

# Shape recognition with spectral distances

Michael M. Bronstein *Member, IEEE*, Alexander M. Bronstein *Member, IEEE*,

**Abstract**—Recent works have shown the use of diffusion geometry for various pattern recognition applications, including non-rigid shape analysis. In this paper, we introduce spectral shape distance as a general framework for distribution-based shape similarity and show that two recent methods for shape similarity due to Rustomov and Mahmoudi & Sapiro are particular cases thereof.

**Index Terms**—diffusion distance, commute time, spectral distance, eigenmap, Laplace-Beltrami operator, heat kernel, distribution, global point signature, non-rigid shapes, similarity.



## 1 INTRODUCTION

Recent works [1], [9], [23], [15], [6], [16] have shown the use of *diffusion geometry* for various pattern recognition applications, including analysis of non-rigid shape and surfaces. In particular, diffusion geometry arising from heat propagation on a surface [24], [3] allows defining shape similarity criteria which are *intrinsic*, i.e., invariant to inelastic deformations of the surface. The latter property is especially important in the comparison and retrieval of non-rigid shapes possessing a high degree of flexibility [4], [22], [11]. Also, diffusion geometry appears to be robust to topological noise [6], unlike other geometries, e.g., geodesic [5].

In this paper, we propose the spectral shape distance, based on distributions of generic diffusion distances, and show that this framework generalizes previously proposed approaches. In particular, our formulation allows to compare the method of Rustomov [23] (modeling shapes as distribution of commute time distances) and of Mahmoudi and Sapiro [15] (modeling shapes as distributions of diffusion distances) and show the relation between them.

The rest of the paper is organized as follows. In Section 2, we introduce the mathematical background (for a comprehensive overview of the notions in diffusion geometry in the continuous and discrete case, we refer the reader to [9], [12] and [20], respectively). In Section 3, we present the construction of the spectral shape distances and discuss their invariance properties. A discussion is dedicated to the comparison of the methods in [23], [15], which are regarded as particular cases of our framework.

In Section 4, we sketch standard approaches to the discretization of diffusion geometry, and in Section 5 show experimental results. Finally, Section 6 concludes the paper.

## 2 BACKGROUND

Let  $X$  be a shape, modeled as a compact Riemannian manifold. In the following, we denote by  $\mu$  the standard area measure on  $X$ . The  $L^2$  norm of a function  $f$  on  $X$  with respect to the measure  $\mu$  is defined by  $\|f\|_{L^2(X)} = \int_X f(x)d\mu(x)$ . Equipped with a metric  $d : X \times X \rightarrow \mathbb{R}_+$ , the pair  $(X, d)$  forms a *metric space* and the triplet  $(X, d, \mu)$  a *metric measure space*.

### 2.1 Diffusion kernels

A function  $k : X \times X \rightarrow \mathbb{R}$  is called a *diffusion kernel* if it satisfies the following properties:

- (K1) *Non-negativity*:  $k(x, x) \geq 0$ .
- (K2) *Symmetry*:  $k(x, y) = k(y, x)$ .
- (K3) *Positive-semidefiniteness*: for every bounded  $f$ ,

$$\int \int k(x, y) f(x) f(y) d\mu(x) d\mu(y) \geq 0.$$

- (K4) *Square integrability*:  $\int \int k^2(x, y) d\mu(x) d\mu(y) < \infty$ .
- (K5) *Conservation*:  $\int k(x, y) d\mu(y) = 1$ .

The value of  $k(x, y)$  can be interpreted as a transition probability from  $x$  to  $y$  by one step of a random walk on  $X$ .

Diffusion kernel defines a linear operator

$$\mathbf{K}f = \int k(x, y) f(y) d\mu(y), \quad (1)$$

which is known to be self-adjoint. Because of (K4),  $\mathbf{K}$  has a finite Hilbert norm and therefore is compact.

• The authors are with the Department of Computer Science, Technion – Israel Institute of Technology, Haifa, 32000, Israel.  
Email: bronstein@ieee.org, alexbron@ieee.org.

As the result, it admits a discrete eigendecomposition  $\mathbf{K}\psi_i = \alpha_i\psi_i$  with eigenfunctions  $\{\psi_i\}_{i=0}^{\infty}$  and eigenvalues  $\{\alpha_i\}_{i=0}^{\infty}$ .  $\alpha_i \geq 0$  by virtue of property (K3), and  $\alpha_i \leq 1$ , by virtue of (K5) and consequence of the Perron-Frobenis theorem.

By the spectral theorem, the diffusion kernel admits the following *spectral decomposition* property

$$k(x, y) = \sum_{i=0}^{\infty} \alpha_i \psi_i(x) \psi_i(y). \quad (2)$$

Since  $\psi_i$  form an orthonormal basis of  $L^2(X)$ ,

$$\int \int k^2(x, y) d\mu(x) d\mu(y) = \sum_{i=0}^{\infty} \alpha_i^2, \quad (3)$$

a fact sometimes referred to as Parseval's theorem. Using these results, properties (K3-5) can be rewritten in the spectral form as  $0 \leq \alpha_i \leq 1$  and  $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$ .

An important property of diffusion operators is the fact that for every  $t \geq 0$ , the operator  $\mathbf{K}^t$  is also a diffusion operator with the eigenbasis of  $\mathbf{K}$  and corresponding eigenvalues  $\{\alpha_i^t\}_{i=0}^{\infty}$ . The kernel of  $\mathbf{K}^t$  expresses the transition probability by random walk of  $t$  steps. This allows to define a scale space of kernels,  $\{k_t(x, y)\}_{t \in T}$ , with the scale parameter  $t$ .

## 2.2 Heat diffusion

There exists a large variety of possibilities to define a diffusion kernel and the related diffusion operator. Here, we restrict our attention to operators describing *heat diffusion*. Heat diffusion on surfaces is governed by the *heat equation*,

$$\left( \Delta_X + \frac{\partial}{\partial t} \right) u = 0, \quad (4)$$

where  $u$  is the distribution of heat on the surface and  $\Delta_X$  is the positive-semidefinite *Laplace-Beltrami operator*, a generalization of the Laplacian to non-Euclidean domains. For compact surfaces, the Laplace-Beltrami operator has discrete eigendecomposition of the form  $\Delta_X \phi_i = \lambda_i \phi_i$ .

Diffusion operators associated with heat propagation processes are diagonalized by the eigenbasis of the Laplace-Beltrami operator, namely  $\mathbf{K}$ 's having  $\psi_i = \phi_i$ . The corresponding diffusion kernels can be expressed as

$$k(x, y) = \sum_{i=0}^{\infty} K(\lambda_i) \phi_i(x) \phi_i(y), \quad (5)$$

where  $K(\lambda)$  is some function such that  $\alpha_i = K(\lambda_i)$ . Since the Laplace-Beltrami eigenvalues can be interpreted as

*frequency*,  $K(\lambda)$  can be thought of as the *transfer function* of a low-pass filter.<sup>1</sup> Using this signal processing analogy, the kernel  $k(x, y)$  can be interpreted as the point spread function at a point  $y$ , and the action of the diffusion operator  $\mathbf{K}f$  on a function  $f$  on  $X$  can be thought of as the application of the point spread function by means of a non shift-invariant version of convolution. The transfer function of the diffusion operator  $\mathbf{K}^t$  is  $K^t(\lambda)$ , which can be interpreted as multiple applications of the filter  $K(\lambda)$ . Such multiple applications decrease the effective bandwidth of the filter and, consequently, increase its effective support in space. Because of this duality, we will freely interchange between  $k(x, y)$  and  $K(\lambda)$  referring to both as kernels.

An important particular case of is the *heat operator*  $\mathbf{H}^t$  defined by the family of transfer functions  $H^t(\lambda) = e^{-t\lambda}$  and the associated *heat kernels*  $h_t(x, y) = \sum_{i=0}^{\infty} e^{-t\lambda_i} \phi_i(x) \phi_i(y)$ . The heat kernel  $h_t(x, y)$  is the solution of the heat equation with point heat source at  $x$  at time  $t = 0$ , i.e., the heat value at point  $y$  after time  $t$ .

## 2.3 Diffusion distances

Since a diffusion kernel  $k(x, y)$  measures the degree of proximity between  $x$  and  $y$ , it can be used to define a metric

$$d^2(x, y) = \|k(x, \cdot) - k(y, \cdot)\|_{L^2(X)}^2, \quad (6)$$

on  $X$ , dubbed the *diffusion distance* by Coifman and Lafon [9], [12]. Another way to interpret the latter distance is by considering the embedding  $\Psi : x \mapsto L^2(X)$  by which each point  $x$  on  $X$  is mapped to the function  $\Psi(x) = k(x, \cdot)$ . The embedding  $\Psi$  is an isometry between  $X$  equipped with diffusion distance and  $L^2(X)$  equipped with the standard  $L^2$  metric, since

$$d(x, y) = \|\Psi(x) - \Psi(y)\|_{L^2(X)}. \quad (7)$$

Because of spectral duality, the diffusion distance can also be written as

$$d^2(x, y) = \sum_{i=0}^{\infty} K^2(\lambda_i) (\phi_i(x) - \phi_i(y))^2. \quad (8)$$

Here as well we can define an isometric embedding  $\Phi : x \mapsto \ell^2$  with  $\Phi(x) = \{K(\lambda_i) \phi_i(x)\}_{i=0}^{\infty}$ , termed as the *diffusion map* by Lafon. The diffusion distance can be cast as  $d(x, y) = \|\Phi(x) - \Phi(y)\|_{\ell^2}$ .

1. For an insightful discussion of the physical interpretation of the Laplace-Beltrami operator, the reader is referred to [13].

The same way a diffusion operator  $\mathbf{K}^t$  defines a scale space, a family of diffusion metrics can be defined for  $t \geq 0$  as

$$\begin{aligned} d_t^2(x, y) &= \|\Phi_t(x) - \Phi_t(y)\|_{\ell_2}^2 \\ &= \sum_{i=0}^{\infty} K^{2t}(\lambda_i)(\phi_i(x) - \phi_i(y))^2, \end{aligned} \quad (9)$$

where  $\Phi_t(x) = \{K^t(\lambda_i)\phi_i(x)\}_{i=0}^{\infty}$ . Interpreting diffusion processes as random walks,  $d_t$  can be related to the ‘‘connectivity’’ of points  $x$  and  $y$  by walks of length  $t$  (the more such walks exist, the smaller is the distance).

## 2.4 Distance distributions

Though diffusion metrics contain significant amount of information about the geometry of the underlying shape, direct comparison of metrics is problematic since it requires computation of correspondence between shapes. A common way to circumvent the need of correspondence is by representing a metric by its distribution, and measuring the similarity of two shapes by comparing the distributions of the respective metrics.

A metric  $d$  on  $X$  naturally pushes forward the product measure  $\mu \times \mu$  on  $X \times X$  (i.e., the measure defined by  $d(\mu \times \mu)(x, y) = d\mu(x)d\mu(y)$ ) to the measure  $F = d_*(\mu \times \mu)$  on  $[0, \infty)$  defined as  $F(I) = (\mu \times \mu)(\{(x, y) : d(x, y) \in I\})$  for every measurable set  $I \subset [0, \infty)$ .  $F$  can be fully described by means of a cumulative distribution function, denoted by

$$F(\delta) = \int_0^{\delta} dP = \int \chi_{d(x,y) \leq \delta} d\mu(x)d\mu(y) \quad (10)$$

with some abuse of notation (here  $\chi$  is the indicator function).  $F(\delta)$  defined this way is the measure of pairs of points the distance between which is no larger than  $\delta$ ;  $F(\infty) = \mu^2(X)$  is the squared area of the surface  $X$ . The density function (histogram) can be defined as the derivative  $f(\delta) = \frac{d}{d\delta}F(\delta)$ . Sometimes, it is convenient to work with normalized distributions,  $\hat{F} = F/F(\infty)$  and the corresponding density functions,  $\hat{f}$ , which can be interpreted as probabilities.

Using this idea, comparison of two metric measure spaces reduces to the comparison of measures on  $[0, \infty)$ , or equivalently, comparison of un-normalized or normalized distributions, which is carried out using one of the standard distribution dissimilarity criteria used in statistics:

$L_p$  **dissimilarity:** for  $1 \leq p < \infty$ ,

$$D_{L_p}(F, G) = \left( \int_0^{\infty} |f - g|^p d\delta \right)^{1/p}; \quad (11)$$

and for  $p = \infty$ ,

$$D_{L_{\infty}}(F, G) = \sup_{\delta \in [0, \infty)} |f - g|. \quad (12)$$

**Kullback-Leibler dissimilarity:**

$$D_{\text{KL}}(F, g) = \int_0^{\infty} \left( f \log \frac{f}{g} + g \log \frac{g}{f} \right) d\delta. \quad (13)$$

**Bhattacharyya dissimilarity:**

$$D_{\text{Bhatt}}(F, G) = -\log \int_0^{\infty} \sqrt{fg} d\delta. \quad (14)$$

$\chi^2$  **dissimilarity:**

$$D_{\chi^2}(F, G) = 2 \int_0^{\infty} \frac{(f - g)^2}{f + g} d\delta. \quad (15)$$

**Earth mover’s distance (EMD)**, which for normalized univariate distributions is known to be equivalent to

$$D_{\text{EMD}}(\hat{F}, \hat{G}) = \int_0^{\infty} |\hat{F} - \hat{G}| d\delta. \quad (16)$$

## 3 SPECTRAL SHAPE DISTANCES

In this paper, we propose the *spectral distance* distributions framework for shape similarity, which will be shown to generalize several state-of-the-art approaches. Figure 1 schematically describes the proposed data flow. Given two shapes, the eigendecomposition of their Laplace-Beltrami operators is performed. The eigenvalues and eigenfunctions are used to define a diffusion kernel  $k$ , which in turn, is used to define a family of scale-dependent diffusion distances between points on the shapes. Next, the distributions of pair-wise distances are computed. The dissimilarity of two shapes is thus converted into dissimilarity of distributions. Multiple scales are merged into a single *spectral shape distance* by means of an aggregation operation, which can be performed after either of the above stages (possible locations of the integration operator are denoted in red in the figure). We individuate four main components in this framework:

**Kernel:** According to (5), the diffusion kernel  $k(x, y)$  is described by the transfer function  $K(\lambda)$ . The dependence on the scale parameter is introduced by taking the powers  $K^t$ . The kernel can be selected to satisfy certain invariance properties, as detailed in the next section.

**Norm:** A generalized diffusion distance  $d_t(x, y)$  is defined as a cross-talk between the diffusion kernels,  $d_t(x, y) = \|k_t(x, \cdot) - k_t(y, \cdot)\|$ , using some norm. For the choice of  $L^2(X)$ , the diffusion distance has an explicit expression according to (9). It is remarkable that some selections of transfer functions may result in kernels

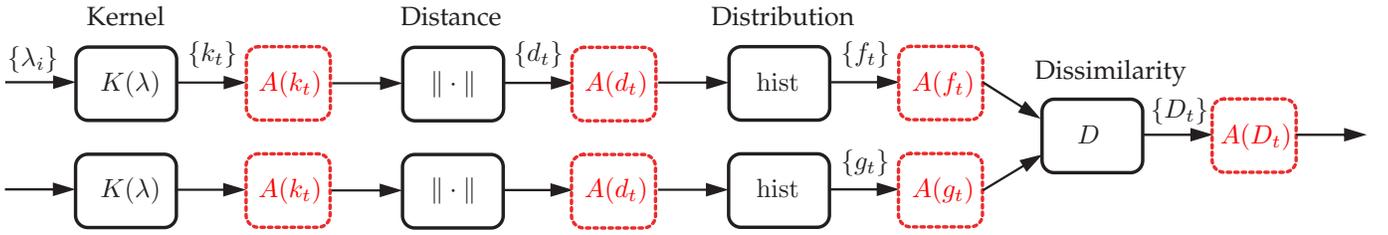


Fig. 1. Data flow in the proposed approach: for each of the compared shapes, a transfer function  $K(\lambda)$  and the norm  $\|\cdot\|$  define a family of pair-wise distances. Shape dissimilarity is cast as distribution dissimilarity using the criterion  $D$ . Aggregation of different scales can be performed at any of these stages as denoted in dashed red.

violating properties of (K1–5) and lead to distances that are not metrics. Comparison of distributions of such distances may still be very useful.

**Dissimilarity**  $D(F, G)$  is used to compare the distance distributions, as detailed in Section 2.4.

**Aggregation:** the specific way to aggregate the scales and its position in the pipeline, as detailed in Section 3.2.

### 3.1 Invariance

The described framework is very generic and different choices of the transfer function  $K$  and the scale integration method lead to different shape similarity criteria. In practical applications, specific choices are driven by desired invariance considerations: it is possible to design  $K$  in such a way that it is invariant to certain classes of shape transformation.

**Deformation invariance:** Since the Laplace-Beltrami operator is intrinsic (i.e., fully expressible in terms of the Riemannian metric of the surface), its eigenfunctions and eigenvalues are invariant to inelastic bendings. As a result, diffusion kernels based on  $\phi_i, \lambda_i$  are deformation-invariant, and so are distributions of the related distances.

**Topological invariance:** In many practical applications, shapes suffer from “topological noise”, resulting in different local connectivity. This is a typical situation when shapes are scanned by a 3D range camera. The change in connectivity alters the diffusion distances by introducing “shortcuts”; however, these artifacts have different effect at different scale. For large  $t$ , the influence of the shortcuts is smaller.

**Scale invariance:** When a shape  $X'$  is obtained by uniformly scaling  $X$  by a factor  $\alpha > 0$ , the eigenvalues of the Laplace-Beltrami operator are scaled according to  $\lambda'_i = \alpha^{-2}\lambda_i$ , and the corresponding  $L^2$ -normalized eigenfunctions become  $\phi'_i(x) = \alpha^{-1}\phi_i(x)$ . A diffusion kernel defined by a transfer function  $K(\lambda)$  therefore

becomes

$$k'(x, y) = \frac{1}{\alpha^2} \sum_{i=0}^{\infty} K(\lambda/\alpha^2) \phi_i(x) \phi_i(y). \quad (17)$$

When, for example, the  $L^p$  norm is used, the corresponding diffusion distance is given by

$$\begin{aligned} d'(x, y) &= \int |k'(x, z) - k'(y, z)|^p d\mu'(z) \\ &= \int \left| \sum_{i=0}^{\infty} \alpha^{2/p-2} K(\lambda/\alpha^2) \phi_i(z) (\phi_i(x) - \phi_i(y)) \right|^p d\mu(z). \end{aligned} \quad (18)$$

In order to obtain a scale-invariant distance, the transfer function  $K(\lambda)$  must satisfy  $K(\lambda/\alpha^2) = \alpha^{2-2/p}K(\lambda)$ . Functions satisfying this property are  $K(\lambda) = \lambda^{-1/p}$ .

In the particular case  $p = 2$ , the kernel  $K(\lambda) = \lambda^{-1/2}$  defines a distance

$$\begin{aligned} d_{\text{CT}}^2(x, y) &= \int \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \phi_i^2(z) (\phi_i(x) - \phi_i(y))^2 d\mu(z) \\ &= \sum_{i=1}^{\infty} \frac{1}{\lambda_i} (\phi_i(x) - \phi_i(y))^2 \end{aligned} \quad (19)$$

(note that the singularity at  $\lambda_0 = 0$  can be removed starting the summation from  $i = 1$ , since  $\phi_0 = \text{const}$ ). Such a distance is known in spectral graph theory as the *commute time distance*, and can be thought of as the average time of a walk starting at  $x$  to go through  $y$  and return back to  $x$  (called the commute time, hence the name). Invoking the relation  $\int_0^{\infty} e^{-\lambda t} dt = \frac{1}{\lambda}$ , the commute time distance can be interpreted as the integral of the squared heat diffusion distance  $d_t^2$  over the entire time scale,

$$\begin{aligned} \int_0^{\infty} d_t^2(x, y) &= \sum_{i=0}^{\infty} \int_0^{\infty} e^{-2\lambda_i t} dt (\phi_i(x) - \phi_i(y))^2 \\ &= \sum_{i=0}^{\infty} \frac{1}{2\lambda_i} (\phi_i(x) - \phi_i(y))^2 = \frac{1}{2} d_{\text{CT}}^2(x, y). \end{aligned} \quad (20)$$

Equivalently, the *commute time kernel*  $K_{\text{CT}}(\lambda) = \lambda^{-1/2}$  can be thought of as the integral of the heat kernel,  $K_{\text{CT}}^2(\lambda) = \int_0^{\infty} H_t(\lambda) dt$ .

### 3.2 Scale aggregation

The derivation of the commute time kernel and the commute time distance from the corresponding heat kernel and diffusion distance can be interpreted as the application of a *scale aggregation* functional on  $\{H_t(\lambda)\}_{t>0}$  or  $\{d_t(x, y)\}_{t>0}$ . Examples of such aggregation functionals are

**Sum** or, more generally, **weighted sum**  $A(f_t) = \int_T w(t) f_t dt$  where  $w$  is some scale-dependent weighting function and  $T$  is some range of scales.

**Sum of squares** or, more generally, **sum of nonlinearities**  $A(f_t) = \int_T \xi(f_t) dt$ , where  $\xi$  is a scalar non-linearity.

**Maximum:**  $A(f_t) = \sup_{t \in T} f_t$ .

Here,  $f_t$  denotes a family of scale-dependent objects. Depending on the stage at which aggregation is performed, aggregation functionals can be classified as

**Kernel aggregation:**  $A$  acting on  $\{K^t(\lambda)\}$  or  $\{k_t(x, y)\}$ .

**Distance aggregation:**  $A$  acting on  $\{d_t(x, y)\}$ .

**Distribution aggregation:**  $A$  acting on  $\{F_t\}$  or  $\{f_t\}$ .

**Dissimilarity aggregation:**  $A$  acting on  $\{D(F_t, G_t)\}$ .

This resembles the spirit of the aggregation of multi-scale dissimilarities of histograms-based descriptors proposed in [14].

### 3.3 Relation to other methods

**Diffusion distance distributions:** In [15], Mahmoudi and Sapiro proposed applying the shape distribution approach to heat diffusion distances. Given two shapes  $X$  and  $Y$  with distribution densities  $f_X$  and  $f_Y$  of the corresponding diffusion distances  $d_{X,t}, d_{Y,t}$ , Mahmoudi and Sapiro defined a similarity criterion as  $d_{MS,t}(X, Y) = \|f_X - f_Y\|_2$ . This criterion can be regarded as a particular case of our spectral shape distance, with the heat kernel  $K^t(\lambda) = e^{-\lambda t}$ ,  $L^2(X)$  norm, and the  $L_2$  distribution dissimilarity. The diffusion distances are taken at some single fixed scale  $t$ ; no scale aggregation is performed. Choosing different  $t$  allows describing small- and large-scale properties of the shape (for example, two shapes can be similar at small scale and dissimilar on large scales or vice versa). At the same time, the scale depends on the shape size and the choice of  $t$  can often be problematic [20].

**Diffusion scale-space distance:** In [7], we proposed a way to address the problem of scale selection by integrating the Mahmoudi-Sapiro distance on a set of scales, resulting in what was termed the *diffusion scale-space distance*,  $d_{DSS}(X, Y) = \sum_{t \in T} d_{MS,t}(X, Y)$ , where  $T$

denotes some set of time scale. Such a distance has the advantage of comparing shape properties at multiple scales, considering each scale separately.

$d_{DSS}$  is also spectral shape distance with the heat kernel  $K^t(\lambda) = e^{-\lambda t}$ , diffusion distance defined with the  $L^2(X)$  norm, and  $L_2$  distribution dissimilarity. The aggregation is performed on distribution dissimilarities using sum  $\sum_{t \in T}$  over a finite set of scales  $T$ .

**Global point signature:** In [23], Rustamov proposed the *global point signatures* (GPS) representation for shapes based on the eigenmap of the form  $\Psi_X(x) = \{\lambda_i^{-1/2} \phi_i(x)\}_{i=1}^\infty$ . Given two shapes  $X, Y$ , Rustamov looked at the corresponding  $\Psi_X$  and  $\Psi_Y$  as Euclidean objects with distances defined by  $d_X(x, y) = \|\Psi_X(x) - \Psi_X(y)\|_{L^2(X)}$  (similarly for  $d_Y$ ), and computed the distance between the shapes  $X, Y$  as the  $L_2$  dissimilarity of the distributions of the distances  $d_X, d_Y$ ,  $d_{GPS}(X, Y) = \|f_X - f_Y\|_2$ . This method can be considered as an application of the shape distributions approach [17] to the embeddings  $\Psi_X, \Psi_Y$ .

The GPS distance can be cast as a particular setting of the spectral shape distance with the commute time kernel  $K(\lambda) = \lambda^{-1/2}$ ,  $L^2(X)$  norm,  $L_2$  distribution dissimilarity, and no scale aggregation, or, alternatively, as a the spectral distance with the heat kernel,  $L^2(X)$  norm, and aggregation of diffusion distances using the sum of squares on  $t \in [0, \infty)$ .

Considering all the three aforementioned methods from the standpoint of our framework, we can clearly see their equivalence up to position of the scale integration operation. In  $d_{MS,t}$ , the heat diffusion distance is computed at a single scale  $t$ , then distribution distance is computed, without doing any scale integration. In  $d_{GPS}$ , integration is performed *before* computing the distribution. In  $d_{DSS}$ , integration is performed *after* computing the distribution distances.

**Scale invariant heat kernel signature:** In [8], it was observed that using the logarithmic scale  $\tau = \log t$ , the heat kernel on  $X$  and its counterpart on  $X'$  uniformly scaled by a factor  $\alpha > 0$  are related by  $\log k'_\tau = \log k_{\tau+\beta} + \beta$ , where  $\beta = -2 \log \alpha$ . Taking the derivative of  $\log k_\tau$  with respect to  $\tau$  undoes the additive constant  $\beta$ , while the shift in  $\tau$  can be undone in the Fourier domain by canceling the phase. The resulting family of kernels,

$$\hat{k}_\tau(x, y) = \mathcal{F}^{-1} \left( \left| \mathcal{F} \left( \frac{\partial}{\partial \tau} \log k_\tau(x, y) \right) \right| \right), \quad (21)$$

is scale invariant. From the point of view of the proposed framework, the above transformation can be thought of

as a scale aggregation functional acting on the kernel. It can be further followed by selection of a single value of  $\tau$  or aggregation of several scales in the previously discussed way. Though the kernel  $\hat{k}_\tau$  is not a valid diffusion kernel (e.g., it can assume negative values), the resulting “distance” is still informative. We intend to investigate this interesting relation in future studies.

## 4 NUMERICAL IMPLEMENTATION

For numerical computation, the shape  $X$  is represented by a finite number of samples  $\hat{X} = \{x_1, \dots, x_N\} \subseteq X$ . The discrete approximation of the Laplace-Beltrami operator has the following generic form

$$(\Delta_{\hat{X}}u)_i = \frac{1}{a_i} \sum_j w_{ij}(u_i - u_j), \quad (22)$$

where  $u = (u_1, \dots, u_N)$  is a scalar function defined on  $\hat{X}$ ,  $w_{ij}$  are weights, and  $a_i$  are normalization coefficients. In matrix notation, Equation (22) can be written as  $\Delta_{\hat{X}}u = A^{-1}Lf$ , where  $A = \text{diag}(a_i)$  and  $L = \text{diag}\left(\sum_{l \neq i} w_{il}\right) - (w_{ij})$  are sparse matrices. The eigenvalues and eigenfunctions of the discretized Laplace-Beltrami operator are computed by solving the *generalized eigendecomposition* problem  $A\Phi = \Lambda L\Phi$ , where  $\Lambda$  is the  $(k+1) \times (k+1)$  diagonal matrix of the smallest eigenvalues  $\lambda_0, \dots, \lambda_k$ , and  $\Phi$  is an  $N \times (k+1)$  matrix of corresponding eigenfunctions [13].

**Discrete Laplace-Beltrami operator.** Different discretizations of the Laplace-Beltrami operator lead to different choice of  $A$  and  $W$  [21]. For triangular meshes, a popular choice is the *cotangent weight* scheme [19], in which  $w_{ij} = \cot \alpha_{ij} + \cot \beta_{ij}$  ( $\alpha_{ij}$  and  $\beta_{ij}$  are the two angles opposite to the edge between vertices  $i$  and  $j$  in the two triangles sharing the edge) for  $j$  in the 1-ring neighborhood of vertex  $i$  and zero otherwise, and  $a_i$  is one third of the area of the triangles sharing the vertex  $x_i$ . This discretization preserves many important properties of the continuous Laplace-Beltrami operator, such as positive semi-definiteness, symmetry, and locality, and in addition it is numerically consistent. For shapes represented as point clouds, the Laplace-Beltrami operator can be approximated using [2].

**Approximation quality and complexity.** In practice, numerical inaccuracy due to the discretization of the Laplace-Beltrami operator imposes a limit on the number  $k$  of eigenvalues and eigenvectors that can be accurately computed. Typical complexity of computing  $k$  first eigenvectors of a sparse symmetric matrix is  $\mathcal{O}(Nk)$  with

the Arnoldi algorithm used in MATLAB function `eigs`. This should be taken into consideration when choosing the transfer function  $K(\lambda)$ : ideally, it should decay fast enough to allow taking a small number of eigenvalues, thus reducing the computation complexity. Specifically in the particular cases discussed in Section 3.3, numerical approximation of the commute time distance (19) is less favorable compared to the heat diffusion distance, since the decay of the terms  $e^{-\lambda t}$  is faster than  $\lambda^{-1/2}$  for typical choice of  $t$ .

## 5 RESULTS

We used a subset of 486 shapes from the Shape-Google database [18], containing different shapes with simulated transformations. The transformations include null (no transformation), isometry (near-isometric bending), topology (connectivity change obtained by welding some of the shape vertices), isometry+topology (combination of the previous two), triangulation (different meshing of the same shape), and partiality (missing information, obtained by making holes and cutting parts off the shape). Multiple instances of each transformation are present for each shape class. We performed a leave-one-out retrieval experiment, in which each class of transformations was queried against a database containing the remaining transformations. Matches were considered correct between different transformations of the same shape. The *average precision* (area under the precision-recall curve) was measured for each query shape, and the *mean average precision* (mAP) was used as a single number quantifying the retrieval performance for each class of transformations. As an additional quality measure, for each class of transformations the *receiver operating characteristic* (ROC) curve was computed, representing a tradeoff between the percentage of similar shapes correctly identified as similar (*true positives rate* or TPR) and the percentage of dissimilar shapes wrongfully identified as similar (*false positive rate* or FPR). For additional details, the reader is referred to [18].

**Experiments.** We performed several experiments to reveal the influence of the choice of the different processing stages in Figure 1. While these experiments do not attempt to entirely cover the entire space of possible choices, we believe they provide a sufficiently interesting insight about the importance of each ingredient. We denote each choice as *Kernel-Norm-Dissimilarity-Aggregation*, according to Section 3. As *Kernel*, we used  $K(\lambda) = e^{-\lambda}$  (*heat kernel*, abbreviated as H),  $\lambda^{-1/2}$

TABLE 1  
mAP (%) of Heat- $L_2$ -Dissimilarity-Single for different distribution dissimilarities.

Transform.	$L_1$	$L_2$	$nL_1$	$nL_2$	K-L	Bhatt	$\chi^2$	EMD
Null	<b>96.19</b>	95.59	<b>96.19</b>	95.82	93.96	94.81	95.10	93.83
Isometry	<b>94.62</b>	93.27	<b>94.62</b>	94.07	91.59	92.71	93.42	90.56
Iso+Topo	<b>89.43</b>	87.32	<b>89.43</b>	87.56	85.73	86.89	87.31	85.13
Partiality	90.49	90.92	90.48	<b>91.38</b>	84.74	88.62	90.06	78.78
Topology	83.40	<b>83.82</b>	83.39	83.73	65.80	68.57	70.07	69.31
Triang.	92.54	91.56	<b>92.55</b>	91.81	90.38	91.04	91.23	84.68
Noise	89.14	89.26	89.16	<b>89.34</b>	85.03	86.67	87.45	86.06
All	90.93	90.22	<b>90.94</b>	90.56	86.12	87.86	88.62	84.23

(commute time kernel, abbreviated as CT), and the Step function. As Norm, we used  $L_1$ ,  $L_2$ , and  $L_\infty$ . The Kernel-Norm pair fully defines the spectral distance used in the distribution computation. For reference, we also show retrieval using distributions of geodesic distances (denoted as Geo) as proposed in [10]. As Dissimilarity, we used the  $L_1$ ,  $L_2$ , normalized  $L_1$  ( $nL_1$ ), normalized  $L_2$  ( $nL_2$ ), KL, Bhattacharyya,  $\chi^2$ , and EMD, as detailed in Section 2.4. For scale Aggregation, we tested a few simple choices applied at the last stage aggregating distribution dissimilarities at different scales, using Single scale selection, Sum, Sum<sup>2</sup>, and Max, as detailed in Section 3.2.

In all experiments, distributions were computed in the interval from 0 to the 95-percentile of the distances on the entire corpus of shapes. The Parzen window technique was used to produce the histograms with the window bandwidth set to  $\frac{1}{256}$  of the interval length. The histograms were sampled on 256 evenly spaced bins.

**Dissimilarity choice.** In this experiment, we evaluated the choice of different dissimilarity functions used to compare between distributions. The Heat- $L_2$  spectral distance was fixed with a single time scale  $t = 1024$ , evaluating the retrieval performance of different choices of the form Heat- $L_2$ -Dissimilarity-Single. Table 1 summarizes the mean average precision of each dissimilarity function in different transformation classes. While there is no absolute winner in all classes, the normalized  $L_1$  dissimilarity performs the best in four out of seven classes, close to the best in other three, and wins in overall performance. This behavior is observed consistently for other choices of spectral distances. Based on this observation, we use the normalized  $L_1$  dissimilarity in all further experiments.

**Kernel-Norm choice.** In this experiment, we fixed the distribution dissimilarity to  $nL_1$  and evaluated the performance of different spectral distances. We evaluated the performance of Kernel-Normal- $nL_1$ -Single with the

TABLE 2

mAP (%) of Kernel-Normal- $nL_1$  for different spectral distances.

Transform.	H- $L_2$	H- $L_1$	H- $L_\infty$	CT- $L_2$	Step- $L_2$	Geo
Null	96.19	95.41	94.61	94.41	<b>96.35</b>	94.85
Isometry	94.62	92.84	90.31	91.89	<b>95.18</b>	90.98
Iso+Topo	89.43	91.48	87.17	83.69	<b>93.41</b>	84.88
Partiality	<b>90.48</b>	82.27	82.85	78.05	87.68	84.29
Topology	83.39	98.29	85.73	86.65	<b>98.92</b>	96.24
Triang.	92.55	91.59	91.57	89.78	<b>92.93</b>	90.14
Noise	89.16	84.68	81.70	<b>92.21</b>	82.53	85.33
All	90.94	89.87	87.17	87.29	<b>91.65</b>	88.32
Scale	30.55	15.58	28.28	<b>100.00</b>	17.27	31.07

TABLE 3

mAP (%) of H- $L_2$ - $nL_1$ -Aggregation, for different aggregation functionals.

Transform.	$t = 1024$	$t = 4096$	Sum	Sum <sup>2</sup>	Max
Null	96.19	93.07	96.42	<b>96.54</b>	96.51
Isometry	<b>94.62</b>	89.06	93.90	94.06	94.24
Iso+Topo	89.43	82.42	89.03	89.20	<b>91.75</b>
Partiality	<b>90.48</b>	80.22	88.75	89.06	90.37
Topology	83.39	65.49	83.37	87.03	<b>98.23</b>
Triang.	92.55	84.43	92.70	92.83	<b>92.91</b>
Noise	89.16	86.48	92.31	92.43	<b>92.45</b>
All	90.94	83.37	90.90	91.37	<b>93.05</b>

heat kernel and  $L_1$ ,  $L_2$  and  $L_\infty$  norms, the commute time and step kernels with the  $L_2$  norm. For comparison, the performance of Geo- $nL_1$ -Single is given. Since geodesic distances are known to be sensitive to large-scale changes in topology, it is usually helpful to restrict the distances within a certain radius. The reported results were obtained with  $d \leq 100$  which maximized performance. Table 2 summarizes the measured mAP broken down by transformation class. Step kernel with the  $L_2$  norm performs the best in five out of seven classes and achieves the highest overall performance. The heat kernel with the  $L_\infty$  norm achieves the lowest overall performance followed by the commute time kernel and geodesic distances. As another reference,

we show the mAP of the compared distances under global scale transformations ranging from 125% to 200%. Poor performance is obtained by all methods with the exception of commute time kernel that scores 100% mAP. This gives an experimental evidence to scale invariance of commute time distances and lack of scale invariance of geodesic and diffusion distances.

**Aggregation choice.** In this experiment, we evaluated the importance of aggregation of distribution dissimilarities at different scales. The  $nL_1$  dissimilarity of the Heat- $L_2$  spectral distance distribution was computed at six time scales  $t \in \{1024, 1351, 1782, 2352, 3104, 4096\}$ . The performance of Heat- $L_2$ - $nL_1$ -Aggregation was evaluated for two single scales  $t = 1024$  and  $4096$  as well as for the sum, sum of squares, and maximum aggregation functions. Table 3 summarizes the results. We conclude that using the maximum function achieves the best overall mAP. An even bigger gain is seen from the ROC curves in Figure 2, where performance is evaluated in terms of false positive and negative rates.

## 6 CONCLUSION

We presented a generic framework for computing the similarity of non-rigid shapes based on the distributions of diffusion distances. Our framework allows building in explicit invariance properties, and also generalizes previous approaches due to Rustamov [23], Mahmoudi and Sapiro [15], and [7]; the presented formulation brings insights on the relations and equivalence of these methods. An important question left undiscussed in this paper is the construction of an *optimal* task-specific kernel  $K(\lambda)$ . We plan to address this problem in follow-up works.

## REFERENCES

- [1] M. Belkin and P. Niyogi, *Laplacian eigenmaps for dimensionality reduction and data representation*, Neural Computation **13** (2003), 1373–1396.
- [2] M. Belkin, J. Sun, and Y. Wang, *Discrete laplace operator on meshed surfaces*, Proc. Symp. Computational Geometry, 2009, pp. 278–287.
- [3] A. Bronstein, M. Bronstein, M. Ovsjanikov, and L. Guibas, *Shape google: a computer vision approach to invariant shape retrieval*, Proc. NORDIA, 2009.
- [4] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, *Numerical geometry of nonrigid shapes*, Springer, 2008.
- [5] ———, *Topology-invariant similarity of nonrigid shapes*, Int’l J. Computer Vision (IJCV) (2008), to appear.
- [6] A. M. Bronstein, M. M. Bronstein, M. Mahmoudi, R. Kimmel, and G. Sapiro, *A Gromov-Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching*, IJCV (2010).
- [7] M. M. Bronstein and A. M. Bronstein, *On a relation between shape recognition algorithms based on distributions of distances*, CIS Technical Report 2009-14, Technion, 2009.
- [8] M. M. Bronstein and I. Kokkinos, *Scale-invariant heat kernel signatures for non-rigid shape recognition*, Proc. CVPR, 2010.
- [9] R. R. Coifman and S. Lafon, *Diffusion maps*, Applied and Computational Harmonic Analysis **21** (2006), 5–30.
- [10] A. B. Hamza and H. Krim, *Geodesic matching of triangulated surfaces*, IEEE Trans. Imag. Proc. **15** (2006), no. 8, 2249.
- [11] A. Ion, N.M. Artner, G. Peyré, W.G. Kropatsch, and L.D. Cohen, *Matching 2D and 3D Articulated Shapes using Eccentricity*, (2009).
- [12] S. Lafon, *Diffusion Maps and Geometric Harmonics*, Ph.D. dissertation, Yale University (2004).
- [13] B. Lévy, *Laplace-Beltrami eigenfunctions towards an algorithm that “understands” geometry*, Int’l Conf. Shape Modeling and Applications, 2006.
- [14] H. Ling and K. Okada, *Diffusion distance for histogram comparison*, Proc. CVPR, 2006.
- [15] M. Mahmoudi and G. Sapiro, *Three-dimensional point cloud recognition via distributions of geometric distances*, Graphical Models **71** (2009), no. 1, 22–31.
- [16] F. Mémoli, *Spectral Gromov-Wasserstein distances for shape matching*, Proc. NORDIA, 2009.
- [17] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, *Shape distributions*, ACM Transactions on Graphics (TOG) **21** (2002), no. 4, 807–832.
- [18] M. Ovsjanikov, A.M. Bronstein, M.M. Bronstein, and L.J. Guibas, *Shape Google: a computer vision approach to invariant shape retrieval*, Proc. NORDIA, 2009.
- [19] U. Pinkall and K. Polthier, *Computing discrete minimal surfaces and their conjugates*, Experimental mathematics **2** (1993), no. 1, 15–36.
- [20] H. Qiu and ER Hancock, *Clustering and embedding using commute times*, Trans. PAMI **29** (2007), no. 11, 1873–1890.
- [21] M. Reuter, S. Biasotti, D. Giorgi, G. Patanè, and M. Spagnuolo, *Discrete Laplace-Beltrami operators for shape analysis and segmentation*, Computers & Graphics **33** (2009), no. 3, 381–390.
- [22] M.R. Ruggeri and D. Saupe, *Isometry-invariant matching of point set surfaces*, Proc. 3DOR, 2008.
- [23] R. M. Rustamov, *Laplace-Beltrami eigenfunctions for deformation invariant shape representation*, Proc. SGP, 2007, pp. 225–233.
- [24] J. Sun, M. Ovsjanikov, and L. Guibas, *A concise and provably informative multi-scale signature based on heat diffusion*, Proc. SGP, 2009.

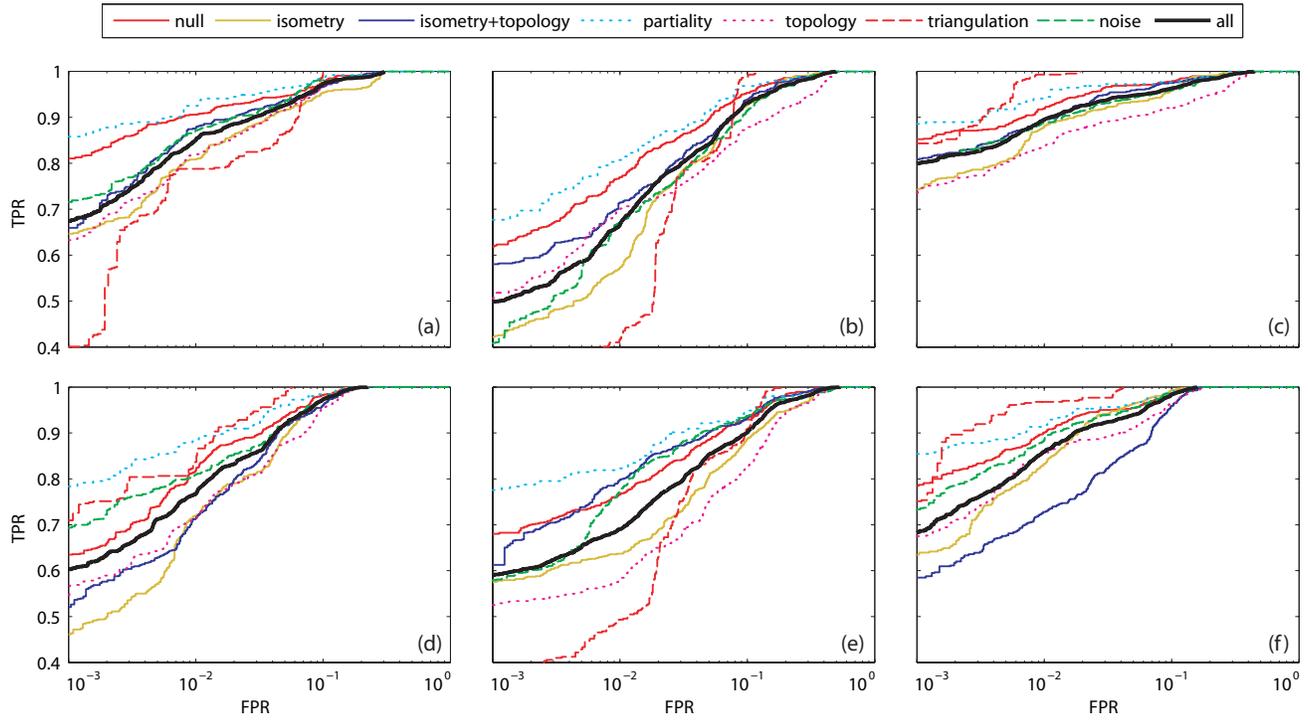


Fig. 2. ROC curves of (a)  $H-L_2-nL_1$ -Single,  $t = 1024$ ; (b)  $H-L_2-nL_1$ -Single,  $t = 4096$ ; (c)  $H-L_2-nL_1$ -Max,  $t \in [1024, 4096]$ ; (d)  $Geo-nL_1$ -Single; (e)  $CT-L_2-nL_1$ -Single; and (F)  $Step-L_2-nL_1$ -Single,  $t = 1024$ .