Graph-Based Descriptor Learning for Non-Rigid 3D Shapes

Yuehan Xiong, Hongkai Xiong
Department of Electronic Engineering, *Shanghai Jiao Tong University*, *Shanghai, China*
{xiongyuehan, xionghongkai}@sjtu.edu.cn

***Abstract*—In this paper, we propose a deformable 3D shape descriptor learning approach that takes into account the spatial correlations among local shape descriptors. By constructing a weighted graph that connects salient points on a shape surface, local feature descriptors on salient points are considered as signals defined on graph vertices, incorporating local surface information into the global graph structure. We then learn a graph structure-aware dictionary for each category of shapes with the multi-graph dictionary learning strategy, capturing similar spectral properties among graph signals generated from different shapes of the same category. Experiments conducted on representative 3D shape benchmark datasets demonstrate that our method improves over the state-of-the-art.**

***Keywords*—**Shape descriptors, sparse coding, graph dictionary learning, non-rigid 3D shapes, spectral graph theory

I. Introduction

The advancement in visualization technologies has given rise to an enormous amount of 3D shapes on the Internet. Moreover, 3D shapes have been widely used in manufacturing, entertainment, medical imaging and virtual reality, etc. Among many challenges in 3D shape analysis, we will focus on the problem of non-rigid 3D shape classification. The increasing availability of 3D shapes makes designing efficient classification approaches an important issue.

Feature descriptors are commonly used in shape recognition tasks. It assigns each point of the shape a vector in some single- or multi-dimensional feature descriptor domain representing local or global geometric properties of that point. Shape descriptors should be discriminative while also invariant to intra-class transformations. The non-rigid shapes are particularly challenging as they include a wide variability of transformations, such as bending and articulated motion.

Following the trends in the image analysis area, where learned representations are favored over hand-crafted ones, learning-based shape descriptors are proposed for the problem of non-rigid shape analysis [1][2][3]. One of the descriptor learning approaches that has been used in shape analysis scenario is dictionary learning and sparse coding [4][5], which is capable of learning a set of over-complete atoms so a feature vector can be represented by a sparse linear combination of dictionary atoms. Dictionaries that take into account the underlying graph structure of graph signals have also been put forward [6][7]. In recent years, a large number of dictionary learning and sparse coding methods have been proposed for various computer vision tasks. In [1], the dictionary and the sparse codes are learned in the supervised regime via bi-level optimization using a task specific objective that promotes invariance desired in the shape retrieval applications.

In this paper, we propose a 3D shape descriptor learning approach that describes the spatial correlations among local descriptors with a weighted graph and employs a graph dictionary learning strategy. We first construct a fully-connected weighted graph for a shape, where each graph vertex is associated with a segmented region. Feature vector is computed for each disjointly segmented region and concatenated into a descriptor matrix. Then each column of the region descriptor matrix can be considered as a signal living on the graph. Assuming that shapes from the same category generate similar graphs and share similar graph signals, a multi-graph dictionary learning method is used to learn a dictionary for each category of shapes. Contrary to previous shape descriptor learning approaches that use sparse coding and dictionary learning while ignoring the underlying structure of feature distribution, our approach encodes the prominent surface structure of a shape with a graph to learn a graph structure-aware dictionary. The framework of the proposed method is presented in Fig. 1.

The remainder of the paper is organized as follows. Section II presents the overall framework and details of the graph-based non-rigid 3D shape descriptor learning approach. Section III provides experiment results on various datasets. Finally, we conclude the paper in Section IV.

II. The Method

In this section we introduce: i) intrinsic graph generation and graph signal computation, ii) solving the correspondence problem for shape regions, and iii) learning class-specific graph dictionaries.

A. Isometry-Invariant Graph Generation

We employ the method presented in [8] to segment each shape into disjoint regions. The segmentation approach is based on persistence-based clustering and is stable under near-isometric deformations. Given a heat kernel signature function [9] computed on each vertex of the shape, first a persistence diagram is computed and then shape vertices are merged, resulting in a disjoint segmentation which associates each vertex of the shape to a salient point.

978-1-7281-0397-6/19/$31.00 ©2019 IEEE

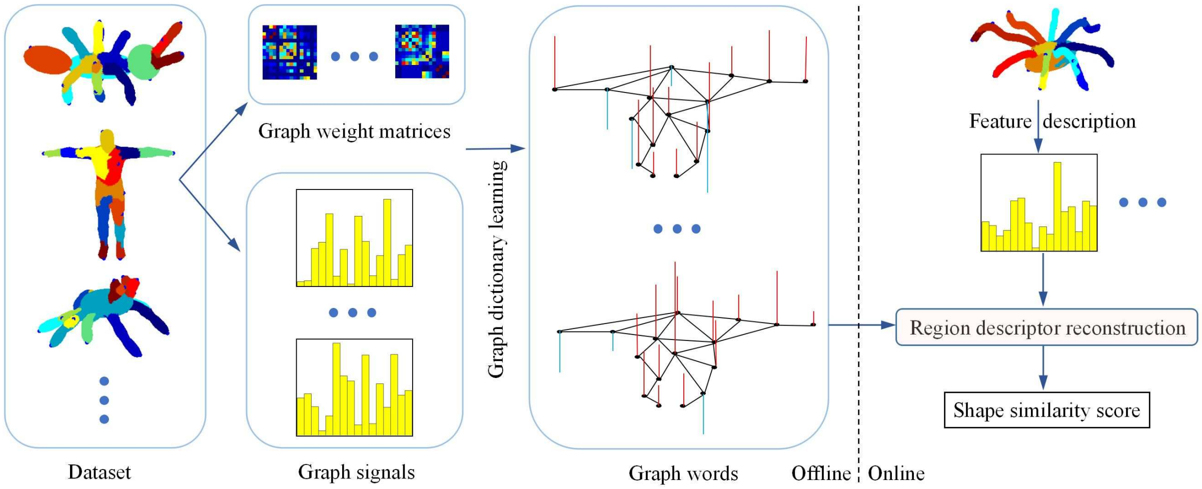

Fig. 1. The proposed graph-based shape descriptor learning framework. The learned graph words have positive signal values as red bars, and negative values as blue bars. The lengths of the bars on the graph vertices indicate the absolute signal values and only the graph edges whose weights are over a certain threshold are depicted.

Suppose that a shape is segmented into N regions and each region is associated with a salient point, by connecting each pair of salient points, a fully-connected weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ with N vertices is constructed for each shape. $\mathcal{V}$ is the vertex set, $\mathcal{E}$ denotes the edge set, W is the positive edge weight matrix computed with the generalized Radial Basis Function

$$\widetilde{W}_{ij} = e^{-\frac{d(v_i, v_j)}{\rho}} \tag{1}$$

where d measures the geodesic distance between a pair of vertices on the mesh, $\widetilde{W}_{ij}$ is the weight of the edge connecting graph vertices v_i and v_j. Defining edge weights with geodesic distances ensures the isometry invariance of the constructed graph.

As graph vertices generated from different shapes of the same class are in different orders, a permutation of the vertices of each graph is carried out with $W = \widetilde{P}\widetilde{W}\widetilde{P}^T$, where the permutation matrix $\widetilde{P}$ is computed as $\widetilde{P}_{ij} = \delta_{i,l_j}$, $\mathbf{l}$ is the region label vector of the associated shape with $l_k \in \{1, \ldots, N\}$ as the label of the kth region.

A signal $f : \mathcal{V} \to \mathbb{R}$ defined on the vertices of the graph $\mathcal{G}$ can be represented as a vector $f \in \mathbb{R}^N$, where the ith component of vector f represents the signal value at the ith vertex in $\mathcal{V}$. In this paper, we compute a descriptor vector for each region of the shape and concatenate the descriptors into a matrix $\widetilde{X}$, where the ith row of $\widetilde{X}$ equals the descriptor of the ith region. In this way, each column of $\widetilde{X}$ can be viewed as a signal defined on $\mathcal{G}$. The rows of descriptor matrix $\widetilde{X}$ are also out-of-order and a permutation is carried out as $X = \widetilde{P}\widetilde{X}$.

B. Solving Region Correspondence Problem

To solve the problem of region correspondence in the testing phase, a random forest is trained for each class of shapes using the region descriptors during the training phase. Given region descriptors of training shapes and the associated region labels $\{(y_\mathbf{s}, l_\mathbf{s})\}_{\mathbf{s}\in R}$, y_s is the descriptor vector of region $\mathbf{s}$ in the region set R, label $l_\mathbf{s}$ is from a canonical label set $\{1, \ldots, N\}$. Each decision tree in the random forest is trained from root to leaves using the information gain based algorithm, with linear classifiers as split functions located at the internal nodes. In the end, each leaf node is assigned a score vector $\mathbf{P}_m = [P_{m1}, \ldots, P_{ml}, \ldots, P_{mN}]$, with P_{ml} measuring the probability among the subset $\mathbb{S}$ reaching the leaf node

$$P_{ml} = \frac{|\{(y, l) \in \mathbb{S}\}|}{|\mathbb{S}|} \tag{2}$$

where $m = 1, \ldots, M$ is the index of the decision tree.

In the testing process, each region of a shape is assigned a label through the trained random forest. Given the descriptor y of a region, each decision tree in the forest routes the descriptor from the root to a leaf node according to the trained split functions of the tree. The prediction of the whole forest on the probability of a label l is computed by averaging the distributions of all trees in the random forest

$$P_l = \frac{1}{M}\sum_{m=1}^{M} P_{ml} \tag{3}$$

where P_{ml} is the computed score of the retrieved leaf node in the mth tree. Thus the permutation matrix $\widetilde{P}$ of the region descriptor matrix of a test shape can be computed as $\widetilde{P}_{ij} = \delta_{i,l_j}$, where l_j is the label of the jth region in the shape.

C. Graph Dictionary Learning

Given the weight matrix W of the constructed graph $\mathcal{G}$, the normalized graph Laplacian is computed as $\mathcal{L} = I - D^{-\frac{1}{2}}WD^{-\frac{1}{2}}$, where D is the diagonal degree matrix whose ith entry is the sum of weights of all edges incident to vertex i [10]. As $\mathcal{L}$ is a real symmetric matrix, it has a finite set of nonnegative eigenvalues and orthonormal eigenvectors.

The eigenvectors are denoted by $\{\mathcal{X}_l\}_{l=0,\ldots,N-1}$, and the eigenvalues by $\sigma(\mathcal{L}) = \{0 = \lambda_0 < \lambda_1 \leq \ldots \leq \lambda_{N-1} \leq 2\}$.

Consider a graph signal f defined on the vertices of graph $\mathcal{G}$, its graph Fourier transform can be defined as

$$\hat{f}(\lambda_l) = \langle f, \mathcal{X}_l \rangle = \sum_{n=1}^{N} f(n)\mathcal{X}_l^*(n) \tag{4}$$

The inverse graph Fourier transform is then given by

$$f(n) = \sum_{l=0}^{N-1} \hat{f}(\lambda_l)\mathcal{X}_l(n) \tag{5}$$

Following the definition of graph Fourier transform and the property that convolution in the vertex domain is equivalent to multiplication in the graph spectral domain, the generalized translation operator can be defined according to the convolution with a Kronecker δ function centered at the nth vertex

$$T_n(g) = \sqrt{N}(g * \delta_n) = \sqrt{N} \sum_{l=0}^{N-1} \hat{g}(\lambda_l)\mathcal{X}_l^*(n)\mathcal{X}_l \tag{6}$$

The localization of $T_n(g)$ around the center vertex is controlled by the smoothness of the kernel $\hat{g}(\cdot)$ [11]. Defining the kernel in the graph spectral domain with a smooth polynomial function of degree K as $\hat{g}(\lambda_l) = \sum_{k=0}^{K} \alpha_k \lambda_l^k,\ l = 0, \ldots, N-1$, the kernel g can be translated to a vertex n as

$$T_n(g) = \sqrt{N} \sum_{l=0}^{N-1} \sum_{k=0}^{K} \alpha_k \lambda_l^k \mathcal{X}_l^*(n)\mathcal{X}_l = \sqrt{N} \sum_{k=0}^{K} \alpha_k (\mathcal{L}^k)_n \tag{7}$$

where $(\mathcal{L}^k)_n$ denotes the nth column of $\mathcal{L}^k$. The concatenation of N columns gives $T(g)$:

$$T(g) = \sqrt{N}\hat{g}(\mathcal{L}) = \sqrt{N} \sum_{k=0}^{K} \alpha_k \mathcal{L}^k \tag{8}$$

The graph dictionary $\mathcal{D} = [\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_S]$ can be defined with the above polynomial of graph Laplacian as

$$\mathcal{D}_s = \hat{g}_s(\mathcal{L}) = \sum_{k=0}^{K} \alpha_{sk} \mathcal{L}^k \tag{9}$$

where column n of $\mathcal{D}_s$ has its support contained in the K-hop neighborhood of vertex n. Each subdictionary $\mathcal{D}_s$ is capable of capturing all possible translations of a certain pattern. Given a signal f defined on $\mathcal{V}$, it can be represented with the dictionary atoms as $f = \mathcal{D}x$, where x is the coefficient vector of linear combination. For efficient representation of a graph signal, vector x should be sparse and should capture the most prominent characteristics of the graph signal [6].

The proposed method processes each dimension of the region descriptor as a graph signal defined on the vertex set $\mathcal{V}$. Assuming that shapes of the same class generate graph signals that share similar spectral properties, we learn a dictionary for each class of shapes with a variation of the multi-graph dictionary learning approach [7], which is capable of capturing similar spectral properties among signals on different graphs.

Given training shapes $\{M_t\}_{t=1}^T$ from category C, we first compute the isometry-invariant graph $\mathcal{G}_t$ and region descriptor matrix X_t of each shape, then learn a spectral graph dictionary $\mathcal{D}_t = [\mathcal{D}_t^1, \mathcal{D}_t^2, \ldots, \mathcal{D}_t^S]$, where $\mathcal{D}_t^s = \sum_{k=0}^{K} \alpha_{sk} \mathcal{L}_t^k$, $\mathcal{L}_t$ is the normalized graph Laplacian of graph $\mathcal{G}_t$, S is the number of subdictionaries. The dictionary learning problem can be formulated as the following optimization problem

$$\begin{aligned} \underset{\alpha \in \mathbb{R}^{(K+1)S}, A_t \in \mathbb{R}^{SN \times q}}{argmin} \quad & \sum_{t=1}^{T} \frac{1}{q} ||X_t - \mathcal{D}_t A_t||_F^2 + \mu ||\alpha||_2^2 \\ s.t. \quad & ||A_t^i||_0 \leq T_0 \quad \forall i \in \{1, \ldots, q\} \\ & \text{support}(A_t^1) = \cdots = \text{support}(A_t^q) \\ & 0 \preceq \mathcal{D}_t^s \preceq c \quad \forall s \in \{1, \ldots, S\} \end{aligned} \tag{10}$$

where q is the dimension of region descriptors, A_t^i denotes the ith column of the coefficient matrix A_t, T_0 is the maximum sparsity level of the coefficients of each training signal, $\alpha = [\alpha_1; \ldots; \alpha_S]$, α_s is the polynomial coefficient vector with $K+1$ entries, $\text{support}(A_t^i)$ denotes the index set of non-zero elements in A_t^i. As different dimensions of region descriptors are correlated, it is preferable to enforce simultaneous sparse coding in the optimization process. The constraint on the spectrum guarantees that the kernels are nonnegative and uniformly bounded by a given constant c. The norm of α is penalized to promote smoothness in the kernels.

The objective function (10) is not jointly convex, but it can be reduced gradually as we alternatively update the sparse codes and the dictionary. First, when fixing the dictionary, we solve with respect to the sparse codes using the simultaneous orthogonal matching pursuit (S-OMP) [12] on each training signal. In the second step, we fix the sparse codes, and update the dictionary by solving the quadratic problem with respect to the parameter vector α.

III. Experiments

We use the datasets of SHREC' 10 and SHREC' 11 non-rigid 3D shape retrieval tracks for evaluation of the proposed approach. The SHREC' 10 dataset consists of 200 selected articulated models evenly divided into 10 classes based on their semantic meanings. Each class of the dataset consists of shapes in different poses deformed from one shape. The names of the classes are ant, crab, hand, human, octopus, plier, snake, spectacle, spider and teddy. SHREC' 11 contains 600 watertight mesh models evenly divided into 30 classes based on their semantic meanings. Each class of the dataset consists of shapes in different poses deformed from a canonical shape. The names of the shape classes are alien, ant, armadillo, bird1, bird2, camel, cat, centaur, dinoskel, dinosaur, dog1, dog2, flamingo, glasses, gorilla, hand, horse, lamp, man, octopus, paper, pliers, rabbit, santa, scissor, shark, snake, spider, twoballs and woman.

The proposed algorithm was compared with state-of-the-art methods including GPS-embedding [13], shape-DNA [14], compact shape-DNA [15], and F1-, F2-, and F3-features [16].

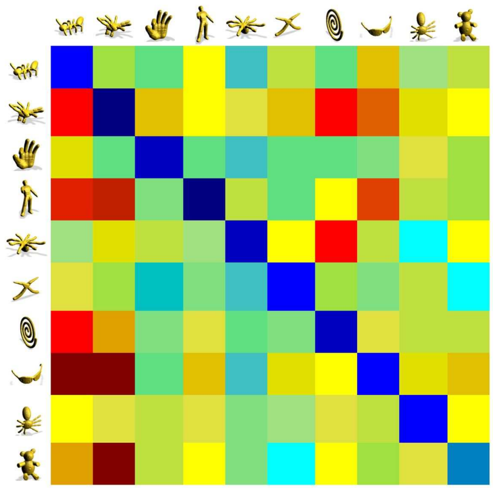

Fig. 2. Dissimilarity matrix SHREC' 10 dataset demonstrating generative errors of different classes of learned graph dictionaries to each class of shapes. Cold colors represent lower values while hot colors represent higher values.

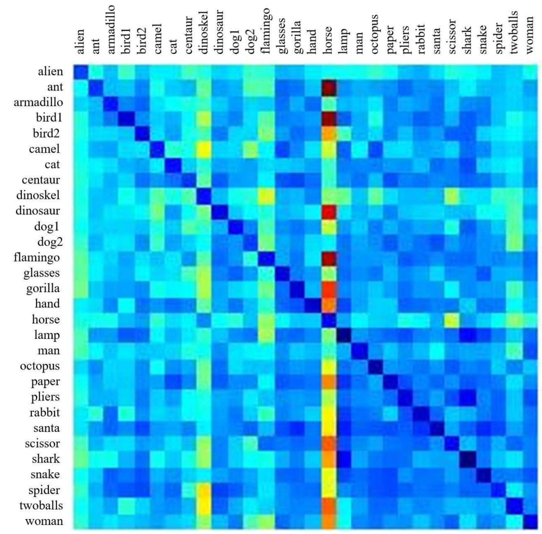

Fig. 3. Dissimilarity matrix of SHREC' 11 dataset demonstrating generative errors of different classes of learned graph dictionaries to each class of shapes.

The classification results are averaged over 10 runs. For both datasets, we randomly select 30% of the shapes for testing and the rest for training in each run. The number of segmented regions of each shape is set to 16 for both datasets. The wave kernel signature [17] of dimension 100 and weightedly averaged by areas of vertices in each region is used as the region descriptor. Each shape is classified into the class of the learned dictionary that gives the smallest generative error for the shape's region descriptor matrix. The classification results of different approaches are given in Table I and Table II, showing that the proposed method outperforms the other approaches on both benchmark datasets. The dissimilarity matrices computed for the datasets are given in Fig. 2 and Fig. 3. Note that the matrix is not a classical dissimilarity matrix, each column represents a class of shapes and each row shows the mean generative errors of a class-specified dictionary for different classes of shapes. Aside from the diagonal values being the lowest, the low values between octopus and pliers and between octopus and ants also indicate the similarity property of the proposed descriptor. The higher diagonal value of teddy shows that the proposed method is less effective with shapes that have more flat areas than protruding areas, where the extracted regions associated to prominent points are less stable.

TABLE I
CLASSIFICATION ACCURACY RESULTS ON THE SHREC'10 DATASET.

Method	Average accuracy %
GPS-embedding	88.33 ± 0.54
Shape-DNA	90.10 ± 0.86
cShape-DNA	91.67 ± 0.93
F1-features	85.16 ± 1.33
F2-features	83.33 ± 0.78
F3-features	86.67 ± 1.05
Proposed	**95.83 ± 1.26**

TABLE II
CLASSIFICATION ACCURACY RESULTS ON THE SHREC'11 DATASET.

Method	Average accuracy %
GPS-embedding	87.78 ± 0.62
Shape-DNA	92.22 ± 0.67
cShape-DNA	93.56 ± 0.85
F1-features	91.11 ± 0.79
F2-features	88.35 ± 1.12
F3-features	91.57 ± 0.91
Proposed	**94.17 ± 0.97**

IV. Conclusion

In this paper, we present a novel approach for non-rigid 3D shape descriptor learning. By constructing a graph connecting segmented regions of a shape and considering region descriptors as signals defined on graph vertices, the multi-graph dictionary learning approach can be employed to train a dictionary for each category of shapes. The proposed approach achieves state-of-the-art performances on standard shape benchmark datasets.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 61425011, Grant 61720106001, Grant 61529101, and in part by Shanghai High Technology Project under Grant 17511106603 and the Program of Shanghai Academic Research Leader under Grant 17XD1401900.

References

[1] R. Litman, A. Bronstein, M. Bronstein *et al.*, "Supervised learning of bag-of-features shape descriptors using sparse coding," in *Computer Graphics Forum*, 2014, vol. 33, no. 5, pp. 127–136.

[2] A. Gasparetto and A. Torsello, "A statistical model of riemannian metric variation for deformable shape analysis," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2015, pp. 1219–1228.

[3] R. Litman and A. M. Bronstein, "Learning spectral descriptors for deformable shape correspondence," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 171–180, 2014.

[4] K. Kreutz-Delgado and J. F. Murray, "Dictionary learning algorithms for sparse representation," *Neural computation*, vol. 15, no. 2, pp. 349–396, 2003.

[5] D. Boscaini and U. Castellani, "A sparse coding approach for local-to-global 3d shape description," *The Visual Computer*, vol. 30, no. 11, pp. 1233–1245, 2014.

[6] D. Thanou, D. I. Shuman, and F. Pascal, "Learning parametric dictionaries for signals on graphs," *IEEE Trans. Signal Processing*, vol. 62, no. 15, pp. 3849–3862, 2014.

[7] D. Thanou and P. Frossard, "Multi-graph learning of spectral graph dictionaries," in *Proc. IEEE Conf. Acoustics, Speech and Signal Processing*, 2015, pp. 3397–3401.

[8] P. Skraba, M. Ovsjanikov, F. Chazal *et al.*, "Persistence-based segmentation of deformable shapes," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops*. IEEE, 2010, pp. 45–52.

[9] J. Sun, M. Ovsjanikov, and L. Guibas, "A concise and provably informative multi-scale signature based on heat diffusion," in *Computer Graphics Forum*, vol. 28, no. 5, 2009, pp. 1383–1392.

[10] F. R. Chung, "Spectral graph theory," *American Mathematical Soc*, no. 92, 1997.

[11] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[12] T. A. Joel, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[13] A. J. Chaudhari, R. M. Leahy, B. L. Wise *et al.*, "Global point signature for shape analysis of carpal bones," *Physics in Medicine & Biology*, vol. 59, no. 4, p. 961, 2014.

[14] M. Reuter, F. E. Wolter, and N. Peinecke, "Laplace-Beltrami spectra as 'Shape-DNA' of surfaces and solids," *Computer-Aided Design*, vol. 38, no. 4, pp. 342–366, 2006.

[15] Z. Gao, Z. Yu, and X. Pang, "A compact shape descriptor for triangular surface meshes," *Computer-Aided Design*, vol. 53, pp. 62–69, 2014.

[16] M. A. Khabou, L. Hermi, and M. B. H. Rhouma, "Shape recognition using eigenvalues of the Dirichlet Laplacian," *Pattern recognition*, vol. 40, no. 1, pp. 141–153, 2007.

[17] M. Aubry, U. Schlickewei, and D. Cremers, "The wave kernel signature: A quantum mechanical approach to shape analysis," in *Proc. IEEE Conf. Computer Vision Workshops*, 2011, pp. 1626–1633.