesh to the sphere via mean curvature flow [KSBC12] and use a simple network consisting of a single 16-channel, $B = 64$ band-limited MCResNet block followed by a global mean pool and a fully-connected layer to give predictions for the 30 shape categories.

We fit our network on both the original SHREC '11 dataset and our conformally-augmented version using 10 samples per class. For comparisons, we report the results of Field Convolutions (FC) [MKK21] and DiffusionNet [SACO20] – two state-of-the-art surface networks – on the original dataset and train both networks on the conformally-augmented version. As inputs, each network takes the Heat Kernel Signature (HKS) [SOG09] computed at 16 different timescales; since the HKS isn't

Method	Accuracy	IoU	
MC (ours)	60.9%	43.3%	Spectral
SWSCNN [EMD20]	58.7%	43.4%	
SphCNN [EABMD18]	52.8%	40.2%	
CubeNet [SR+21]	62.5%	45.0%	Spatial
HexNet [ZLSC19]	58.6%	43.3%	
UGSCNN [JHK+18]	54.7%	38.3%	

Table 7-II. Omni-directional image segmentation

conformally-invariant, we use the values computed on the original meshes when training and testing on their conformally-augmented counterparts.

Results are shown in Table 7-I in the form of the mean classification accuracy over three randomly sampled test-train splits. Our simple Möbius convolution network matches the state-of-the-art performance of FC and DiffusionNet on the original dataset and significantly outperforms both on the conformally augmented version, despite the fact that the transformations between the spherically-parameterized meshes aren't perfect Möbius transformations. Like FC and DiffusionNet, our method is equivariant to isometric deformations of the meshes as they manifest as Möbius transformations after parameterization to the sphere, which serves to explain our strong performance on the original dataset. However, in addition, Möbius-equivariance allows our rudimentary network to better account for conformal deformations between similar shapes and suggests that a new class of *conformally-equivariant* surface networks may outperform existing *isometry-equivariant* networks in challenging shape analysis and recognition tasks.

7.6.3 Omni-directional Image Segmentation

Last, we demonstrate the utility of Möbius convolutions by moving from geometry to image processing, where we apply them to semantically segment omni-directional images from the Stanford 2D3DS dataset [ASZS17]. Here, we use MCResNet blocks to construct a U-Net [RFB15] architecture with 32, 64, 128, 256, 128, 64, 32 channels per layer, applying max pooling or nearest neighboring upsampling before each increase or decrease in channel width. As with other state-of-the-art equivariant spherical networks [EMD20, SR+21], we find our method performs best as a feature extractor for a small network of standard convolutional layers due to the consistent latitudinal orientation of the images; we append six 3 × 3 2D convolutions to the end of our network to predict labels. To measure performance, we report the mean per-class segmentation accuracy and intersection over union (IoU) averaged over three official folds.

Results are shown in Table 7-II, and we attain performance comparable to the state-of-the-art. Existing successful spherical networks compute convolutions either in the spatial domain [SR+21, ZLSC19, JHK+18] or, like our method, in the spectral domain via expansions in spherical basis functions. In the latter case, efficiency and scalable filter support comes at the cost of fidelity, as some degree of high-frequency information is lost when computing the forward SHT due to the fixed band-limit assumption. This puts spectral methods at a disadvantage in precision labeling tasks like segmentation, where spectral aliasing can blur sharp boundaries and over-smooth the kinds of hyper-localized features necessary to make accurate predictions in the presence of a large class imbalance. Compounded by the devaluation of equivariance due to the consistent orientation of the images, this is a challenging task for our framework. However, we outperform existing rotation-equivariant spectral approaches, demonstrating that we are able to achieve equivariance to a more complex group of transformations without sacrificing descriptiveness.

7.7 Conclusion

Using Möbius-equivariant extended convolution, we develop the foundations for Möbius-equivariant spherical CNNs and demonstrate the utility of this framework by achieving strong results in both geometry and spherical-image processing tasks. More generally, this work represents an effort to move both image and surface convolutional neural networks beyond standard rotation- and isometry-equivariance and into the realm of conformal-equivariance. In particular, our experiments suggest that the latter transition may be especially relevant in the context of shape analysis and recognition and we hope this work serves to catalyze the development of a new generation of surface networks better able to handle the kinds of complex deformations found in real-world shape data.

Chapter 8

Conclusion

This thesis presents extended convolution – a unified framework for transformation-equivariant convolutions on arbitrary homogeneous spaces and 2D Riemannian manifolds. Extended convolution is based on a key observation: to achieve equivariance to a given group of transformations we only need to consider the stabilizer subgroup which deforms the positions of points as seen in the frames of their neighbors. By defining an equivariant frame operator at each point with which we align the filter, we correct for the change in the relative positions induced by subgroup. To compute convolutions, input features are mapped to a density distribution, controlling for the change in area measure, and integrated against the aligned filters over homogeneous space, rather than the group itself.

Extended convolution is highly flexible and descriptive - the construction places no constraints on the kinds of filters that can be used. Furthermore, the framework can handle arbitrary transformation groups, including higher-dimensional non-compact groups that act non-linearly on the domain, such as Möbius transformations of the sphere. Critically, extended convolution naturally generalizes to arbitrary 2D Riemannian manifolds – such as the surfaces of 3D shapes – as it does not need a canonical coordinate system to be applied.

The power and utility of the extended convolution framework is demonstrated in several applications. A unified family ECHO (Extended Convolution Histogram of Orientations) of local image and surface descriptors is proposed, constructed by rasterizing the filter maximizing the response of the extended convolution at a keypoint. The ECHO image descriptor matches the performance of SIFT on a challenging, large-scale image dataset and 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.

Field convolution generalizes extended convolution to an operator on surface vector fields, offing a rich notion of convolution that combines invariant spatial weighting with the parallel transport of features in a scattering operation while placing no constraints on the filters themselves. Field convolution is isometry-equivariant, highly descriptive, insensitive to a variety of nuisance factors, and straight-forward to implement; with it, we construct simple networks that achieve state-of-the-art results in fundamental geometry-processing tasks including shape classification, segmentation, dense correspondence, and feature matching.

Last, we move beyond rotations and isometries and realize the full potential of the extended convolution framework in providing a recipe for constructing convolutional operators equivariant to high-dimensional non-compact transformation groups that act non-linearly. Specifically, Möbius-equivariant extended convolution is used to develop the foundations for Möbius-equivariant spherical CNNs. More generally, this work represents an effort to move both image and surface convolutional networks into the realm of conformal-equivariance, and we believe our network to be the first of its kind in this regard. In particular, our experiments suggest that the latter transition may be especially relevant in the context of shape analysis and recognition. Looking forward, we hope this work serves to catalyze the development of a new generation of surface networks better able to handle the kinds of complex deformations found in real-world shape data.

Appendix I

Transformation of Functions on $\widehat{\mathbb{C}}$

Given a filter f parameterized as in Equation (7.8),

$$f = \sum_{m=-M}^{M} \sum_{s=-N}^{N} b_{ms}\, \mathbf{B}_{ms}^{t},$$

where $\mathbf{B}_{ms}^{t}$ are the spherical log-polar functions

$$\mathbf{B}_{ms}^{t}(z) \equiv \frac{|z|^{is}}{|z|^{t}} \left(\frac{z}{|z|}\right)^{m},$$

we derive the expansion of the transformation of the filter by a lower triangular matrix $L = \left[\begin{smallmatrix} a & 0 \\ n & a^{-1} \end{smallmatrix}\right] \in \mathrm{L} \subset \mathrm{SL}(2, \mathbb{C})$ given in Equation (7.9):

$$L\, f = \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{j=1}^{3} {}_{j}\boldsymbol{\zeta}_{ms}^{t\sigma_j}(L, \mathbf{b})\, \mathbf{B}_{ms}^{\sigma_j}\, ds.$$

Noting that ${}_{j}\boldsymbol{\zeta}_{ms}^{t\sigma_j}(L, \mathbf{b})$ is a linear function in $\mathbf{b}$, we first derive an expansion for the transformation of the log-polar bases $L\,\mathbf{B}_{ms}^{t}$. Then, we substitute this expression into the filter parameterization Equation (7.8), to recover ${}_{j}\boldsymbol{\zeta}_{ms}^{t\sigma_j}(L, \mathbf{b})$. Afterwards we discuss how we approximate the expansion in practice as in Equation (7.10).

A. Transformation of Spherical Log-Polar Bases

Here we derive an expansion of the transformation of the spherical log-polar bases $\mathbf{B}_{ms}^{t}$ by a lower triangular matrix $L \in \mathrm{L}$. Specifically, we seek an expansion which expresses $L\,\mathbf{B}^t_{ms}$ as a linear combination of log-polar bases depending on z, indexed in angular and radial frequencies u and ω and localization variables σ – $\mathbf{B}^{\sigma}_{u\omega}$ – with a set of coefficient functions depending *only* on L.

We consider elements $L \in \mathrm{L} \subset \mathrm{SL}(2, \mathbb{C})$ of the form

$$L = \begin{bmatrix} a & 0 \\ n & a^{-1} \end{bmatrix} \qquad \text{with} \qquad a, n \in \mathbb{C},\ a \neq 0. \tag{I.1}$$

We treat separately the cases where $n = 0$ and $n \neq 0$, first finding an expansion of $L\,\mathbf{B}^t_{ms}$ for each and afterwards combining the two to form an expansion of $L\,\mathbf{B}^t_{ms}$ which holds for all $L \in \mathrm{L}$.

Case 1 $(n = 0)$ **:** If $n = 0$, then for any $z \in \widehat{\mathbb{C}}$

$$L^{-1}z = a^{-2}z,$$

and directly evaluating $\left[L\,\mathbf{B}^t_{ms}\right](z) = \mathbf{B}^t_{ms}(L^{-1}z)$ gives

$$\begin{aligned}
\left[L\,\mathbf{B}^t_{ms}\right](z) &\overset{(7.7)}{=} \frac{|a^{-2}z|^{is}}{|a^{-2}z|^{t}} \left(\frac{a^{-2}z}{|a^{-2}z|}\right)^m \\
&= \frac{|a^{-2}|^{is}}{|a^{-2}|^{t}} \frac{|z|^{is}}{|z|^{t}} \left(\frac{a^{-2}}{|a^{-2}|}\right)^m \left(\frac{z}{|z|}\right)^m \\
&= \frac{|a^{2}|^{-is}}{|a^{2}|^{-t}} \left(\frac{a^{2}}{|a^{2}|}\right)^{-m} \frac{|z|^{is}}{|z|^{t}} \left(\frac{z}{|z|}\right)^m \\
&\overset{(7.7)}{=} \mathbf{B}^{-t}_{-m-s}(a^2)\,\mathbf{B}^t_{ms}(z),
\end{aligned} \tag{I.2}$$

where the last equality provides the desired expansion.

Case 2 $(n \neq 0)$: Here, finding an expansion for $L\,\mathbf{B}^t_{ms}$ is significantly more involved as L^{-1} acts projectively on $\widehat{\mathbb{C}}$. Specifically,

$$L^{-1}z = \frac{a^{-1}z}{a - nz} = \frac{z}{a^2 - anz},$$

which does not allow for a straight-forward separation of variables as in Equation (I.2). Instead, we begin by making two observations. First, denoting $J = \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right] \in$

$\mathrm{SL}(2, \mathbb{C})$, it is easy to show that

$$L^{-1} = J^{-1} L^{\top} J, \tag{I.3}$$

where $L^{\top} \in \mathrm{T}$ is the transpose of L, belonging to the subgroup $\mathrm{U} \in \mathrm{SL}(2, \mathbb{C})$ consisting of all *upper-triangular* elements. Critically, $L^{\top}$ acts on $\widehat{\mathbb{C}}$ not projectively but as an affine transformation,

$$L^{\top} z = a^2 z + an, \tag{I.4}$$

equivalent to a planar rotation and dilation, followed by a translation. Second, for all $z \in \widehat{\mathbb{C}}$,

$$Jz = J^{-1} z = -z^{-1}$$

from which it follows that

$$J\, \mathbf{B}^t_{ms} = \; J^{-1}\, \mathbf{B}^t_{ms} = e^{is\pi}\, \mathbf{B}^{-t}_{-m-s}. \tag{I.5}$$

Combining the observations in Equations (I.3) and (I.5), we have

$$\begin{aligned} L\, \mathbf{B}^t_{ms} &\overset{(I.3)}{=} J^{-1} L^{-\top} J\, \mathbf{B}^t_{ms} \\ &\overset{(I.5)}{=} e^{is\pi} \big[J^{-1} L^{-\top} \mathbf{B}^{-t}_{-m-s} \big], \end{aligned} \tag{I.6}$$

Our strategy now becomes clear. Using Equation (I.6), we can view the transformation of $\mathbf{B}^t_{ms}$ by L as the transformation of $\mathbf{B}^{-t}_{-m-s}$ by $L^{-\top}$, followed by J. This allows us to replace the projective action of L with the affine action of the upper-triangular subgroup U – the group of rotations, translations, and dilations – whose representations are better understood [Vil78, VK91]. Our goal is now to find an expansion of $L^{-\top}\, \mathbf{B}^{-t}_{-m-s}$, which we can convert to the desired expansion for $L\, \mathbf{B}^t_{ms} \in \mathrm{L}$ via the simple action of J^{-1} in Equation (I.5).

We recover an expansion of $L^{-\top}\, \mathbf{B}^{-t}_{-m-s}$ as follows: First, we apply the Hankel transform in the radial dimension which will allow us to represent $\mathbf{B}^{-t}_{-m-s}$ in terms of the irreducible unitary representations (**IURs**) of SE(2) – the group of planar rotations and translations. From here we can use the regular representation of the group to separate the rotational and translational components of $L^{-\top}$. To handle the remaining dilation, we apply the Mellin transform, which results in the desired expansion.

To simplify notation, we convert to polar coordinates

$$z \mapsto (|z|, \operatorname{Arg} z) \equiv (r, \vartheta).$$

In these coordinates $\mathbf{B}^t_{ms}$ becomes

$$\mathbf{B}^t_{ms}(z) \mapsto \mathbf{B}^t_{ms}(r, \vartheta) = r^{is-t}\, e^{im\vartheta}.$$

Similarly, we express a^2, the rotational and dilational component $L^\top z$, and an, the translational component of $L^\top z$, as

$$a^2 = \alpha e^{i\varphi} \qquad \text{and} \qquad an = \tau e^{i\varkappa}$$

for some $\alpha, \tau \in \mathbb{R}_{>0}$ and $\varphi, \varkappa \in [0, 2\pi)$.

The following calculations were performed with the aid of Mathematica 13.0 [Wol21]. We begin by expressing r^{-is+t} in terms of its Hankel expansion in the angular frequency $-m$

$$r^{-is+t} = 2^{1-is+t}\, \mathbf{R}_{ms} \int_0^\infty \varrho^{is-1-t}\, J_{-m}(\varrho r)\, d\varrho, \tag{I.7}$$

where

$$\mathbf{R}_{ms} = \begin{cases} \dfrac{\Gamma\left(1 - \frac{m-t+is}{2}\right)}{\Gamma\left(\frac{-m-t+is}{2}\right)} & m \leq 0 \\[2ex] (-1)^m \dfrac{\Gamma\left(1 - \frac{-m-t+is}{2}\right)}{\Gamma\left(\frac{m-t+is}{2}\right)} & m > 0 \end{cases}. \tag{I.8}$$

and Γ and J_{-m} denote the Gamma function and Bessel functions of the first kind, respectively. Substituting the Hankel expansion of r^{-is+t} in Equation (I.7) into the polar coordinate expression for $\mathbf{B}^{-t}_{-m-s}$ gives

$$\mathbf{B}^{-t}_{-m-s}(r,\vartheta) = 2^{1-is+t}\, e^{-im\vartheta}\, \mathbf{R}_{ms} \int_0^\infty \varrho^{is-1-t}\, J_{-m}(\varrho r)\, d\varrho. \tag{I.9}$$

The matrix elements of the irreducible unitary representations of SE(2) are given by [CK16]

$$h^{\varrho}_{mn}(r,\vartheta,\phi) = i^{n-m}\, e^{-in\phi - i(m-n)\vartheta}\, J_{n-m}(\varrho r), \tag{I.10}$$

where ϕ is the angle of rotation and r and ϑ are the magnitude and polar angle of the translation, respectively. It follows that Equation (I.9) can be written as

$$\mathbf{B}^{-t}_{-m-s}(r,\vartheta) = 2^{1-is+t}\, i^{m}\, \mathbf{R}_{sm} \int_0^\infty \varrho^{is-1-t}\, h^{\varrho}_{m0}(r,\vartheta,0)\, d\varrho. \tag{I.11}$$

From here we can use the regular representation of the group to separate the rotational and translational components of $L^{-\top}$, expanding $L^{-\top}\,\mathbf{B}^{-t}_{-m-s}$ as

$$\begin{aligned}
&\left[L^{-\top}\,\mathbf{B}^{-t}_{-m-s}\right](r,\vartheta) = \\
&2^{1-is+t}\, i^{m}\, \mathbf{R}_{ms} \\
&\times \sum_{u=-\infty}^{\infty} \int_0^\infty \varrho^{is-1-t}\, h^{\varrho}_{mu}(\tau,\varkappa,\varphi)\, h^{\varrho}_{u0}(\alpha r,\vartheta,0)\, d\varrho.
\end{aligned} \tag{I.12}$$

Expanding the integral in Equation (I.12) gives

$$\begin{aligned}
&\int_0^\infty \varrho^{is-1-t} h^{\varrho}_{mu}(\tau,\varkappa,\varphi)\, h^{\varrho}_{u0}(\alpha r,\vartheta,0)\, d\rho \\
&\overset{(I.10)}{=} i^{-m}\, e^{-iu(\varphi+\vartheta)-i(m-u)\varkappa} \\
&\quad \times \int_0^\infty \varrho^{is-1-t}\, J_{m-u}(\tau\varrho)\, J_{-u}(\alpha r\varrho)\, d\varrho \\
&= i^{-m}\, e^{-iu(\varphi+\vartheta)-i(m-u)\varkappa}\, \alpha^{-is+t} \\
&\quad \times \int_0^\infty \varrho^{is-1-t}\, J_{m-u}(\alpha^{-1}\tau\varrho)\, J_{-u}(r\varrho)\, d\varrho && \text{(I.13)} \\
&= 2^{is-1-t}\, i^{-m}\, e^{-iu(\varphi+\vartheta)-i(m-u)\varkappa}\, \tau^{-is+t}\, M^{t}_{smu}(\alpha^2\tau^{-2}r^2), && \text{(I.14)}
\end{aligned}$$

where the second equality follows from the change of variables $r \mapsto \alpha r$, and the third from evaluation of the integral (the inverse Hankel transform in the $(m-u)-$th frequency). Here,

$$M^t_{msu}(r^2) = \begin{cases} G^{1,1}_{2,2}\left(\begin{matrix}\mathbf{x}^t_{msu} \\ \mathbf{y}_u\end{matrix}\middle|\, r^2\right) & u \geq m \\ (-1)^{u-m}\, G^{1,1}_{2,2}\left(\begin{matrix}\mathbf{x}^t_{-ms-u} \\ \mathbf{y}_u\end{matrix}\middle|\, r^2\right) & u < m \end{cases},$$

$$\mathbf{x}^t_{msu} = \left[\frac{1}{2}(2-u-is+m+t),\ \frac{1}{2}(2+u-is-m+t)\right],$$

$$\mathbf{y}_u = \left[-\frac{1}{2}u,\ \frac{1}{2}u\right],$$

with $G^{m,n}_{p,q}\left(\begin{matrix}\mathbf{x} \\ \mathbf{y}\end{matrix}\middle|\, z\right)$ denoting the Meijer G-function [Bat53]. Plugging the expression for the integral in Equation (I.14) into the expression for $L^{-\top}\,\mathbf{B}^{-t}_{-m-s}$ in Equation (I.12) gives

$$\left[L^{-\top}\,\mathbf{B}^{-t}_{-m-s}\right](r,\vartheta) = \mathbf{R}_{ms}\sum_{u=-\infty}^{\infty} e^{-iu(\varphi+\vartheta)-i(m-u)\varkappa}\,\tau^{-is+t}\,M^t_{msu}(\alpha^2\tau^{-2}r^2) \tag{I.15}$$

The above expansion factors out the the rotational and translational components of $L^\top$ as desired, and the final step is to factor out the scale term $\alpha^2\tau^{-2}$ acting on r^2 in the argument of the function M^t_{msu}.

To do so, we decompose M^t_{msu} using the Mellin transform. The basis functions of the Mellin transform $r^{i\omega-\sigma}$ are the irreducable unitary representations of the group of dilations acting via multiplication on the positive real line. By decomposing M^t_{msu} in terms of these bases, we factor out the $\alpha^2\tau^{-2}$ term in the argument using the regular representation of the group in the same manner as was done in Equation (I.2). Specifically, for $0 \leq t < 1$, and real numbers σ_1, σ_2 satisfying

$$t < \sigma_1 < 2 \qquad \text{and} \qquad t-1 < \sigma_2 < 0 \tag{I.16}$$

$M^{t}_{msu}(r^2)$ can be decomposed as a sum of two Mellin transform expansions

$$M^{t}_{msu}(r^2) = \frac{1}{2\pi}\sum_{j=1}^{2}\int_{0}^{\infty} {}_{j}\mathbf{M}^{t,\sigma_j}_{msu}(\omega)\, r^{\sigma_j - i\omega} d\omega. \tag{I.17}$$

where

$$\begin{aligned}&{}_{1}\mathbf{M}^{t,\sigma_1}_{msu}(\omega) = \\ &\begin{cases} \dfrac{\Gamma\left(\frac{\sigma_1+i\omega-u}{2}\right)\Gamma\left(\frac{u+is-m-\sigma_1-i\omega-t}{2}\right)}{2\,\Gamma\left(\frac{2-u-\sigma_1-i\omega}{2}\right)\Gamma\left(\frac{2+u-is-m+\sigma_1+i\omega+t}{2}\right)} & u \geq m,\, u < 0 \\ (-1)^{u}\dfrac{\Gamma\left(\frac{u+\sigma_1+i\omega}{2}\right)\Gamma\left(\frac{u+is-m-\sigma_1-i\omega-t}{2}\right)}{2\,\Gamma\left(\frac{2+u-\sigma_1-i\omega}{2}\right)\Gamma\left(\frac{2+u-is-m+\sigma_1+i\omega+t}{2}\right)} & u \geq m,\, u > 0 \\ (-1)^{u-m}\dfrac{\Gamma\left(\frac{\sigma_1+i\omega-u}{2}\right)\Gamma\left(\frac{-u+is+m-\sigma_1-i\omega-t}{2}\right)}{2\,\Gamma\left(\frac{2-u-\sigma_1-i\omega}{2}\right)\Gamma\left(\frac{2-u-is+m+\sigma_1+i\omega+t}{2}\right)} & u < m,\, u < 0 \\ (-1)^{m}\dfrac{\Gamma\left(\frac{u+\sigma_1+i\omega}{2}\right)\Gamma\left(\frac{-u+is+m-\sigma_1-i\omega-t}{2}\right)}{2\,\Gamma\left(\frac{2+u-\sigma_1-i\omega}{2}\right)\Gamma\left(\frac{2-u-is+m+\sigma_1+i\omega+t}{2}\right)} & u < m,\, u > 0 \\ 0 & u = 0, m \neq 0 \\ \dfrac{\Gamma\left(\frac{2+\sigma_1+i\omega}{2}\right)\Gamma\left(\frac{is-\sigma_1-i\omega-t}{2}\right)}{2\,\left(1-\frac{2-is+t}{2}\right)\Gamma\left(\frac{2-\sigma_1-i\omega}{2}\right)\Gamma\left(\frac{2-is+\sigma_1+i\omega+t}{2}\right)} & u = m = 0 \end{cases}\end{aligned} \tag{I.18}$$

and

$$\begin{aligned}&{}_{2}\mathbf{M}^{t,\sigma_2}_{msu}(\omega) = \\ &\begin{cases} 0 & u \neq 0 \\ \dfrac{\Gamma\left(\frac{\sigma_2+i\omega}{2}\right)\Gamma\left(\frac{is-m-\sigma_2-i\omega-t}{2}\right)}{2\,\Gamma\left(\frac{2-\sigma_2-i\omega}{2}\right)\Gamma\left(\frac{2-is-m+\sigma_2+i\omega+t}{2}\right)} & u = 0,\, m < 0 \\ (-1)^{m}\dfrac{\Gamma\left(\frac{\sigma_2+i\omega}{2}\right)\Gamma\left(\frac{is+m-\sigma_2-i\omega-t}{2}\right)}{2\,\Gamma\left(\frac{2-\sigma_2-i\omega}{2}\right)\Gamma\left(\frac{2-is+m+\sigma_2+i\omega+t}{2}\right)} & u = 0,\, m > 0 \\ \dfrac{\Gamma\left(\frac{\sigma_2+i\omega}{2}\right)\Gamma\left(\frac{2+is-\sigma_2-i\omega-t}{2}\right)}{2\,\left(1-\frac{2-is+t}{2}\right)\Gamma\left(\frac{2-\sigma_2-i\omega}{2}\right)\Gamma\left(\frac{2-is+\sigma_2+i\omega+t}{2}\right)} & u = m = 0 \end{cases}.\end{aligned} \tag{I.19}$$

Here, the bounds of σ_1 and σ_2 are due to the particular properties of the Mellin transform and ensure that the M^t_{msu} can be recovered from the coefficients of the forward transform [VK91]. Then, replacing $M^t_{msu}(\alpha^2\tau^{-2}r^2)$ in Equation (I.15) with its Mellin decomposition in Equation (I.17), rearranging terms, and converting back to complex coordinates $(r, \vartheta) \mapsto (|z|, \operatorname{Arg} z)$ gives

$$\begin{aligned}
&\left[L^{-\top}\, \mathbf{B}^{-t}_{-m-s}\right](z) = \\
&\frac{\mathbf{R}_{sm}}{2\pi} \sum_{u=-\infty}^{\infty} \int_0^{\infty} \sum_{j=1}^{2} \mathbf{B}^{-\sigma_j}_{-u-\omega}(a^2)\, \mathbf{B}^{\sigma_j - t}_{u-m\,\omega-s}(an) \\
&\qquad\qquad \times {}_j\mathbf{M}^{t,\sigma_j}_{msu}(\omega)\, \mathbf{B}^{-\sigma_j}_{-u-\omega}(z)\, d\omega.
\end{aligned} \tag{I.20}$$

Finally, substituting this expression into Equation (I.6) and using Equation (I.5) we arrive at the desired expansion of $L\, \mathbf{B}^t_{ms}$:

$$\begin{aligned}
&\left[L\, \mathbf{B}^{t}_{ms}\right](z) = \\
&\frac{\mathbf{R}_{sm}}{2\pi} \sum_{u=-\infty}^{\infty} \int_0^{\infty} \sum_{j=1}^{2} \mathbf{B}^{-\sigma_j}_{-u-\omega}(a^2)\, \mathbf{B}^{\sigma_j - t}_{u-m\,\omega-s}(an) \\
&\qquad\qquad \times {}_j\mathbf{M}^{t,\sigma_j}_{msu}(\omega)\, \mathbf{B}^{\sigma_j}_{u\omega}(z)\, d\omega.
\end{aligned} \tag{I.21}$$

General Case $(n \in \mathbb{C})$ **:** We combine the expansions of $L\, \mathbf{B}^t_{ms}$ for the cases $n = 0$ in Equation (I.2) and $n \neq 0$ in Equation (I.21) into a general form holding for all $L \in \mathrm{L}$. Specifically, we define the following functions mapping a lower-triangular matrix to a set of filter coefficients,

$$\begin{aligned}
{}^{u\omega}_{\;1}\boldsymbol{\xi}^{t\sigma_1}_{ms}(L) = (1 - \delta_{|n|0})\frac{\mathbf{R}_{sm}}{2\pi}\mathbf{B}^{-\sigma_1}_{-u-\omega}(a^2)& \\
\times\, \mathbf{B}^{\sigma_1 - t}_{u-m\,\omega-s}(an)\, {}_1\mathbf{M}^{t,\sigma_j}_{msu}(\omega),&
\end{aligned} \tag{I.22}$$

$$\begin{aligned}
{}^{u\omega}_{\;2}\boldsymbol{\xi}^{t\sigma_2}_{ms}(L) = (1 - \delta_{|n|0})\frac{\mathbf{R}_{sm}}{2\pi}\mathbf{B}^{-\sigma_2}_{-u-\omega}(a^2)& \\
\times\, \mathbf{B}^{\sigma_2 - t}_{u-m\,\omega-s}(an)\, {}_2\mathbf{M}^{t,\sigma_2}_{msu}(\omega),&
\end{aligned} \tag{I.23}$$

$${}^{u\omega}_{\;3}\boldsymbol{\xi}^{t\sigma_3}_{ms}(L) = \delta_{|n|0}\delta_{mu}\delta(s - \omega)\, \mathbf{B}^{-\sigma_3}_{-m-s}(a^2), \tag{I.24}$$

where δ_{xy} and $\delta(x)$ denote the Kronecker and Dirac delta functions, respectively, and σ_1, σ_2 satisfy the conditions in Equation (I.16). Given $\mathbf{B}^t_{ms}$ for some $t \in (0,1)$ and setting $\sigma_3 = t$, it follows from Equations (I.2) and (I.21) that for any $L \in \mathrm{L}$, $L\,\mathbf{B}^t_{ms}$ can be expanded as

$$L\,\mathbf{B}^t_{ms} = \sum_{u=-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{j=1}^{3} {}^{u\omega}_{\;j}\boldsymbol{\xi}^{t\sigma_j}_{ms}(L)\,\mathbf{B}^{\sigma_j}_{u\omega}\,d\omega \tag{I.25}$$

In practice, we only sum over the first two indices of j, replacing an with a small constant factor $\varepsilon = 0.05$ whenever $|an|$ nears zero.

B. Transformation of Filters

Using the expansion of the transformation of basis functions in Equation (I.25), it is straight-forward to recover the expansion of the transformation of filters of the form

$$f = \sum_{m=-M}^{M} \sum_{s=-N}^{N} b_{ms}\,\mathbf{B}^t_{ms},$$

by elements of L. Namely,

$$\begin{aligned}
L\,f &= \sum_{m=-M}^{M} \sum_{s=-N}^{N} b_{ms}\,L\,\mathbf{B}^t_{ms} \tag{I.26}\\
&\overset{(I.25)}{=} \sum_{m=-M}^{M} \sum_{s=-N}^{N} b_{ms} \sum_{u=-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{j=1}^{3} {}^{u\omega}_{\;j}\boldsymbol{\xi}^{t\sigma_j}_{ms}(L)\,\mathbf{B}^{\sigma_j}_{u\omega}\,d\omega \tag{I.27}\\
&= \sum_{u=-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{j=1}^{3} \underbrace{\left[\sum_{m=-M}^{M} \sum_{s=-N}^{N} b_{ms}\;{}^{u\omega}_{\;j}\boldsymbol{\xi}^{t\sigma_j}_{ms}(L)\right]}_{{}_j\boldsymbol{\zeta}^{t\sigma_j}_{u\omega}(L,\mathbf{b})} \mathbf{B}^{\sigma_j}_{u\omega}\,d\omega, \tag{I.28}
\end{aligned}$$

where $\mathbf{b}$ is the $(2M+1)\times(2N+1)$-dimensional vector of coefficients of f and ${}_j\boldsymbol{\zeta}^{t\sigma_j}_{u\omega}$ maps a lower-triangular element and a set of filter coefficients to the coefficient of $\mathbf{B}^{\sigma_j}_{us\omega}$ in the expansion.

C. Implementation

As discussed in §7.4 we approximate the expansion of $L f$ by truncating the summation over u, and replacing the integration over the real line with a discrete approximation using quadrature. The summation is a consequence of the addition theorem for Bessel functions [Wat95]. Here it arises in the regular representation of SE(2) used in Equation (I.12) to factor the rotational and translational components of the transformations from the argument of the basis functions. Fortunately, it converges rapidly for low angular basis frequencies m and we typically find truncation at $M+1$ terms to be sufficient.

Approximating the integral in the expansion is less straight-forward. For example, the reader may have noticed that the second to last equality – Equation (I.13) – in the expansion of the integral in Equation (I.12) provides a seemingly suitable separation of variables for our purposes, raising the question of why we expend the additional effort dealing with the Mellin transform. The problem with the expansion offered by Equation (I.13) is that the product of Bessel functions in the integrand is highly-oscillatory, and decays either very rapidly or very slowly depending on the values of α, τ and r making a low-order numerical integration scheme impossible.

However, it turns out that first recollecting the separated terms by evaluating the integral (the inverse Hankel transform) – Equation (I.14) – and then expanding the solution again using the Mellin transform – Equation (I.17) – gives us something we can handle numerically. (Equivalently, we could have first expanded $J_{-u}(r\rho)$ in Equation (I.13) using the Mellin transform, then applied the inverse Hankel transform to arrive at a similar expression). Despite being aesthetically-challenged, the Mellin transform coefficients ${}_{j}\mathbf{M}_{msu}^{t,\sigma_j}$ in Equations (I.18 - I.19) have several nice properties which make possible a low-order quadrature approximation of the integral. Specifically, they decay rapidly, are relatively smooth, and retain these properties even with increasing values of $|m|, |s|$ and $|u|$. Furthermore, for a given choice of $t \in (0, 1)$ determining the localization of the filters, the smoothness and decay of the Mellin coefficients can further be controlled by the choices of σ_1 and σ_2 satisfying Equation (I.16). In our implementations with $t = 0.15$, we set $\sigma_1 = -0.35$ and $\sigma_2 = 0.15$ as we heuristically observe they improve the smoothness and localisation of the coefficients, allowing us to better approximate the integral with fewer quadrature samples. Furthermore, since the Mellin transform coefficients ${}_j\mathbf{M}^{t,\sigma_j}_{msu}$ in Equations (I.22-I.23) depend only on the radial frequency ω and are independent of both L and the filter coefficients $\mathbf{b}$ they are computed in a pre-processing step, avoiding the evaluation of the Gamma functions at run-time.

References Cited

[ABD12] Pablo Fernández Alcantarilla, Adrien Bartoli, and Andrew J Davison. KAZE features. In *European Conference on Computer Vision.*, pages 214–227, 2012.

[Ado16] Adobe. Adobe mixamo 3D characters, 2016. www.mixamo.com.

[AML18] Matan Atzmon, Haggai Maron, and Yaron Lipman. Point convolutional neural networks by extension operators. *arXiv preprint arXiv:1803.10091*, 2018.

[AS11] Pablo F Alcantarilla and T Solutions. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *Transactions on Pattern Analysis and Machine Intelligence*, 34:1281–1298, 2011.

[ASK+05] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. Scape: Shape completion and animation of people. *Transactions on Graphics*, 24(3):408–416, 2005.

[ASZS17] Iro Armeni, Sasha Sax, Amir R Zamir, and Silvio Savarese. Joint 2d-3d-semantic data for indoor scene understanding. *arXiv preprint arXiv:1702.01105*, 2017.

[BAO+20] Alexander Bogatskiy, Brandon Anderson, Jan Offermann, Marwah Roussi, David Miller, and Risi Kondor. Lorentz group equivariant neural network for particle physics. In *International Conference on Machine Learning*, pages 992–1002. PMLR, 2020.

[Bat53] Harry Bateman. *Higher transcendental functions Vol. 1*, volume 1. McGraw-Hill Book Company, 1953.

[BBB+11] Edmond Boyer, Alexander M Bronstein, Michael M Bronstein, Benjamin Bustos, Tal Darom, Radu Horaud, Ingrid Hotz, Yosi Keller, Johannes Keustermans, Artiom Kovnatsky, et al. SHREC 2011: Robust feature detection and description benchmarkk. In *Eurographics Workshop on 3D Object Retrieval*, 2011.

[BBC+10] Alex Bronstein, Michael Bronstein, Umberto Castellani, Anastasia Dubrovina, LJ Guibas, RP Horaud, Ron Kimmel, David Knossow, Etienne Von Lavante, Diana Mateus, et al. SHREC 2010: Robust feature detection and description benchmark. In *Eurographics Workshop on 3D Object Retrieval*, 2010.

[BBK08] Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. *Numerical Geometry of Non-Rigid Shapes*. Springer Verlag, 2008.

[BBL+17] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

[BCK18] Alex Baden, Keenan Crane, and Misha Kazhdan. Möbius registration. In *Computer Graphics Forum*, volume 37, pages 211–220. Wiley Online Library, 2018.

[BETVG08] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110:346–359, 2008.

[BK10] Michael M Bronstein and Iasonas Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *Computer Vision and Pattern Recognition*, pages 1704–1711, 2010.

[BMR+16] Davide Boscaini, Jonathan Masci, Emanuele Rodolà, Michael M Bronstein, and Daniel Cremers. Anisotropic diffusion descriptors. In *Computer Graphics Forum*, volume 35, pages 431–441. Wiley Online Library, 2016.

[BMRB16] Davide Boscaini, Jonathan Masci, Emanuele Rodolà, and Michael Bronstein. Learning shape correspondence with anisotropic convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 3189–3197, 2016.

[Bra00] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[BRLB14] Federica Bogo, Javier Romero, Matthew Loper, and Michael J. Black. FAUST: Dataset and evaluation for 3D mesh registration. In *Computer Vision and Pattern Recognition*. IEEE, 2014.

[BTVG06] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. SURF: Speeded up robust features. In *European Conference on Computer Vision.*, pages 404–417, 2006.

[CGKW18] Taco S Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical cnns. *arXiv preprint arXiv:1801.10130*, 2018.

[CGW19] Taco S Cohen, Mario Geiger, and Maurice Weiler. A general theory of equivariant cnns on homogeneous spaces. In *Advances in Neural Information Processing Systems*, pages 9145–9156, 2019.

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

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

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

[CPK19] Christopher Choy, Jaesik Park, and Vladlen Koltun. Fully convolutional geometric features. In *International Conference on Computer Vision*, pages 8958–8966, 2019.

[CPS11] Keenan Crane, Ulrich Pinkall, and Peter Schröder. Spin transformations of discrete surfaces. *Transactions on Graphics*, 30, 2011.

[CRB$^+$16] Luca Cosmo, Emanuele Rodolà, Michael M Bronstein, Andrea Torsello, Daniel Cremers, and Y Sahillioglu. SHREC '16: Partial matching of deformable shapes. *Eurographics Workshop on 3D Object Retrieval*, 2016.

[CW16] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on Machine learning*, pages 2990–2999. PMLR, 2016.

[CWKW19a] Taco Cohen, Maurice Weiler, Berkay Kicanaoglu, and Max Welling. Gauge equivariant convolutional networks and the icosahedral CNN. In *International conference on Machine learning*, pages 1321–1330. PMLR, 2019.

[CWKW19b] Taco Cohen, Maurice Weiler, Berkay Kicanaoglu, and Max Welling. Gauge equivariant convolutional networks and the icosahedral CNN. In *International conference on Machine learning*, volume 97, pages 1321–1330, 2019.

[DBI18] Haowen Deng, Tolga Birdal, and Slobodan Ilic. PPFnet: Global context aware local features for robust 3d point matching. In *Computer Vision and Pattern Recognition*, pages 195–205, 2018.

[DBI19] Haowen Deng, Tolga Birdal, and Slobodan Ilic. 3D local features for direct pairwise registration. In *Computer Vision and Pattern Recogni-*

tion, pages 3244–3253, 2019.

[DBV16] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*, page 3844–3852, 2016.

[DH94] James R Driscoll and Dennis M Healy. Computing fourier transforms and convolutions on the 2-sphere. *Advances in applied mathematics*, 15(2):202–250, 1994.

[dHWCW20] Pim de Haan, Maurice Weiler, Taco Cohen, and Max Welling. Gauge equivariant mesh CNNs: Anisotropic convolutions on geometric graphs. *arXiv preprint arXiv:2003.05425*, 2020.

[DSL$^+$19] R. M. Dyke, C. Stride, Y.-K. Lai, P. L. Rosin, M. Aubry, A. Boyarski, A. M. Bronstein, M. M. Bronstein, D. Cremers, M. Fisher, T. Groueix, D. Guo, V. G. Kim, R. Kimmel, Z. Lähner, K. Li, O. Litany, T. Remez, E. Rodolà, B. C. Russell, Y. Sahillioğlu, R. Slossberg, G. K. L. Tam, M. Vestner, Z. Wu, and J. Yang. Shape correspondence with isometric and non-isometric deformations. In Silvia Biasotti, Guillaume Lavoué, and Remco Veltkamp, editors, *Eurographics Workshop on 3D Object Retrieval*. The Eurographics Association, 2019.

[DSO20] Nicolas Donati, Abhishek Sharma, and Maks Ovsjanikov. Deep geometric functional maps: Robust feature learning for shape correspondence. In *Computer Vision and Pattern Recognition*, pages 8592–8601, 2020.

[EABMD18] Carlos Esteves, Christine Allen-Blanchette, Ameesh Makadia, and Kostas Daniilidis. Learning so (3) equivariant representations with spherical cnns. In *European Conference on Computer Vision.*, pages

52–68, 2018.

[EMD20] Carlos Esteves, Ameesh Makadia, and Kostas Daniilidis. Spin-weighted spherical cnns. *Advances in Neural Information Processing Systems*, 2020.

[EMSJB14] Ghina El Mir, Christophe Saint-Jean, and Michel Berthier. Conformal geometry for viewpoint change representation. *Advances in Applied Clifford Algebras*, 24(2):443–463, 2014.

[ESKBC17] Danielle Ezuz, Justin Solomon, Vladimir G. Kim, and Mirela Ben-Chen. GWCNN: A metric alignment layer for deep shape analysis. *Computer Graphics Forum*, 36(5):49–57, 2017.

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

[FELWM18] Matthias Fey, Jan Eric Lenssen, Frank Weichert, and Heinrich Müller. SplineCNN: Fast geometric deep learning with continuous B-spline kernels. In *Computer Vision and Pattern Recognition*, pages 869–877, 2018.

[FL19] Matthias Fey and Jan E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

[FM82] Kunihiko Fukushima and Sei Miyake. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets*, pages 267–285. Springer, 1982.

[FSIW20] Marc Finzi, Samuel Stanton, Pavel Izmailov, and Andrew Gordon Wilson. Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data. In *International conference on Machine learning*, pages 3165–3176. PMLR, 2020.

[FWW21] Marc Finzi, Max Welling, and Andrew Gordon Wilson. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. *arXiv preprint arXiv:2104.09459*, 2021.

[GBP07] Daniela Giorgi, Silvia Biasotti, and Laura Paraboschi. Shape retrieval contest 2007: Watertight models track. *SHREC competition*, 8(7), 2007.

[GBS+16] Yulan Guo, Mohammed Bennamoun, Ferdous Sohel, Min Lu, Jianwei Wan, and Ngai Ming Kwok. A comprehensive performance evaluation of 3D local feature descriptors. *International Journal of Computer Vision*, 116:66–89, 2016.

[GCBZ19] Shunwang Gong, Lei Chen, Michael Bronstein, and Stefanos Zafeiriou. Spiralnet++: A fast and highly efficient mesh convolution operator. In *Computer Vision and Pattern Recognition Workshops*, 2019.

[GFK+18] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan C Russell, and Mathieu Aubry. 3d-coded: 3d correspondences by deep deformation. In *European Conference on Computer Vision.*, pages 230–246, 2018.

[GQ20] Jean H Gallier and Jocelyn Quaintance. *Differential Geometry and Lie Groups: A Computational Perspective*, volume 12. Springer Nature, 2020.

[GSB+13] Yulan Guo, Ferdous Sohel, Mohammed Bennamoun, Min Lu, and Jianwei Wan. Rotational projection statistics for 3D local surface description and object recognition. *International Journal of Computer Vision*, 105:63–86, 2013.

[GWC+04] Xianfeng Gu, Yalin Wang, T.F. Chan, P.M. Thompson, and Shing-Tung

Yau. Genus zero surface conformal mapping and its application to brain surface mapping. *IEEE Transactions on Medical Imaging*, 23(8):949–958, 2004.

[HA19] Philipp Herholz and Marc Alexa. Efficient computation of smoothed exponential maps. *Computer Graphics Forum*, 38:79–90, 2019.

[HHF+19] Rana Hanocka, Amir Hertz, Noa Fish, Raja Giryes, Shachar Fleishman, and Daniel Cohen-Or. MeshCNN: A network with an edge. *Transactions on Graphics*, 38(4):90, 2019.

[HJZS20] Wenchong He, Zhe Jiang, Chengming Zhang, and Arpan Man Sainju. *CurvaNet: Geometric Deep Learning Based on Directional Curvature for 3D Shape Analysis*, page 2214–2224. Association for Computing Machinery, New York, NY, USA, 2020.

[HLS18] Kun He, Yan Lu, and Stan Sclaroff. Local descriptors optimized for average precision. In *Computer Vision and Pattern Recognition*, pages 596–605, 2018.

[HSBH+19] Niv Haim, Nimrod Segol, Heli Ben-Hamu, Haggai Maron, and Yaron Lipman. Surface networks via general covers. In *International Conference on Computer Vision*, pages 632–641, 2019.

[HZ03] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge books online. Cambridge University Press, 2003.

[HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[JHK+18] Chiyu Max Jiang, Jingwei Huang, Karthik Kashinath, Philip Marcus, Matthias Niessner, et al. Spherical cnns on unstructured grids. In *International Conference on Learning Representations*, 2018.

[JMM+20] Yuhe Jin, Dmytro Mishkin, Anastasiia Mishchuk, Jiri Matas, Pascal Fua, Kwang Moo Yi, and Eduard Trulls. Image matching across wide baselines: From paper to practice. *arXiv preprint arXiv:2003.01587*, 2020.

[KB15] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.

[KBLB12] Iasonas Kokkinos, Michael M Bronstein, Roee Litman, and Alex M Bronstein. Intrinsic shape context descriptors for deformable shapes. In *Computer Vision and Pattern Recognition*, pages 159–166, 2012.

[KCPS13] Felix Knöppel, Keenan Crane, Ulrich Pinkall, and Peter Schröder. Globally optimal direction fields. *Transactions on Graphics*, 32(4), 2013.

[KFR04] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Symmetry descriptors and 3D shape matching. In *Eurographics Symposium on Geometry Processing 2004*, volume 2, pages 116–125, 2004.

[KJM05] V. Kumar, R. Juday, and A. Mahalanobis. *Correlation Pattern Recognition*. Cambridge University Press, 2005.

[KPS17] Ebrahim Karami, Siva Prasad, and Mohamed Shehata. Image matching using SIFT, SURF, BRIEF and ORB: performance comparison for distorted images. *arXiv preprint arXiv:1710.02726*, 2017.

[KR08] Peter J Kostelec and Daniel N Rockmore. Ffts on the rotation group. *Journal of Fourier analysis and applications*, 14(2):145–179, 2008.

[KS01] A. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. Society of Industrial and Applied Mathematics, 2001.

[KS04] Yan Ke and Rahul Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition*, volume 2, pages 506–513. IEEE, 2004.

[KS06] Y. Keller and Y. Shkolnisky. A signal processing approach to symmetry detection. *Transactions on Image Processing*, 15:2198–2207, 2006.

[KSBC12] Michael Kazhdan, Jake Solomon, and Mirela Ben-Chen. Can mean-curvature flow be modified to be non-singular? In *Computer Graphics Forum*, volume 31, pages 1745–1754. Wiley Online Library, 2012.

[KZK17] Marc Khoury, Qian-Yi Zhou, and Vladlen Koltun. Learning compact geometric features. In *International Conference on Computer Vision*, pages 153–161, 2017.

[LB+95] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

[LB13] Roee Litman and Alexander M Bronstein. Learning spectral descriptors for deformable shape correspondence. *Transactions on Pattern Analysis and Machine Intelligence*, 36(1):171–180, 2013.

[LBD+89] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[LCS11] Stefan Leutenegger, Margarita Chli, and Roland Y Siegwart. BRISK: Binary robust invariant scalable keypoints. In *International Conference on Computer Vision*, pages 2548–2555, 2011.

[Lee12] John M. Lee. *Smooth Manifolds*. Springer New York, New York, NY, 2012.

[LGB+11] Zhouhui Lian, Afzal Godil, Benjamin Bustos, Mohamed Daoudi, Jeroen Hermans, Shun Kawamura, Yukinori Kurita, Guillaume Lavoué, Hien Nguyen, Ryutarou Ohbuchi, Yuki Ohkita, Yuya Ohishi, Fatih Porikli, Martin Reuter, Ivan Sipiran, Dirk Smeets, Paul Suetens, Hedi Tabia, and Dirk Vandermeulen. Shrec '11 track: Shape retrieval on non-rigid 3d watertight meshes. In *Eurographics Workshop on 3D Object Retrieval*, pages 79–88, 01 2011.

[LH17] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts, 2017.

[LLHL20] Qinsong Li, Shengjun Liu, Ling Hu, and Xinru Liu. Shape correspondence using anisotropic chebyshev spectral CNNs. In *Computer Vision and Pattern Recognition*, pages 14658–14667, 2020.

[LMBB18] Ron Levie, Federico Monti, Xavier Bresson, and Michael M Bronstein. Cayleynets: Graph convolutional neural networks with complex rational spectral filters. *Transactions on Signal Processing*, 67(1):97–109, 2018.

[Low99] David G Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, volume 2, pages 1150–1157, 1999.

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

[LPRM02] Bruno Lévy, Sylvain Petitjean, Nicolas Ray, and Jérome Maillot. Least squares conformal maps for automatic texture atlas generation. *ACM Trans. Graph.*, 21(3):362–371, 2002.

[LRB+16] Zorah Lähner, Emanuele Rodolà, Michael M Bronstein, Daniel Cremers, Oliver Burghard, Luca Cosmo, Andreas Dieckmann, Reinhard Klein, and Yusuf Sahillioglu. SHREC'16: Matching of deformable shapes with topological noise. *Eurographics Workshop on 3D Object Retrieval*, 2016.

[LRF10] Yaron Lipman, Raif M Rustamov, and Thomas A Funkhouser. Biharmonic distance. *Transactions on Graphics*, 29:1–11, 2010.

[LSZ+18] Zixin Luo, Tianwei Shen, Lei Zhou, Siyu Zhu, Runze Zhang, Yao Yao, Tian Fang, and Long Quan. Geodesc: Learning local descriptors by integrating geometry constraints. In *European Conference on Computer Vision.*, pages 168–183, 2018.

[LT20] Alon Lahav and Ayellet Tal. Meshwalker: Deep mesh understanding by random walks. *Transactions on Graphics*, 39(6):1–13, 2020.

[LW21] Leon Lang and Maurice Weiler. A wigner-eckart theorem for group equivariant convolution kernels. In *International conference on Machine learning*, 2021.

[LYLG18] Kun Li, Jingyu Yang, Yu-Kun Lai, and Daoliang Guo. Robust nonrigid registration with reweighted position and transformation sparsity. *IEEE Transactions on Visualization and Computer Graphics*, 25(6):2255–2269, 2018.

[MAKK22] Thomas W. Mitchel, Noam Aigerman, Vladimir G. Kim, and Michael Kazhdan. Möbius Convolutions for Spherical CNNs. *arXiv preprint arXiv:2201.12212*, 2022.

[MBBV15] Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. Geodesic convolutional neural networks on riemannian manifolds. In *International Conference on Computer Vision*, pages 37–45, 2015.

[MBK+20] Thomas W. Mitchel, Benedict Brown, David Koller, Tim Weyrich, Szymon Rusi-nkiewicz, and Michael Kazhdan. Efficient Spatially Adaptive Convolution and Correlation. *arXiv preprint arXiv:2006.13188*, 2020.

[MBM+17] Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda, and Michael M Bronstein. Geometric deep learning on graphs and manifolds using mixture model CNNs. In *Computer Vision and Pattern Recognition*, volume 1, page 3. IEEE, 2017.

[MGA+17] Haggai Maron, Meirav Galun, Noam Aigerman, Miri Trope, Nadav Dym, Ersin Yumer, Vladimir G. Kim, and Yaron Lipman. Convolutional neural networks on surfaces via seamless toric covers. *Transactions on Graphics*, 36(4):71, 2017.

[Mis19] Diganta Misra. Mish: A self regularized non-monotonic activation function. *arXiv preprint arXiv:1908.08681*, 2019.

[MK10] Jan Möbius and Leif Kobbelt. OpenFlipper: an open source geometry processing and rendering framework. In *International Conference on Curves and Surfaces*, pages 488–500. Springer, 2010.

[MKK21] Thomas W. Mitchel, Vladimir G. Kim, and Michael Kazhdan. Field Convolutions for Surface CNNs. In *International Conference on Computer Vision*, pages 10001–10011, 2021.

[MLR+20] Francesco Milano, Antonio Loquercio, Antoni Rosinol, Davide Scaramuzza, and Luca Carlone. Primal-dual mesh convolutional neural networks, 2020.

[MMRM17] Anastasiia Mishchuk, Dmytro Mishkin, Filip Radenovic, and Jiri Matas. Working hard to know your neighbor's margins: Local descriptor learning loss. In *Advances in Neural Information Processing Systems*, pages 4826–4837, 2017.

[MR12] Eivind Lyche Melvær and Martin Reimers. Geodesic polar coordinates on polygonal meshes. *Computer Graphics Forum*, 31:2423–2435, 2012.

[MRCK21] Thomas W. Mitchel, Szymon Rusinkiewicz, Gregory S. Chirikjian, and

Michael Kazhdan. ECHO: Extended Convolution Histogram of Orientations for Local Surface Description. *Computer Graphics Forum*, 40(1):180–194, 2021.

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

[Nat01] F. Natterer. *The Mathematics of Computerized Tomography*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 2001.

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

[PGM+19] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32:8026–8037, 2019.

[PO18] Adrien Poulenard and Maks Ovsjanikov. Multi-directional geodesic neural networks via equivariant convolution. *Transactions on Graphics*, 37(6):236:1–236:14, 2018.

[PRPO19] Adrien Poulenard, Marie-Julie Rakotosaona, Yann Ponty, and Maks Ovsjanikov. Effective rotation-invariant point CNN with spherical harmonics kernels. In *3D Vision*, pages 47–56, 2019.

[PVG+11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[QSMG17] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Computer Vision and Pattern Recognition*, pages 652–660, 2017.

[QYSG17] Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv preprint arXiv:1706.02413*, 2017.

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

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

[RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[RRKB11] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. ORB: An efficient alternative to SIFT or SURF. In *International Conference on Computer Vision*, pages 2564–2571, 2011.

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

[SACO20] Nicholas Sharp, Souhaib Attaiki, Keenan Crane, and Maks Ovsjanikov. Diffusion is all you need for learning on surfaces. *arXiv preprint arXiv:2012.00888*, 2020.

[SBR16] Ayan Sinha, Jing Bai, and Karthik Ramani. Deep learning 3d shape surfaces using geometry images. In *European Conference on Computer Vision.*, pages 223–240. Springer, 2016.

[SBS06] O. Schall, A. Belyaev, and H.P. Seidel. Adaptive Fourier-based surface reconstruction. In *Geometric Modeling and Processing*, volume 4, pages 34–44, 2006.

[SC20] Nicholas Sharp and Keenan Crane. A Laplacian for Nonmanifold Triangle Meshes. *Computer Graphics Forum*, 39(5), 2020.

[SDL18] Stefan C Schonsheck, Bin Dong, and Rongjie Lai. Parallel transport convolution: A new tool for convolutional neural networks on manifolds. *arXiv preprint arXiv:1805.07857*, 2018.

[SF96] Eero P Simoncelli and Hany Farid. Steerable wedge filters for local orientation analysis. *Transactions on Image Processing*, 5:1377–1382, 1996.

[SGW06] Ryan Schmidt, Cindy Grimm, and Brian Wyvill. Interactive decal compositing with discrete exponential maps. *Transactions on Graphics*, 25:605–613, 2006.

[SHG$^+$20] Zhiyu Sun, Yusen He, Andrey Gritsenko, Amaury Lendasse, and Stephen Baek. Embedded spectral descriptors: learning the pointwise correspondence metric via siamese neural networks. *Journal of Computational Design and Engineering*, 7(1):18–29, 2020.

[SHSP17] Johannes L Schonberger, Hans Hardmeier, Torsten Sattler, and Marc Pollefeys. Comparative evaluation of hand-crafted and learned local features. In *Computer Vision and Pattern Recognition*, pages 1482–1491, 2017.

[SK20] Saurabh Singh and Shankar Krishnan. Filter response normalization layer: Eliminating batch dependence in the training of deep neural networks. In *Computer Vision and Pattern Recognition*, pages 11237–11246, 2020.

[SMDH13] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147. PMLR, 2013.

[SMKF04] Philip Shilane, Patrick Min, Michael Kazhdan, and Thomas Funkhouser. The Princeton shape benchmark. In *Proceedings Shape Modeling Applications*, pages 167–178, 2004.

[SMKLM15] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *International Conference on Computer Vision*, pages 945–953, 2015.

[SOG09] Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *Computer Graphics Forum*, volume 28, pages 1383–1392. Wiley Online Library, 2009.

[SR$^+$21] Mehran Shakerinava, Siamak Ravanbakhsh, et al. Equivariant networks for pixelized spheres. *arXiv preprint arXiv:2106.06662*, 2021.

[SS16] Saul Schleimer and Henry Segerman. Squares that look round: Transforming spherical images. *CoRR*, abs/1605.01396, 2016.

[SSC19a] Nicholas Sharp, Yousuf Soliman, and Keenan Crane. The vector heat method. *Transactions on Graphics*, 38:1–19, 2019.

[SSC19b] Nicholas Sharp, Yousuf Soliman, and Keenan Crane. The vector heat method. *Transactions on Graphics*, 38(3), 2019.

[SSDS19] Riccardo Spezialetti, Samuele Salti, and Luigi Di Stefano. Performance evaluation of learned 3D features. In *International Conference on Image Analysis and Processing*, pages 519–531. Springer, 2019.

[SSS19a] Ivan Sosnovik, Michał Szmaja, and Arnold Smeulders. Scale-equivariant steerable networks. In *International Conference on Learning Representations*, 2019.

[SSS19b] Riccardo Spezialetti, Samuele Salti, and Luigi Di Stefano. Learning an effective equivariant 3D descriptor without supervision. In *International Conference on Computer Vision*, pages 6401–6410, 2019.

[STDS14] Samuele Salti, Federico Tombari, and Luigi Di Stefano. Shot: Unique signatures of histograms for surface and texture description. *Computer Vision and Image Understanding*, 125:251–264, 2014.

[SVI+16] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.

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

[TLF09] Engin Tola, Vincent Lepetit, and Pascal Fua. DAISY: An efficient dense descriptor applied to wide-baseline stereo. *Transactions on Pattern Analysis and Machine Intelligence*, 32:815–830, 2009.

[TQD+19] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Computer Vision and Pattern Recognition*, pages 6411–6420, 2019.

[TS18] Shaharyar Ahmed Khan Tareen and Zahra Saleem. A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. In *2018 International Conference on Computing, Mathematics and Engineering Tech-*

nologies, pages 1–10, 2018.

[TSDS10a] Federico Tombari, Samuele Salti, and Luigi Di Stefano. Unique signatures of histograms for local surface description. In *European Conference on Computer Vision.*, pages 356–369, 2010.

[TSDS10b] Federico Tombari, Samuele Salti, and Luigi Di Stefano. Unique signatures of histograms for local surface description. In *European Conference on Computer Vision.*, pages 356–369. Springer, 2010.

[TSK+18] Nathaniel Thomas, Tess Smidt, Steven M. Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. *arXiv:1802.08219*, 2018.

[VBMP08] Daniel Vlasic, Ilya Baran, Wojciech Matusik, and Jovan Popovic. Articulated mesh animation from multi-view silhouettes. *Transactions on Graphics*, 27(3), 2008.

[VBV18] Nitika Verma, Edmond Boyer, and Jakob Verbeek. Feastnet: Feature-steered graph convolutions for 3d shape analysis. In *Computer Vision and Pattern Recognition*, pages 2598–2606. IEEE, 2018.

[VdMH08] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11), 2008.

[VHSF+21] Soledad Villar, David Hogg, Kate Storey-Fisher, Weichi Yao, and Ben Blum-Smith. Scalars are universal: Equivariant machine learning, structured like classical physics. *Advances in Neural Information Processing Systems*, 34, 2021.

[Vil78] N. Ja. Vilenkin. *Special functions and the theory of group representations*, volume 22. American Mathematical Soc., 1978.

[VK91] N. Ja. Vilenkin and A. U. Klimyk. Representation of lie groups and special functions: Volume 1: Simplest lie groups, special functions and integral transforms (mathematics and its applications), 1991.

[VLB+17] Matthias Vestner, Zorah Lähner, Amit Boyarski, Or Litany, Ron Slossberg, Tal Remez, Emanuele Rodola, Alex Bronstein, Michael Bronstein, Ron Kimmel, et al. Efficient deformable shape correspondence via kernel matching. In *3D Vision*, pages 517–526, 2017.

[Wal91] G. Wallace. The JPEG still picture compression standard. *Communications of the ACM*, 34:30–44, 1991.

[Wat95] George Neville Watson. *A treatise on the theory of Bessel functions*. Cambridge university press, 1995.

[WC19] Maurice Weiler and Gabriele Cesa. General e (2)-equivariant steerable cnns. *Advances in Neural Information Processing Systems*, 32:14334–14345, 2019.

[WEH20] Ruben Wiersma, Elmar Eisemann, and Klaus Hildebrandt. CNNs on surfaces using rotation-equivariant features. *Transactions on Graphics*, 39(4):92–1, 2020.

[WGTB17] Daniel E Worrall, Stephan J Garbin, Daniyar Turmukhambetov, and Gabriel J Brostow. Harmonic networks: Deep translation and rotation equivariance. In *Computer Vision and Pattern Recognition*, pages 5028–5037, 2017.

[WGY+18] Hanyu Wang, Jianwei Guo, Dong-Ming Yan, Weize Quan, and Xiaopeng Zhang. Learning 3D keypoint descriptors for non-rigid shape matching. In *European Conference on Computer Vision.*, pages 3–19, 2018.

[WNEH21] Ruben Wiersma, Ahmad Nasikun, Elmar Eisemann, and Klaus Hildebrandt. Deltaconv: Anisotropic point cloud learning with exterior calculus. *arXiv preprint arXiv:2111.08799*, 2021.

[Wol21] Wolfram Research Inc. Mathematica, Version 13.0.0, 2021. Champaign, IL.

[WSK+15] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920, 2015.

[WSL+19] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph CNN for learning on point clouds. *Transactions on Graphics*, 38(5):1–12, 2019.

[WW19] Daniel E Worrall and Max Welling. Deep scale-spaces: Equivariance over scale. *arXiv preprint arXiv:1905.11697*, 2019.

[YLP+20] Yuqi Yang, Shilin Liu, Hao Pan, Yang Liu, and Xin Tong. PFCNN: Convolutional neural networks on 3d surfaces using parallel frames. In *Computer Vision and Pattern Recognition*, June 2020.

[YSGG17] Li Yi, Hao Su, Xingwen Guo, and Leonidas J. Guibas. Syncspeccnn: Synchronized spectral cnn for 3d shape segmentation. In *Computer Vision and Pattern Recognition*, July 2017.

[ZBH12] Andrei Zaharescu, Edmond Boyer, and Radu Horaud. Keypoints and local descriptors of scalar functions on 2D manifolds. *International Journal of Computer Vision*, 100:78–98, 2012.

[ZBL+20] Yongheng Zhao, Tolga Birdal, Jan Eric Lenssen, Emanuele Menegatti, Leonidas Guibas, and Federico Tombari. Quaternion equivariant capsule networks for 3d point clouds. In *European Conference on Computer Vision.*, pages 1–19. Springer, 2020.

[ZFR19] Linguang Zhang, Adam Finkelstein, and Szymon Rusinkiewicz. High-precision localization using ground texture. In *International Conference on Robotics and Automation*, 2019.

[ZLSC19] Chao Zhang, Stephan Liwicki, William Smith, and Roberto Cipolla. Orientation-aware semantic segmentation on icosahedron spheres. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3533–3541, 2019.

[ZMT05] Eugene Zhang, Konstantin Mischaikow, and Greg Turk. Feature-based surface parameterization and texture mapping. *Transactions on Graphics*, 24:1–27, 2005.

[ZR19] Linguang Zhang and Szymon Rusinkiewicz. Learning local descriptors with a CDF-based dynamic soft margin. In *International Conference on Computer Vision*, 2019.