A Robust Local Spectral Descriptor for Matching Non-Rigid Shapes with Incompatible Shape Structures

Yiqun Wang[1,2], Jianwei Guo[1,2]*, Dong-Ming Yan[1,2]*, Kai Wang[3], Xiaopeng Zhang[1,2]

[1]NLPR, Institute of Automation, Chinese Academy of Sciences
[2]University of Chinese Academy of Sciences
[3]University Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab

Abstract

Constructing a robust and discriminative local descriptor for 3D shapes is a key component of many computer vision applications. Although existing learning-based approaches can achieve good performance in specific benchmarks, they usually fail to learn sufficient information from shapes with different types and structures (e.g., spatial resolution, connectivity, and transformation).

To solve this issue, we present a more discriminative local descriptor for deformable 3D shapes with incompatible structures. Based on spectral embedding using the Laplace–Beltrami framework on the surface, we construct a novel local spectral feature that exhibits high resilience to changes in mesh resolution, triangulation, and transformation. The multiscale local spectral features around each vertex are then encoded into a geometry image called vertex spectral image in a compact manner. Such vertex spectral images can be efficiently trained to learn local descriptors using a triplet neural network. Then, we present a new benchmark dataset for training and evaluation by extending the widely used FAUST dataset. We utilize a remeshing approach to generate modified shapes with different structures. Furthermore, we evaluate the proposed approach thoroughly and conduct an in-depth comparison to demonstrate that our approach outperforms recent state-of-the-art methods on this benchmark.

1. Introduction

Obtaining detailed 3D shapes has become easy with the advancement of 3D scanning devices and computer vision reconstruction techniques. Accordingly, the importance of 3D shape analysis (e.g., shape matching, segmentation, correspondence, and retrieval) has increased remarkably. Designing local descriptors for surface points is a fundamental problem in computer vision, computer graphics and
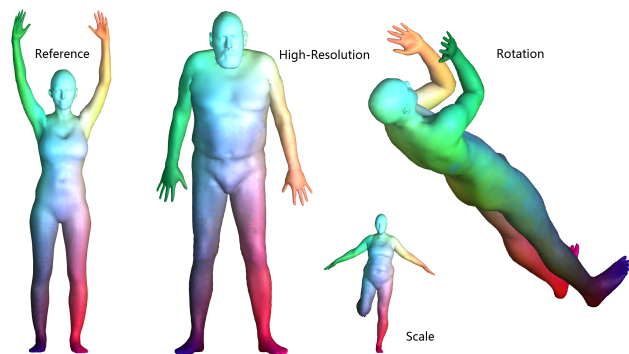
Figure 1. Shape matching results between the reference shape (6890 points) and different shapes with high-resolution (12K points) and other incompatible structures.

robotics, and it is a building block of various shape analysis and geometry processing applications.

Many local shape descriptors, including hand-crafted shape feature [13] and deep learning-based descriptors [8], have been proposed for different scenarios. Compared with hand-crafted methods that are usually limited by the representation power of hand-tuned parameters, deep learning methods innovatively apply deep convolution to non-Euclidean geometric data and exhibit better performance. However, current popular datasets used for descriptor learning, such as FAUST [3], SCAPE [1], and TOSCA [7], have a compatible shape structure, that is, mesh resolution and connectivity are identical across all shapes. The acquired/reconstructed surface shapes usually have incompatible shape structures even when the same acquisition technique is used, and their resolution, triangulation, and transformation may differ. Constructing robust and informative local descriptors for such shapes thus remains challenging. Furthermore, existing deep learning methods have serious limitations. First, training and testing usually require the same number of vertices on the input shapes because of the fixed network architecture. Second, if the features fed into the network are sensitive to triangulation, ensuring that the learned features are robust to triangulation cannot be guaranteed. Third, the model needs to be normalized to the same

scale before training. However, sometimes, two models are normalized incorrectly such as in partial matching.

In this study, we propose a new dense local descriptor for matching non-rigid shapes with incompatible structures. In our approach, we theoretically derive a novel local spectral feature called *Local Point Signature* (LPS) from Dirichlet energy. Different from traditional spectral descriptors (e.g., *Global Point Signature* (GPS), *Heat Kernel Signature* (HKS), and *Wave Kernel Signature* (WKS)), we introduce extra vertex coordinate information to capture distinctive attributes in a local patch. The multiscale LPS features around each vertex are then compactly encoded into the pixels of vertex spectral images, on which a classic triplet neural network can be applied directly to obtain a more discriminative local descriptor. Extensive experimental evaluation shows that the proposed approach outperforms recent state-of-the-art local descriptors.

Contributions. The main contributions of this work are as follows:

- We present a novel local spectral feature that exhibits high resilience to changes in mesh resolution, triangulation, scale, and rotation.

- We design a new representation form of geometry image (GI) called vertex spectral image, in which local spectral features can be encoded in a compact and memory-efficient manner.

- We provide a new benchmark dataset by extending the well-known FAUST dataset, which contains six resolution shapes with different connectivities and transformations.

2. Related Work

We present a brief overview of local descriptors from three aspects.

Spatial domain-based approaches. *Spin images* (SI) [17] and *3D shape context* (3DSC) [11] create accumulators by dividing the local space into different bins and calculating the number of points that fall into each bin to construct descriptors. *Signature of histogram of orientations* (SHOT) [31] generates descriptors by accumulating the normal angles of the key and neighboring points in the neighborhood space. Unlike the SHOT descriptor, the Mesh-HOG [37] descriptor is a histogram based on the orientations of the gradients on the mesh. The *rotational projection statistics* (RoPS) [14] descriptor is generated by rotationally projecting neighboring points onto 2D planes and calculating a set of statistics. Obviously, these traditional approaches are based on the spatial domain feature and susceptible to resolution and shape deformation. Recently, the *dual spin image* (DSI) [36] descriptor was proposed by introducing 3D critical net on the surface, which is robust

against non-rigid deformation. However, this method does not consider the matching between different resolutions.

Embedding-based approaches. Many intrinsic descriptors have been proposed to deal with isometric deformations. The most common method for generating embedded descriptors is based on the Laplace–Beltrami operator. Shape–DNA [28] considers the spectrum of the Laplace–Beltrami operator as the descriptor because the spectrum is isometry-invariant and independent of spatial position. GPS [29] combines the spectrum and eigenfunctions to obtain the global descriptor of a vertex. HKS [30], scale-invariant HKS [9], WKS [2], and *optimal spectral descriptors* (OSD) [23] were proposed based on diffusion geometry. Hu *et al*. [15] extracted salient geometric feature points in the Laplace–Beltrami spectral domain and considered the Laplace–Beltrami spectrum in the local region a local shape descriptor. However, most of these embedding methods are based on global intrinsic features and are not robust enough for local detail descriptions.

Deep learning-based approaches. 3D deep learning methods have recently been used to extract shape descriptors. Wei *et al*. [34] generated invariant descriptors by using a large dataset of depth maps for training. Huang *et al*. [16] extracted local descriptors by training multiple rendered views in multiple scales. Zeng *et al*. [38] used 3D volumetric convolutional neural networks to generate local descriptors for robustly matching RGB-D data. The method of *compact geometric features* (CGF) [18] maps high-dimensional histograms into a low-dimensional Euclidean space to generate descriptors on unstructured point clouds. Wang *et al*. [32] presented a deep learning framework using local GIs to learn descriptors.

In addition, geometric deep learning has been developed for directly training on manifolds. *Anisotropic diffusion descriptors* [6] are learned by using a fully connected neural network to construct direction-sensitive features. GCNN [24], ACNN [5], and MoNet [26] train the network effectively on manifolds by defining the convolution operation in the spatial domain, and they are mainly suitable for shape correspondence applications. FMNet [22] is another framework that introduces deep learning into functional maps. LSCNN [4] generalizes windowed Fourier transform to deeply learn local descriptors, and it can be regarded as a time–frequency analysis on manifolds. Lim *et al*. [21] avoided resampling to learn correspondences on unstructured surface meshes by using a spiral operator on neighborhood vertices. However, given that these methods learn information related to shape structures (e.g., mesh scale and spatial resolution) that vary in shape, their generalization capability is defective. We focus on these issues and present a novel approach to generate a local spectral feature that is robust to different resolutions, triangulations, and transformations.
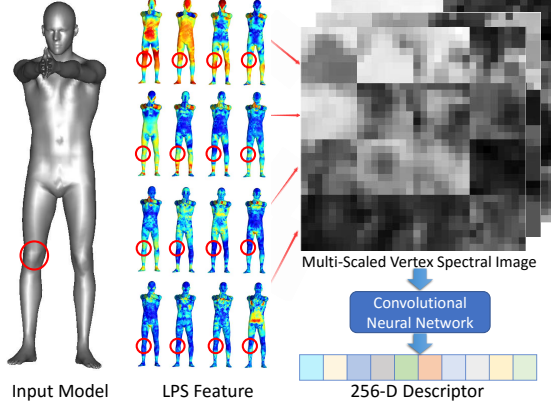
Figure 2. Overview of our descriptor generation framework. Given an input shape (left), we first compute a 16-dimensional local spectral feature for each vertex of the mesh to construct feature models, and the scalar value of each dimension is color-coded in the middle images. For a point of interest, we then generate vertex spectral images by packing the 16-dimensional LPS in a compact manner. Finally, we send the vertex spectral images to the convolutional neural network to obtain our descriptor.

3. Problem Statement and Overview

Given a mesh $\mathcal{M}$ with irregular sampling on the underlying smooth surface defined as (V, E), where $V = \{v_i | i = 1, ..., n\}$ and E are the sets of vertices and edges, respectively. Real function $X = (x_1, x_2, x_3) : V \rightarrow R^3$ on the vertices represents vertex coordinates. Our goal is to obtain a discriminative local descriptor $\mathrm{f}(v_i) \in R^d$ for any given vertex v_i.

The proposed method of generating the local descriptor is divided into two stages, namely, local spectral feature construction and local descriptor learning. As shown in Fig. 2, in the first stage, we propose a local point signature for each vertex in the frequency domain that is robust to resolution, triangulation, scale, and rotation. The second stage is for descriptor learning. Inspired by the recent deep learning method [32], we design a local vertex spectral image using GI to encode the proposed local spectral feature. Benefiting from the designed local spectral feature and vertex spectral image, our method can exploit all vertex information instead of only the sampling vertices for training.

4. Local Spectral Feature

This section introduces the Laplace–Beltrami framework. Then, the local spectral feature is proposed, and several properties of our proposed feature are analyzed.

4.1. Laplace–Beltrami Framework

Given a smooth real function $f : \mathcal{S} \rightarrow R$ on the surface, where $\mathcal{S}$ is a continuous surface, the following Dirichlet en-

ergy is a measure of the change of function f over $\mathcal{S}$ [8].

$$E(f) = \int_S |\nabla f(v)|^2 dv = \int_S f(v) \Delta f(v) dv \quad (1)$$

We often seek function f as smooth as possible, which minimizes the Dirichlet energy. If we obtain a set of orthonormal basis functions $\{\phi_i | i = 0, 1, ..., k-1\}$ on $\mathcal{S}$ and the basis function minimizes Dirichlet energy, then the solution is the first k eigenfunctions of Δ [8]. In this study, Δ is the Laplace–Beltrami operator $\Delta f = -\mathrm{div}(\nabla f)$, which satisfies

$$\Delta \phi_i = \lambda_i \phi_i, i = 0, 1, ..., k-1, \quad (2)$$

where $\{\lambda_i | i = 0, 1, ..., k-1\}$ is the smallest k eigenvalues in increasing order.

Given that the data we process involve a 3D mesh, the discrete case is highly important, and this means that the function $\tilde{f} : V \rightarrow R$ on mesh $\mathcal{M}$ becomes discrete [25]. On the meshes, the value of Δf on v_i can be written as

$$\Delta \tilde{f}(v_i) = \frac{1}{a_i} \sum_{j \in N(v_i)} \frac{\cot \alpha_{ij} + \cot \beta_{ij}}{2} \left(\tilde{f}(v_i) - \tilde{f}(v_j) \right),$$

where a_i is the area of the Voronoi region of v_i. $N(v_i)$ represents the one-ring neighborhood vertex set of v_i. α_{ij} and β_{ij} are the two angles opposite the edge $\{i, j\}$ on the triangle mesh. Fig. 3 shows the angles and Voronoi area of the vertex.
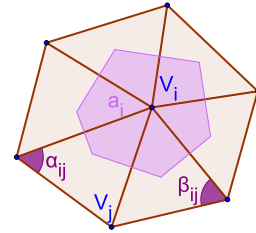

Figure 3. Illustration of the angles and Voronoi area in the discrete Laplace–Beltrami operator.

Therefore, Laplace–Beltrami matrix $\mathbf{L}$ can approximate Laplace–Beltrami operator Δ in a discrete manner and can be written as

$$\mathbf{L}_{ij} = \begin{cases} -\frac{\cot \alpha_{ij} + \cot \beta_{ij}}{2a_i} & \text{if} \quad i, j \ are \ adjacent \\ \sum_k \frac{\cot \alpha_{ik} + \cot \beta_{ik}}{2a_i} & \text{if} \quad i = j \\ 0 & otherwise. \end{cases}$$

The orthogonal basis of minimizing Dirichlet energy can be solved in a discrete form.

$$\mathbf{L}\mathbf{\Phi}_i = \lambda_i \mathbf{\Phi}_i, i = 0, 1, ..., k-1 \quad (3)$$

If the Laplace–Beltrami matrix is positive semi-definite, then the eigenvalues λ_i are nonnegative, and their minimum is 0. These eigenvalues can be sorted from small to

large and interpreted as frequencies in the signal processing community. In addition, the symmetry property of $\mathbf{L}$ facilitates numerical calculations. Unfortunately, $\mathbf{L}$ is asymmetrical, which may lead to imaginary numbers in the numerical computation. Equation (3) can be rewritten as a generalized eigenvalue problem by introducing two symmetric matrices $\mathbf{T}$ and $\mathbf{A}$, and the rewritten form is presented as

$$\mathbf{T}\mathbf{\Phi_i} = \lambda_i \mathbf{A}\mathbf{\Phi_i}, i = 0, 1, ..., k-1, \qquad (4)$$

where

$$\mathbf{T}_{ij} = \begin{cases} -\frac{\cot \alpha_{ij} + \cot \beta_{ij}}{2} & \text{if } i,j \text{ are adjacent} \\ \sum_k \frac{\cot \alpha_{ik} + \cot \beta_{ik}}{2} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

and $\mathbf{A}$ is a diagonal matrix, $\mathbf{A}_{ii} = a_i$.

By using the Arnoldi method of ARPACK [20], the eigenvectors of discrete Laplacian are real, and the eigenvalues are real and nonnegative.

4.2. Local Point Signature

After calculating the eigenvectors $\mathbf{\Phi_i}$ in the generalized eigenvalue problem, $\mathbf{\Phi_i}$ are orthogonal in terms of the $\mathbf{A}$-dot product.

$$\langle \mathbf{\Phi_i}, \mathbf{\Phi_j} \rangle_{\mathbf{A}} = \mathbf{\Phi}_i^T \mathbf{A} \mathbf{\Phi}_j \qquad (5)$$

Similar to the Fourier series, the functions defined on mesh vertices can be expanded as follows:

$$f = \sum_{j=0}^{+\infty} \sigma_j \phi_j, \qquad (6)$$

where σ_j is the coefficient of the frequency domain on the surface. Therefore, we know that the function on the surface can be transformed into a set of coefficients in the frequency domain by a basis function. In the discrete case, the coefficients can be calculated by

$$\sigma_j = \left\langle \tilde{f}, \mathbf{\Phi_j} \right\rangle_{\mathbf{A}} = \tilde{f}^T \mathbf{A} \mathbf{\Phi}_j. \qquad (7)$$

We now introduce the derivation of the proposed spectral features.

Recall the formula of Dirichlet energy, Equation (1), and its discrete form $\tilde{f}^{\mathrm{T}} \mathbf{A} \mathbf{L} \tilde{f}$ in terms of the $\mathbf{A}$-dot product. Combined with Equation (3) and Equation (6), the Dirichlet energy of the function can be expressed as

$$\widetilde{E}\left(\tilde{f}\right) = \tilde{f}^{\mathrm{T}} \mathbf{A} \mathbf{L} \tilde{f} = \sum_{j=0}^{N-1} \sigma_j^2 \lambda_j, \qquad (8)$$

where N is the number of vertices.

If a more generalized function $F = (f_1, f_2, ..., f_d) : S \rightarrow R^d$ is obtained on the surface, the Dirichlet energy over S is

$$E\left(F\right) = \int_S |\nabla F\left(v\right)|^2 dv = \sum_{i=1}^{d} \int_S |\nabla f_i\left(v\right)|^2 dv. \qquad (9)$$

Equation (8) is transformed into the discrete form of Equation (9) and presented as

$$\widetilde{E}\left(\widetilde{F}\right) = \sum_{i=1}^{d} \sum_{j=0}^{N-1} \sigma_{ij}^2 \lambda_j \qquad (10)$$

$$= \sum_{j=0}^{N-1} \lambda_j \sum_{i=1}^{d} \sigma_{ij}^2. \qquad (11)$$

The Dirichlet energy of this generalized function can be composed of different frequency band energies $\lambda_j \sum_{i=1}^{d} \sigma_{ij}^2$ that can be considered a feature of the surface in this band.

Therefore, we remove λ_0 with an eigenvalue of 0. The proposed spectral feature can be expressed as follows:

$$sf = \left(\lambda_1 \sum_{i=1}^{d} \sigma_{i1}^2, \lambda_2 \sum_{i=1}^{d} \sigma_{i2}^2, ..., \lambda_{N-1} \sum_{i=1}^{d} \sigma_{iN-1}^2 \right). \qquad (12)$$

Thus far, we have proposed the spectral feature on the surface. Next, we demonstrate how to design a local spectral feature using vertex coordinates with a rigid transformation-invariant property. It is worth mentioning that we can design other features by using different information as the input.

For a vertex v_i, we extract the neighbor triangles around the vertex and build a local patch mesh. Neighborhood selection uses a geodesic disk whose radius is a fraction of a geodesic diameter (the longest geodesic path on the mesh). Parameter selection is shown in the supplementary material. We simply extract a local mesh in the geodesic disk. After the local patch mesh is obtained, we construct the Laplace–Beltrami matrix on the patch. For this patch, we have the function $X = (x_1, x_2, x_3) : V \rightarrow R^3$ in the vertices representing the vertex coordinates.

In Equation (9), the Dirichlet energy over patch P can be rewritten as

$$E\left(X\right) = \int_P |\nabla X\left(v\right)|^2 dv = \sum_{i=1}^{3} \int_P |\nabla x_i\left(v\right)|^2 dv. \qquad (13)$$

However, for the case with a boundary, we cannot assume that Equation (1) is satisfied because of boundary energy. We take the solution of Hu *et al.* [15]. By attaching the same surface patch with opposite normals to the open boundary patch, we obtain a watertight surface with zero volume. Therefore, Equation (1) is naturally satisfied.

Therefore, the equation can be written as follows:

$$\widetilde{E}(X) = \sum_{j=0}^{N-1} \lambda_j \sum_{i=1}^{3} \sigma_{ij}^2. \qquad (14)$$

We select the feature of 16 dimensions corresponding to the smallest 16 eigenvalues without 0 as our local spectral features, and we define them as *Local Point Signature* (LPS). The reason for choosing 16 is twofold. First, a large amount of information is concentrated at a low frequency. Hence, we choose 16 dimensions to avoid the influence of high-frequency noise as much as possible. Second, constructing our vertex spectral images using 16 dimensions is convenient. In addition, an intuitive reason for robustness can be explained. With the input of vertex coordinates, the energy represents the area of the surface that is robust to triangulation and coordinate-independent despite using coordinate information. Hence, the descriptors that we expand the energy in the frequency domain should be insensitive to these incompatible shape structures. This trait will be proven through experiments later. In addition, to obtain scale invariance, we multiply each energy band by its eigenvalue λ_j and calculate square root of the energy. The reason and theoretical proof will be specified in the supplementary material. The expression of the proposed feature is as follows.

$$LPS = \left(\lambda_1 \sqrt{\sum_{i=1}^{3} \sigma_{i1}^2}, \lambda_2 \sqrt{\sum_{i=1}^{3} \sigma_{i2}^2}, .., \lambda_{16} \sqrt{\sum_{i=1}^{3} \sigma_{i16}^2} \right) \quad (15)$$

4.3. Discussion

Several spectral features, such as GPS, HKS, and WKS have been proposed for previous methods. These methods avoid relying on any extrinsic features to obtain global intrinsic features on the mesh, but the capability of their local description is not strong. Although intrinsic features are invariant under (near-) isometric deformations, they still encounter problems when dealing with non-isometric deformations. In practice, several studies [22, 26] have obtained better results by using non-intrinsic descriptors, such as SHOT. In our approach, we introduce vertex coordinate information to design a new spectral feature by utilizing the Dirichlet energy framework. Therefore, we combine the advantages of intrinsic and extrinsic descriptors to obtain a more discriminative local spectral feature.

In addition, many extrinsic (e.g., SHOT) and intrinsic (e.g., shape DNA, HKS, and WKS) features are sensitive to scale transformation. In comparison, our proposed features can theoretically guarantee scale invariance. The proof of our invariant properties is provided in the supplemental materials.
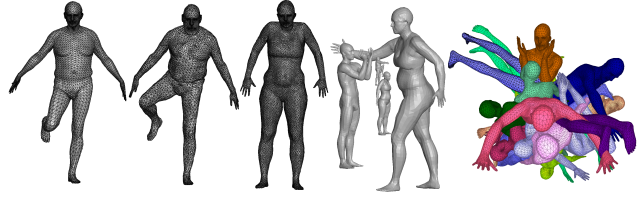

Figure 4. Sample shapes from our dataset. The three models on the left have different resolutions (6890, 8K, and 15K) and triangulations. We show the models with different scalings (middle) and rotations (right).

5. Dataset Creation and Descriptor Learning

An appropriate dataset containing 3D shapes with incompatible structures is required to demonstrate the validity of the proposed spectral feature. We describe how to generate such shapes automatically. Then, we propose a novel vertex spectral image encoding the previous LPS feature, which can be fed into an existing learning framework. Section 5.3 describes the architecture of our network.

5.1. Dataset

The FAUST dataset is a challenging and widely used standard benchmark for solving shape-matching tasks, and it has richer details and more accurate ground-truth correspondences than other datasets. Our benchmark expands upon this dataset to generate shapes with different levels of resolution, triangulation, scale, and rotation.

Resolution and triangulation. All models in FAUST have the same triangulation and resolution with 6890 vertices. We refer to [10, 33] for details on remeshing and generate multi-resolution shapes with irregular triangulation while ensuring that each shape contains user-specified vertices.

Specifically, to retain the ground-truth correspondence across all shapes for accurate evaluation, we mark each original vertex with a "lock" flag, which indicates that the original vertices cannot be moved or deleted during remeshing. Next, the number of user-specified points is satisfied by randomly adding or removing vertices via edge splitting or edge collapsing to increase or decrease the mesh resolution, respectively. We then apply a random edge flipping operation iteratively to obtain an irregular triangulation. Several vertex smoothing operations may be applied to avoid too many badly shaped triangles. These operations will maintain the detail features of the model by considering the Hausdorff distance. For coarse resolution, we use the models with 5K vertices in [27] that are independently remeshed by the centroidal Voronoi tessellation (CVT) approach [35]. Six resolutions are created for the benchmark, namely 5K, original 6890, 8K, 10K, 12K, and 15K vertices.

Scale and rotation. For scaling, we randomly scale the shapes with one of the following five scaling factors: 0.2, 0.5, 1.0, 2.0, and 4.0. For rotation, we generate the rotated shapes by randomly rotating the models. Fig. 4 presents

several selected examples from our dataset.

5.2. Vertex Spectral Images

Geometry image (GI) [12] is a good tool for converting irregular geometry into regular 2D images. Wang *et al*. [32] generated a local GI for descriptor learning. In their approach, the vertex normal direction and principal curvatures are encoded into the pixels of a GI. However, this method requires a large amount of memory to store multichannel GIs. For example, an image with a size of 32×32 is required to store only one element of the vertex normal. In addition, the low-level features used in their approach are more sensitive to shape structures.

To become memory-efficient, we introduce a new representation of the local GI, called vertex spectral image to encode our proposed LPS feature. Fig. 5 shows examples of vertex spectral images, and each of them is composed of 16 smaller GIs. To build such a vertex spectral image around a point $\mathbf{p}$, we regularly sample 64 points from the local patch of $\mathbf{p}$. Notably, each point is equipped with a 16-dimensional LPS feature. Hence, we can obtain 16 small GIs with a size of 8×8, and each GI encodes one element of the LPS feature. Our LPS feature is generated from the local patch instead of the original vertex information. Thus, 64 sampling points are sufficient to describe the local details, and the amount of data is reduced. Then, we pack these small GIs into a large vertex spectral image (32×32) by rearranging them in a row-major order, thus enabling the use of convolution operation to learn information between frequency bands. As a result, small GIs with the feature corresponding to the smallest and largest eigenvalues are placed in the upper left and lower right corners of the vertex spectral image, respectively. This approach significantly reduces the amount of memory used to encode the geometric information of a vertex. In addition, this splicing operation only applies to coordinate-independent features, such as ours. If the feature is coordinate-dependent (e.g., vertex normal direction), then the results will be affected by the rotation because of the ordering problem. Given that our LPS is coordinate-independent, which is unrelated to the embedding of the model, the case of model rotation can be solved.

Finally, we extract the local patch at three scales to construct multiscale vertex spectral images with neighbor geodesic radii of 2.1, 2.7, and $3.5\rho_0$, respectively. ρ_0 is computed as 1% of the geodesic diameter of the entire mesh. In addition, we also generate 12 vertex spectral images around each point for training by rotating the small GIs at $30°$ intervals.

5.3. Training Architecture Details

To learn our descriptor, we adopt the triplet neural network presented in [32] that consists of three identical convolutional networks ("ConvNet" for simplicity). For each
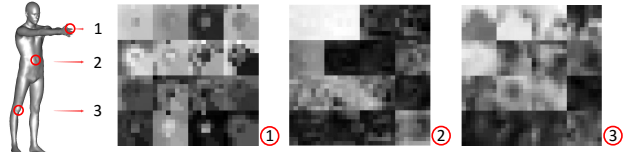

Figure 5. Vertex spectral images around three vertices: (1) middle finger, (2) belly button, and (3) right knee.

ConvNet, we utilize a network constructed as CONV128-3x3-/2 + CONV256-3x3-/2 + CONV512-3x3-/2 + FC512 + FC256. CONVx refers to a convolutional layer that has an x-dimensional output of feature maps, wherein 3x3 is the size of the convolution kernel, /2 is a pooling operation with a stride of 2, and FCx refers to a fully connected layer that outputs a vector with x-dimension. The output feature is expanded to 256 dimensions because we introduce the vertex spectral images of all points as training data. The main difference between our learning process and that of [32] is that the anchor-positive and anchor-negative pairs are selected not only as the key points but also as all the points encoded by utilizing the proposed vertex spectral image.

6. Experimental Results

We conduct an extensive evaluation to verify the effectiveness of the proposed local descriptor. An in-depth comparison with other related state-of-the-art approaches is performed to demonstrate the advantages of our approach. We also show that the proposed spectral feature can be used as a drop-in replacement for tasks involving a search for shape correspondence. All results presented in this study are obtained using an Intel Core i7-7700 processor with 4.2 GHz and 16 GB RAM. Offline training is run on an NVIDIA GeForce GTX 1080Ti (11 GB memory) GPU.

Evaluation metrics. We evaluate descriptor performance by using the standard *cumulative match characteristic* (CMC) and *Princeton protocol* (PP) [19]. CMC evaluates the probability of finding a correct match among $k-$nearest neighbors in the descriptor space. PP is a popular correspondence quality metric that measures correspondence quality by plotting the percentage of nearest-neighbor matches that are at most r-geodesically distant from the ground-truth correspondence.

6.1. Descriptor Evaluation

Invariant to resolution and triangulation. Literature [29] and [15] mentioned that frequency–domain features are robust to resolution and triangulation. In our configuration, we know that they affect the $\mathbf{T}$ and $\mathbf{A}$ matrices in Equation (4). To show that our proposed spectral feature is invariant to shape resolution and triangulation, we select four kinds of shapes, namely, skinny man, fat woman, fat man, and skinny girl for the experiments. Each kind of shape has five models with different resolutions and triangulations.
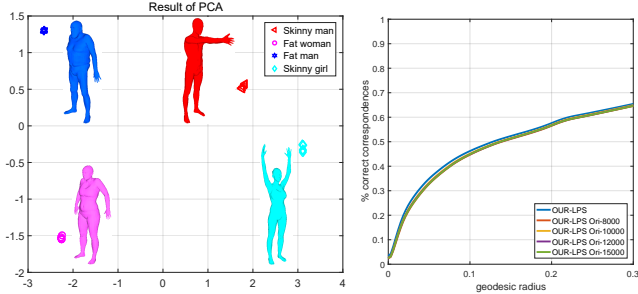
Figure 6. Impact of resolution and triangulation on the proposed LPS feature. Left: PCA results on the LPS feature. Right: shape matching results of multi-resolution using the LPS feature.
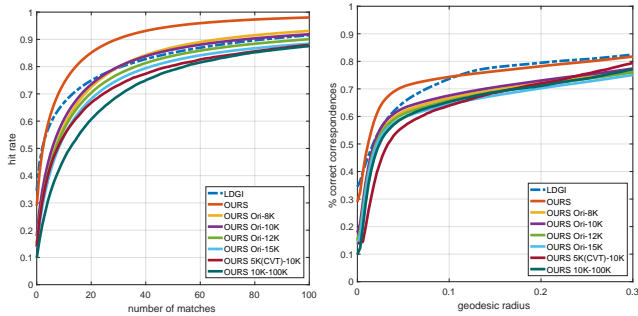

Figure 7. Impact of resolution and triangulation on our newly-learned descriptor. Left: CMC curve. Right: correspondence quality of the geodesic error.
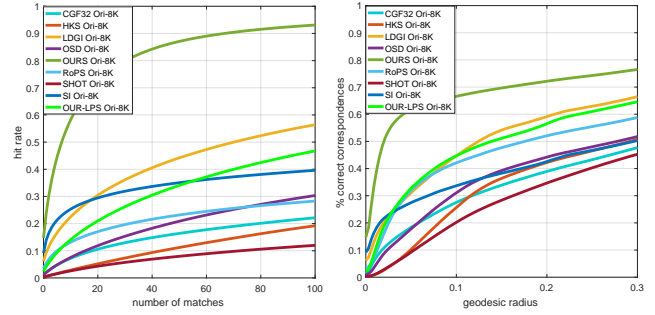

Figure 8. Performances of different descriptors for dense matching between resolution 6890 and 8K. Left: CMC curve. Right: dense correspondence quality of geodesic error.
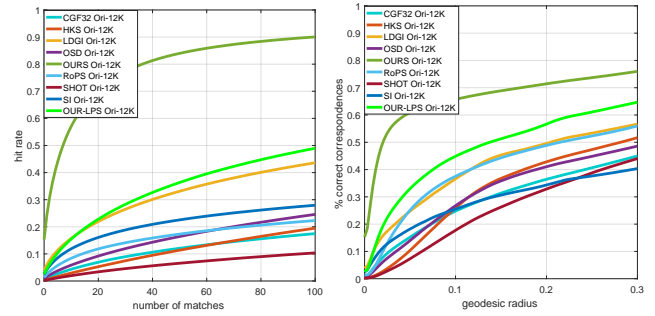

Figure 9. Performance of different descriptors in dense matching between the original 6890 and 12K resolutions. Left: CMC curve. Right: correspondence quality of the geodesic error.

We calculate the spectral feature with 16 dimensions for each specific model. Then, the classical *principal component analysis* (PCA) projection of the features is computed. The dimension reduction results are visualized in Fig. 6. The features of the five resolutions belonging to the same category are clustered together, indicating that our spectral feature is not sensitive to spatial resolution and triangulation. The right of this figure also shows the shape-matching results obtained using our unlearned LPS feature, further demonstrating the robustness of the proposed spectral feature.

The above experiment illustrates the resolution and triangulation invariance of our designed spectral feature. Next, we demonstrate the discriminative power and robustness of our newly learned shape descriptor to establish dense matching between multi-resolution shapes. Accordingly, a challenging task is conducted, whereby we compute local descriptors on low-resolution and high-resolution shapes, and the descriptors are used for shape matching. The deep neural network is trained by only using the original FAUST shapes with 6890 points. Fig. 7 shows the matching results, in which "OURS" means shape matching between original shapes and "OURS Ori-8K" means shape matching between the original shape and a high-resolution shape with 8K points. Our algorithm is applied to different resolutions, especially two challenging tests "10K-100K" and "5K(CVT)-10K". In the test with CVT models, we select

the points with the nearest geodesic distance as the ground-truth correspondence. Furthermore, given that we use the same triplet network in the method of the local descriptor using GIs (LDGI) [32], we also show its matching result. This figure shows that our method is better than LDGI when testing on the original FAUST shapes. When our method is trained at low resolution and tested on high resolution, the results show that no significant performance degradation exists.

We further evaluate the performance of our method by comparing it with various state-of-the-art techniques. Many recent deep-learned methods (such as MoNet [26] and FM-Net [22]) can only be trained and tested at the same resolution because of their fixed network structure. Therefore, these methods are unsuitable for matching tasks between different resolutions. In this test, three learned descriptors (OSD [23], CGF32 [18], and LDGI [32]) and four hand-crafted alternatives (SI [17], SHOT [31], RoPS [14], and HKS [30]) are selected for comparison. All these methods can perform shape matching with different resolutions. Figs. 8. and 9 show the matching results between the original shape with 8k and 12k resolutions, respectively. Other comparisons and analyses are provided in the supplemental materials. All the learned descriptors are trained on the original FAUST dataset and applied to high resolutions. Previous methods have difficulty handling different resolu-
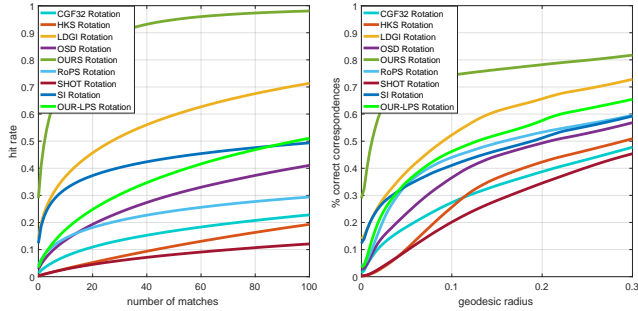
Figure 10. Performance of different descriptors on randomly rotated models. Left: CMC curve. Right: correspondence quality of the geodesic error.

Figure 11. Performance of different descriptors on randomly scaled models. Left: CMC curve. Right: correspondence quality of the geodesic error.

tions, whereas the newly learned descriptor exhibits excellent performance.

Invariant to scale and rotation. Scale and rotation invariance is critical in shape matching because 3D surfaces have many unique spatial factors, such as translation, rotation, and scaling. Many approaches need to normalize the model before processing, but our method is theoretically guaranteed to have no effect. Figs. 10 and 11 show the comparison results obtained on shapes with different rotations and scalings. As shown, our approach significantly outperforms state-of-the-art methods.

6.2. Shape Correspondence

Obtaining the shape correspondence is a different task from the construction of the local descriptor. Although computing the similarity of local descriptors helps in identifying corresponding points, recent learning-based approaches, such as GCNN [24] and MoNet [26], demonstrate better performance by considering the shape correspondence as a labeling problem. To demonstrate that our proposed local spectral feature can be used to compute the shape correspondence, we compare it with GCNN [24] and MoNet [26]. Our approach is tested in two configurations. "GCNN(OURS)" refers to the use of the GCNN framework, but the input feature in GCNN is our proposed LPS feature. This configuration shows that LPS can be used as a drop-in replacement for any existing local feature. In the "OURS-CORR" configuration, we utilize the ConvNet used for our descriptor learning but change the triplet loss to a cross-entropy loss for tagging. Fig. 12 shows the performance of shape correspondence on the FAUST dataset. Both configurations are competitive with existing state-of-the-art methods.

7. Conclusions

We presented a novel local point signature for shape matching of incompatible shape structures. We experimentally demonstrated that this spectral feature is robust to resolution and triangulation, and its invariance to rigid transformations (scale and rotation) can be proven theoretically.
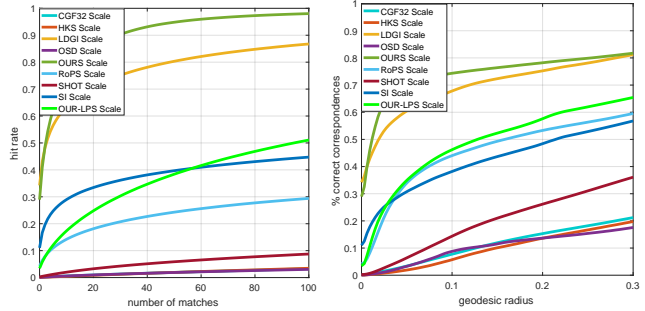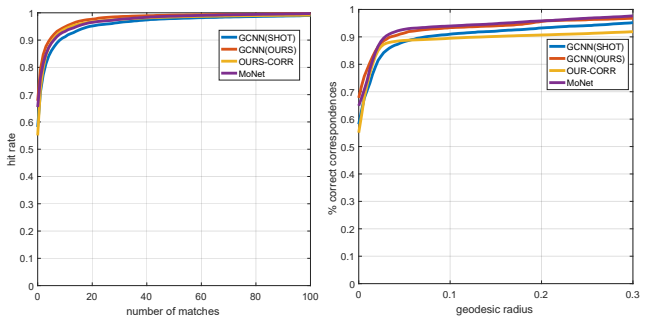

Figure 12. Performance of shape correspondence on the original FAUST dataset.

We drew inspiration from 2D Fourier transform of an image and introduced the local vertex spectral image to encode the information of LPS features in the local region. Then, a deep metric learning model using vertex spectral images was applied to compute local shape descriptors. We also provided a new benchmark dataset for evaluation by extending the well-known FAUST dataset. We demonstrated the advantages of our method in shape matching applications with incompatible shape structures that have not been seen in training data.

Limitation and future work. Given that the eigen-decomposition of every local vertex must be computed, taking a longer time than full eigen-decomposition on the entire model is reasonable. However, this process can be easily parallelized. We will further improve efficiency in the future. Moreover, we will extend our framework to handle point cloud data for a wide range of applications by considering different calculations of the Laplace–Beltrami operator.

References

[1] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. Scape: shape completion and animation of people. *ACM Trans. on Graphics (Proc. SIGGRAPH)*, 24(3):408–416, 2005.

[2] M. Aubry, U. Schlickewei, and D. Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. In *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 1626–1633. IEEE, 2011.

[3] F. Bogo, J. Romero, M. Loper, and M. J. Black. Faust: Dataset and evaluation for 3d mesh registration. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 3794–3801, 2014.

[4] D. Boscaini, J. Masci, S. Melzi, M. M. Bronstein, U. Castellani, and P. Vandergheynst. Learning class-specific descriptors for deformable shapes using localized spectral convolutional networks. *Computer Graphics Forum*, 34(5):13–23, 2015.

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

[6] D. Boscaini, J. Masci, E. Rodolà, M. M. Bronstein, and D. Cremers. Anisotropic diffusion descriptors. *Computer Graphics Forum*, 35(2):431–441, 2016.

[7] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. *Numerical geometry of non-rigid shapes*. Springer Science & Business Media, 2008.

[8] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

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

[10] M. Dunyach, D. Vanderhaeghe, L. Barthe, and M. Botsch. Adaptive remeshing for real-time mesh deformation. In *Eurographics*. The Eurographics Association, 2013.

[11] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik. Recognizing objects in range data using regional point descriptors. In *European Conference on Computer Vision (ECCV)*, pages 224–237. Springer, 2004.

[12] X. Gu, S. J. Gortler, and H. Hoppe. Geometry images. *ACM Trans. on Graphics*, 21(3):355–361, 2002.

[13] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, J. Wan, and N. M. Kwok. A comprehensive performance evaluation of 3d local feature descriptors. *Int. Journal of Computer Vision*, 116(1):66–89, 2016.

[14] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, and J. Wan. Rotational projection statistics for 3d local surface description and object recognition. *Int. Journal of Computer Vision*, 105(1):63–86, 2013.

[15] J. Hu and J. Hua. Salient spectral geometric features for shape matching and retrieval. *The Visual Computer*, 25(5-7):667–675, 2009.

[16] H. Huang, E. Kalogerakis, S. Chaudhuri, D. Ceylan, V. G. Kim, and E. Yumer. Learning local shape descriptors from part correspondences with multiview convolutional networks. *ACM Trans. on Graphics*, 37(1):6, 2018.

[17] A. E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(5):433–449, May 1999.

[18] M. Khoury, Q.-Y. Zhou, and V. Koltun. Learning compact geometric features. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 153–61, 2017.

[19] V. G. Kim, Y. Lipman, and T. Funkhouser. Blended intrinsic maps. *ACM Trans. on Graphics*, 30(4):79:1–79:12, July 2011.

[20] R. B. Lehoucq, D. C. Sorensen, and C. Yang. *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*, volume 6. SIAM, 1998.

[21] I. Lim, A. Dielen, M. Campen, and L. Kobbelt. A simple approach to intrinsic correspondence learning on unstructured 3d meshes. In *European Conference on Computer Vision Workshops (ECCV Workshops)*, pages 349–362. Springer, 2018.

[22] O. Litany, T. Remez, E. Rodolà, A. M. Bronstein, and M. M. Bronstein. Deep functional maps: Structured prediction for dense shape correspondence. In *IEEE International Conference on Computer Vision (ICCV)*, pages 5660–5668, 2017.

[23] R. Litman and A. M. Bronstein. Learning spectral descriptors for deformable shape correspondence. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 36(1):171–180, 2014.

[24] J. Masci, D. Boscaini, M. Bronstein, and P. Vandergheynst. Geodesic convolutional neural networks on riemannian manifolds. In *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 37–45, 2015.

[25] M. Meyer, M. Desbrun, P. Schröder, and A. H. Barr. Discrete differential-geometry operators for triangulated 2-manifolds. In *Visualization and mathematics III*, pages 35–57. Springer, 2003.

[26] F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, and M. M. Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[27] J. Ren, A. Poulenard, P. Wonka, and M. Ovsjanikov. Continuous and orientation-preserving correspondences via functional maps. *ACM Trans. on Graphics*, 37(6):248:1–248:16, Dec. 2018.

[28] M. Reuter, F.-E. Wolter, and N. Peinecke. Laplace–beltrami spectra as 'Shape-DNA' of surfaces and solids. *Computer-Aided Design*, 38(4):342–366, 2006.

[29] R. M. Rustamov. Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Proc. of Symp. of Geometry Processing*, pages 225–233. Eurographics Association, 2007.

[30] J. Sun, M. Ovsjanikov, and L. J. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *Computer Graphics Forum*, 28(5):1383–1392, 2010.

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

[32] H. Wang, J. Guo, D.-M. Yan, W. Quan, and X. Zhang. Learning 3d keypoint descriptors for non-rigid shape matching. In *European Conference on Computer Vision (ECCV)*, pages 3–20. Springer, 2018.

[33] Y. Wang, D.-M. Yan, X. Liu, C. Tang, J. Guo, X. Zhang, and P. Wonka. Isotropic surface remeshing without large and small angles. *IEEE Trans. on Vis. and Comp. Graphics*, 2018.

[34] L. Wei, Q. Huang, D. Ceylan, E. Vouga, and H. Li. Dense human body correspondences using convolutional networks. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1544–1553, 2016.

[35] D.-M. Yan, G. Bao, X. Zhang, and P. Wonka. Low-resolution remeshing using the localized restricted voronoi diagram. *IEEE Trans. on Vis. and Comp. Graphics*, 20(10):1418–1427, 2014.

[36] S. You and D. Zhang. Think locally, fit globally: Robust and fast 3d shape matching via adaptive algebraic fitting. *Neurocomputing*, 259:119–129, 2017.

[37] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud. Surface feature detection and description with applications to mesh matching. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 373–380. IEEE, 2009.

[38] A. Zeng, S. Song, M. Nießner, M. Fisher, J. Xiao, and T. Funkhouser. 3dmatch: Learning local geometric descriptors from rgb-d reconstructions. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 199–208. IEEE, 2017.