Multi-modality Learning for Non-rigid 3D Shape Retrieval via Structured Sparsity Regularizations

Luqing Luo, Lulu Tang, Rui Liu, Xiaoli Zhang, and Zhi-Xin Yang*, *Member, IEEE*.

***Abstract*— Big challenges are usually occurring in non-rigid 3D shape retrieval, for the shapes undergoing arbitrarily non-affine transformations. In this work, a novel design of feature learning approach is proposed for non-rigid 3D shape retrieval, dubbed Structured Sparsity Regularized Multi-Modality Method (SSR-MM). The shape signatures which capture the deformation-invariant characteristics are averaged and stacked for a multi-modality machine learning approach, and a transform matrix based on the structure sparsity regularization is utilized to map those signatures obtaining the discriminative features for retrieval. The proposed framework is evaluated on the publicly available non-rigid 3D human benchmarks, and the experimental results show the efficacy of our contributions and the advantages of our method over existing ones.**

***Index Terms*— 3D shape retrieval, multi-modality learning, non-rigid shapes**

I. Introduction

As an effective way to describe the physical objects, 3D shape representations are burgeoning in a variety of fields, such as computer vision, geometry processing, and shape analysis [29], [38]. Among the broad range of applications, 3D shape retrieval is one of the most important tasks aiming to seek a set of shapes similar to given query in dataset. For past decades, enormous attention has been paid to the retrieval task, including for rigid shapes and non-rigid shapes. Different from the rigid shapes, which cannot be pushed to make another shape, retrieving 3D non-rigid shapes is particularly challenging for there is a certain degree of flexibility coming from the natural properties. That is to say, all kinds of deformations of a non-rigid 3D shape yield different shapes that look dissimilar. Fortunately, although a non-rigid 3D shape can assume abound of variant forms via isometric deformations, the intrinsic geometry of shape is preserved among the variants. Thereby, the key to understanding the 3D non-rigid shapes is being able to analyze their characteristics and describe their behavior [5].

To this end, it is desirable to derive the deformation invariant component which characterizes the object by remaining consistent with deformations. As a result, similarity between shapes is considered as an effective criterion by identifying the attributions of deformation. Meanwhile, as an essential clue for the retrieval task, the similarity between shapes also plays a crucial role in 3D shape semantic learning. Comparing similarity between shapes can be paraphrased as a problem of looking for a quantitative measure of "distance" between two shapes [5]. Under such a metric, the shapes can be considered similar when the distance is small. This inspires us to resort to an efficient shape descriptor, which is evaluated by "distance", to capture the informative features of 3D shapes and keep invariant to isometric transformations or topological noises.

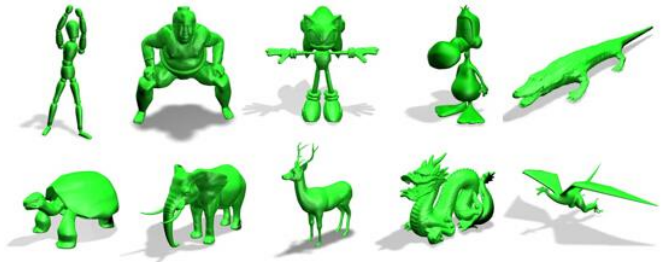


Fig. 1. Examples of the non-rigid 3D shapes from the SHREC dataset.

In this work, we are motivated to develop a retrieval solution for the 3D non-rigid shapes by leveraging appropriate shape descriptors being invariant under isometric transformations. Following the analyzing framework of multi-modal learning [33], we identify factors critical to the objective by obtaining a mapping to project the shape descriptors to the transformed space, where similar shapes with short distances are clustered together, and different shapes are dispersed farther in light of long distances. To implement the motivation, we propose the Structured Sparsity Regularized Multi-Modality Method (SSR-MM), which leverages the stacked and averaged shape signatures by multi-modality machine learning and obtains the discriminative features by the weight matrix with structured sparsity regularizations. The ablation studies

Manuscript received May 3, 2021. This work was funded in part by the Science and Technology Development Fund, Macau SAR (Grant no. 0018/2019/AKP, 0008/2019/AGJ, and SKL-IOTSC-2021-2023), in part by the Ministry of Science and Technology of China (Grant no. 2019YFB1600700), in part by the Guangdong Basic and Applied Basic Research Foundation (Grant no. 2020B1515130001).

Luqing Luo, Lulu Tang, and Zhi-Xin Yang are with State Key Laboratory of Internet of Things for Smart City, University of Macau, Macau SAR, 999078, P. R. China (e-mail: gabrielle.luo@connect.um.edu.mo; lulu.tang@connect.um.edu.mo; zxyang@um.edu.mo).

Rui Liu is with The cognitive robotics and AI lab (CRAI), College of Aeronautics and Engineering, Kent State University, Ohio, USA (e-mail: ruiliu.robotics@gmail.com).

Xiaoli Zhang is with the Department of Mechanical Engineering, Colorado School of Mines, Golden, USA (e-mail: xlzhang@mines.edu).

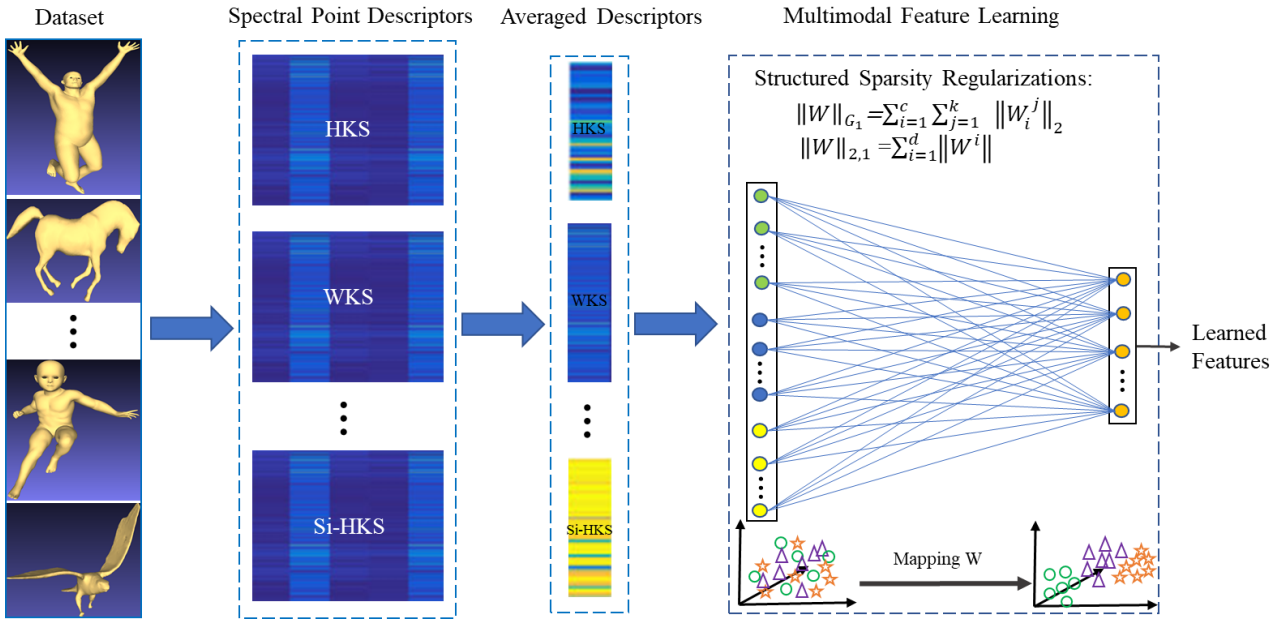


Fig. 2. The framework of the proposed SSR-MM method: a collection of labeled 3D shapes, spectral point descriptors, weighted average shape signatures, and multimodal feature learning with structure sparsity regularizations.

are conducted to confirm the efficacy of our novel designs. Comparative experiments with existing methods show the quantitative and qualitative advantages of our method in terms of acquiring discriminative features for effective retrieval.

The rest of this paper is organized as follows. The related works and basics of involved methods are briefly introduced in section II. The proposed method, Structured Sparsity Regularized Multi-Modality feature learning (SSR-MM) for 3D shape retrieval is explained in detail in section III. The experiments and analysis are discussed in section IV, and the whole study is remarked in section V.

II. Related Works

A. Non-rigid 3D Shape Descriptors

For the tasks of 3D shape retrieval and the other 3D shape analysis, various machine learning methods are developed to learn the optimal point descriptors [8], [9], [11], [19], [25], [35], [37]. For example, Bag-of-features (BoF) assigns the point descriptors to the nearest cluster center by K-means clustering and encodes the shapes as the histogram over the visual vocabulary constructed by the cluster centers [4]. The Geodesic-Aware Bags of Features (GA-BoF) uses a geodesic exponential kernel to substitute the heat kernel in the Spatially Sensitive Bags of Features (SS-BoF) for 3D shape representation [7]. A supervised learning method is introduced to generate the sparse representations for 3D shapes by obtaining Bag-of-features (BoF) signatures via a dictionary of point descriptors [18]. Spectral Graph Wavelets (SGW) constructs a global descriptor by aggregating mid-level features, which are obtained by embedding local descriptors from the spectral graph wavelet transform, resulting in a matrix representation describing the frequency of appearance of nearby codewords in the vocabulary [20]. A multi-stage feature learning scheme is developed to learn high-level shape features from mid-level features, which are generated via the BoF model measuring the spatial relationship between each pair of the BoF descriptors based on the SGW [13]. However, BoF series methods suffer from high computational costs, preventing them to apply to the large-scale dataset.

Meanwhile, a couple of shape descriptors based on the spectral-geometric methods for non-rigid 3D shape retrieval are prosperous [17]. Shape-DNA is one of the global shape descriptors defined on the entire 3D shape based on the Laplace-Beltrami operator (LBO), which utilizes a truncated sequence of eigenvalues as numerical signatures for a given shape [24]. Shape-DNA realizes to recognize the 3D shapes in different poses but struggles to distinguish the compact non-isometric shapes in the same spectra. Global Point Signature (GPS) [28] addresses the issue by integrating the eigenvalues and eigenvectors of each vertex over the entire shape but leaves the problem of sign correction to the eigenfunctions. Heat Kernel Signature (HKS) [32] devises an exponentially-weighted combination of the LBO eigenfunctions based on the Laplacian eigenbasis to alleviate the problem, meanwhile brings in the merits of invariant to isometric transformations and robust to small perturbations. Moreover, an extended version of HKS, Scale-Invariant Heat Kernel Signature (SIHKS) is proposed to guarantee the invariance for multiple shape scales, which is defined on a logarithmically sampled scale-space based on the Fourier transforms magnitude [6]. From the perspective of Fourier Transform, HKS is dominated by features related to low frequencies, while WKS [1] attempts to capture the local structure around the neighborhood, so as to learn high-frequency features with more fine-grained details of the non-rigid 3D shapes. Inspired by quantum mechanics, with certain energy distribution, WKS indicates the average probability to locate a particle at the surface. However, most of

the existing 3D shape descriptors employ one individual shape signature only, lacking diverse information contained from the combinations, especially for those signatures involving different aspects of characteristics.

B. Multimodal Feature Learning

Multi-modality feature learning has been widely used in heterogeneous data integration and multiple data selection. Since different representations of the same data source give rise to different kernel functions, Multiple Kernel Learning (MKL) [14], [31] assembles the different kernels toward a specific application efficiently. Several variants of MKL are proposed and a survey [12] investigates their performance in computer vision tasks. However, MKL series methods train the weights for each type of feature and assign all the features from the same type to be the same, which makes the method error-prone dealing with combining multiple types of features.

Sparse Multimodal Learning (SMML) [33] addresses the issue by assigning different importances for different features. The heterogeneous features from the same data source are integrated by structured sparsity regularizations to learn a sparse weight matrix, which attempts to explore both individual and group-wise correlations of each feature for different classes for feature selection.

III. The Proposed Method

A. Theoretical Preliminaries

In this section, Riemannian manifold [3], [15] and Laplace-Beltrami Operator [10], [26] are introduced briefly, which are the theoretical preliminaries of the proposed method.

1) Riemannian Manifold: A Riemannian manifold $(\mathcal{M}, g)$ is a smooth manifold $\mathcal{M}$ endowed with a Riemannian metric g. If $\mathcal{M}$ is N-dimensional differentiable, the tangent space $T_x(\mathcal{M})$ consists of all possible tangent vectors at point x on the manifold, and the tangent bundle $T(\mathcal{M})$ combines all the tangent spaces, which is $T(\mathcal{M}) = \bigcup_{x\in\mathcal{M}} T_x(\mathcal{M})$. The positive-definite inner product $g_x = \langle\cdot,\cdot\rangle_x : T_x(\mathcal{M}) \times T_x(\mathcal{M}) \to \mathbb{R}$ is defined on the tangent space for each $x \in \mathcal{M}$, and the family of smoothly varying g_x on $T(\mathcal{M})$ is denoted as the Riemannian metric g. Note that since g_x is a bilinear form on $T_x(\mathcal{M})$, it is an element of $T_x(\mathcal{M}) \times T_x(\mathcal{M})$. To say that g varies smoothly just means that g is a smooth section of the bundle $T(\mathcal{M})$.

Equipped with the manifold $\mathcal{M}$ and the Riemannian metric g, a variety of geometric notions on the manifold can be defined, such as the curvature, the angle between two curves, the gradients, and the length of a curve. 3D shapes can be represented by triangle meshes, and the triangle meshes can be defined on manifold as $\mathcal{M} = (V, E)$ or $\mathcal{M} = (V, F)$, where $V = \{v_1, v_2, ..., v_n\}$ is a set of vertices sampled from a two-dimensional Riemannian manifold , $E = \{e_1, e_2, ..., e_k\}$, $e_i \in V \times V$ is a set of edges connecting adjacent vertices, and $F = \{f_1, f_2, ..., f_m\}$, $f_i \in V \times V \times V$ is a set of triangles constructing the triangle meshes.

2) Laplace-Beltrami Operator: To measure the variation on Riemannian manifolds, the LaplaceBeltrami operator (LBO) is defined as the divergence of the gradient on $\mathcal{M}$, which is a second-order differential operator. Given a smooth and real-valued function $f : \mathcal{M} \to \mathbb{R}$ defined on $\mathcal{M}$, LBO is,

$$\triangle f = \mathrm{div}(\bigtriangledown f), \tag{1}$$

where div is the divergence operator, and $\bigtriangledown f$ is the gradient of f. $\triangle f$ measures the difference between $\triangle f$ and its average value in a small neighborhood, which helps to analyze geometry information on the manifold.

According to the Hodge Theorem for Functions [27], for a compact connected oriented Riemannian manifold $\mathcal{M}$, there exists an orthonormal basis consisting of eigenfunctions of the Laplacian,

$$\triangle\phi_i = \lambda_i\phi_i, \quad \lambda_i \in \mathbb{R}. \tag{2}$$

where the eigenvalues $\lambda_i \in \mathbb{R}$ are non-negative discrete set $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq +\infty$, and their corresponding eigenfunctions $\phi_i : \mathcal{M} \to \mathbb{R}$ are often viewed as real functions whose domain is the set of vertices, which can be devised as a smoothest orthonormal eigenbasis, satisfying:

$$\langle\phi_i, \phi_j\rangle = \int_{\mathcal{M}} \phi_i(x)\phi_j(x)dx = \begin{cases} 0, & \text{if} \quad i \neq j, \\ 1, & \text{if} \quad i = j. \end{cases} \tag{3}$$

LBO has been successfully used in various computer graphics applications, such as shape analysis, synthesis, and correspondence. More detailed discussion about the properties of LBO can be found in [23], [34].

B. Structured Sparsity Regularized Multi-Modality Method (SSR-MM) for Learning Non-rigid 3D Shapes

1) Multi-modality Data Integration: To learn non-rigid 3D shape representation, the shapes are modeled as a Riemannian manifold. However, the underlying manifold is usually unknown, thus the heat conduction properties are used as the features to describe shapes. Heat Kernel Signature (HKS) is one of the shape descriptors with heat conduction property, which is derived from the heat diffusion on a compact Riemannian manifold possibly with boundaries. Take a two-dimensional manifold $\mathcal{M}$ as an example, the heat diffusion process on the shape is governed by the heat equation,

$$\left(\frac{\partial}{\partial t} + \Delta_y\right) k(t, x, y) = 0, \tag{4}$$

where Δ_y denotes the positive semi-definite LBO on $\mathcal{M}$, and for any function $f(\cdot) \in \mathcal{M}$, it satisfies,

$$\lim_{t\to 0} \int_{\mathcal{M}} k(t, x, y)f(y)dy = f(x). \tag{5}$$

Let ϕ_i be an orthonormal basis of $\mathcal{M}$ satisfying $\triangle\phi_i = \lambda_i\phi_i$. With equality on $\mathcal{M}$ in the variable y by fixing t and x, we have,

$$k(t, x, \cdot) = \sum f_i(t, x)\phi_i(\cdot). \tag{6}$$

Therefore, $f_i(t,x) = \int_{\mathcal{M}} k(t,x,y)\phi_i(y)dy$. According to (4) and (6), the derivative of $f_i(t,x)$ with respect to time t is,

$$\begin{aligned}\frac{\partial}{\partial t} f_i(t,x) &= \frac{\partial}{\partial t}\int_{\mathcal{M}} k(t,x,y)\phi_i(y)dy \\ &= -\int_{\mathcal{M}} \Delta_y k(t,x,y)\cdot\phi_i(y)dy \\ &= -\int_{\mathcal{M}} k(t,x,y)\Delta_y\phi_i(y)dy \\ &= -\lambda_i\int_{\mathcal{M}} k(t,x,y)\phi_i(y)dy \\ &= -\lambda_i f_i(t,x)\end{aligned} \tag{7}$$

and the solution of (7) is,

$$f_i(t,x) = \omega_i(x)e^{-\lambda_i t}. \tag{8}$$

Suppose an arbitrary function $f(\cdot)\in\mathcal{M}$ is $f=\sum a_i\phi_i$, by substituting (6) and (8) into (5), we have,

$$\begin{aligned}f(x) &= \lim_{t\to 0}\int_{\mathcal{M}} k(t,x,y)f(y)dy \\ &= \lim_{t\to 0}\int_{\mathcal{M}}\sum_i e^{-\lambda_i t}\omega_i(x)\phi_i(y)\sum_j a_j\phi_j(y)dy \\ &= \lim_{t\to 0}\sum_i e^{-\lambda_i t}\omega_i(x)a_i \\ &= \sum_i \omega_i(x)a_i,\end{aligned} \tag{9}$$

which implies $\omega_i(x)=\phi_i(x)$. Thus the heat kernel $k(t,x,y)$ can be expressed as,

$$k(t,x,y) = \sum_i e^{-\lambda_i t}\phi_i(x)\phi_i(y), \tag{10}$$

on $\mathcal{M}$ in the y variable for fixed t and x. As a result, there exists a sequence $k\to\infty$ such that,

$$\sum_{k=0}^{\infty} e^{-\lambda_i t}\phi_i(x)\phi_i(y) \to k(t,x,y), \tag{11}$$

pointwise for any t, x, and almost all y. The heat kernel can be considered as the amount of heat transferred from x to y in time t by given a unit source at x. By setting $y=x$, the Heat Kernel Signature (HKS) is formed as,

$$h(x,t) = k(t,x,x) = \sum_{i=0}^{\infty} e^{-\lambda_i t}\phi_i(x)^2, \tag{12}$$

which represents the amount of heat remaining at point x after time t. The HKS descriptor at point x is obtained by concatenating the solutions of different times $\{t_1,t_2,\cdots,t_n\}$,

$$\mathrm{HKS}(x) = \left(h(x,t_1), h(x,t_2),\cdots,h(x,t_n)\right)^\top. \tag{13}$$

as a discrete analogue of the heat equation.

HKS plays a fundamental role in geometric modeling and shape analysis, the heat kernel is the isometric invariant of the Riemannian metric, making it effective to analyze shapes undergoing isometric deformations, especially for matching non-rigid 3D shapes. Besides, there are a couple of merits of HKS, such as informativeness, stability against noise, and multi-scale representations. The informativeness indicates that HKS fully characterizes shapes up to isometry by including all the information of intrinsic geometry of the shape. Moreover, as a multi-scale signature, it represents the local structure of the surface around point x within a small amount of t, while it reflects the global properties of the surface within a large amount of t [32]. However, HKS is sensitive to scales, resulting in scale-invariant HKS (SIHKS) to be adopted.

Suppose there is a 3D shape $\mathcal{M}$ and its scaled version $\mathcal{M}'=\alpha\mathcal{M}$, the eigen decomposition of shape $\mathcal{M}'$ satisfies $\lambda'=\alpha^2\lambda$ and $\phi'=\alpha\phi$. According to (13), HKS of the scaled version $\mathcal{M}'$ is,

$$h'(x,t) = \sum_{i=0}^{\infty} e^{-\alpha^2\lambda_i t}\alpha^2\phi_i(x)^2 = \alpha^2 k(x,\alpha^2 t). \tag{14}$$

in which the influence of scale factor α should be removed to achieve the scale-invariance. Firstly, by rewriting time t as β^τ, we have $k_\tau = k(x,\beta^\tau)$. According to (14), scaling the shape by α leads to amplitude changing by α^2 and time-shifting by $s=2\log_\beta\alpha$, then $k'_\tau=\alpha^2 k_{\tau+s}$. By imposing logarithm on both sides, we get,

$$\log(k'_\tau) = 2\log\alpha + \log(k_{\tau+s}). \tag{15}$$

Taking the discrete derivative with respect to τ, the constant α^2 can be removed,

$$\dot{h}'_\tau = \dot{h}_{\tau+s}. \tag{16}$$

Moreover, taking the Fourier transform of (16), we get,

$$H'(\omega) = H(\omega)e^{j\omega s} \quad \omega\in[0,2\pi], \tag{17}$$

where H' and H are the discrete-time Fourier transform of $\dot{h}$ and $\dot{h}'$, respectively. Taking the absolute value of $H(\omega)$ and sampling $|H(\omega)|$ with p frequencies $\{\omega_1,\omega_2,\cdots,\omega_p\}$, SIHKS descriptor is formed as,

$$\mathrm{SIHKS}(x) = \left(|H(\omega_1)|, |H(\omega_2)|,\cdots,|H(\omega_p)|\right)^\top. \tag{18}$$

which is invariant to the scale [6].

With the perspective of graph Fourier, HKS [32] and SIHKS [6] are more sensitive to low-frequency components, which contain fine-grained details of the 3D shapes. To capture the overall information across the whole shape, both local and global features are seminal. As a result, Wave Kernel Signature (WKS) [2] is introduced to learn high-frequency information, by utilizing the Schrödinger equation governing the evolution of a quantum particle on the surface,

$$\left(\frac{\partial}{\partial t} + i\Delta\right)\psi(x,t) = 0, \tag{19}$$

where $\psi(x,t)$ is the wave function of the particle. Although the formula looks similar to the heat equation, heat diffusion is a process of dissipation, while the quantum particle follows

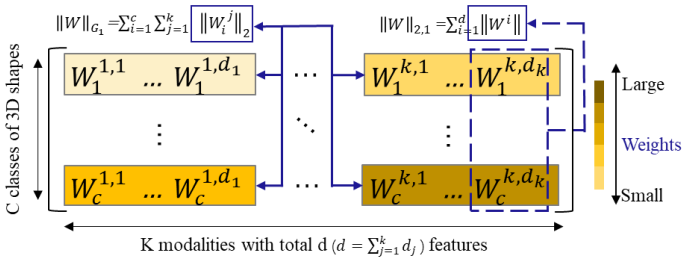


Fig. 3. Visualization of the weight matrix $\mathbf{W}$: suppose there are $\boldsymbol{c}$ classes and $\boldsymbol{k}$ modalities, $\boldsymbol{G_1}$-norm focuses on the group-wise weights for a specific type of features within each class, and $\boldsymbol{L_{21}}$-norm emphasizes individual weights cross multiple classes. Combining these two regularizes, discriminative views, and important features in non-discriminative views obtain significant weights.

a process of oscillations for its unknown positions on the surface. The solution to (19) can be expressed as,

$$\psi_E(x,t) = \sum_{i=0}^{\infty} e^{iE_k t}\phi_k(x) f_E(\lambda_k), \tag{20}$$

where E is the approximate measurement of the energy at time $t = 0$, and f_E is the energy distribution. Suppose the probability of the particle locating at point x is $|\psi_E(x,t)|^2$, the average probability during the time T under a certain energy distribution f_E is,

$$w(x) = \lim_{T\to\infty} \frac{1}{T}\int_0^T |\psi_E(x,t)|^2 = \sum_{k=0}^{\infty} \phi_k(x)^2 f_E(E_k)^2 \tag{21}$$

where the second equality is obtained in terms of the orthogonal property of the functions $e^{iE_k t}$, and $f_E(\cdot)^2$ is the energy probability distribution with expectation value E. This average probability $w(x)$ is defined as WKS and a discrete sequence of WKS are obtained by fixing a family of energy distributions $\{f_{e_1}, f_{e_2}, \cdots, f_{e_q}\}$ as,

$$\text{WKS}(x) = (w_{e_1}(x), w_{e_2}(x), \cdots, w_{e_q}(x))^\top. \tag{22}$$

Thus far, both global and local shape descriptors are acquired. Since each shape descriptor consists of thousands of vertices, a compact averaged shape signature C_f is computed all over the shape descriptors of each mesh by following the technique used in [9], as shown in Fig. 2. Moreover, it is often ambiguous to determine the best descriptor under a certain circumstance, since different shape descriptors include diverse and partly independent information. C_f are further set as the combinations of those averaged descriptors and a multi-modality learning approach is utilized to integrate the heterogeneous signatures C_f.

2) Multimodal Feature Learning: To make the shape signatures C_f more distinguishable, a mapping $\mathbf{W}$ is needed to project the signatures to a higher dimensional feature space, where the signatures of the same category are clustered closer, and different ones are dispersed farther, as shown in the multimodal feature learning module of Fig. 2.

Suppose there are n training 3D shapes $\{(x_i, y_i)\}_{i=1}^n$, where $x_i = (x_i^1, \cdots, x_i^k)^\top \in \mathbb{R}^d$ is the input data including all the features from k different modalities, $y_i \in \mathbb{R}^c$ is the corresponding label of c different classes. Each modality j corresponds to a feature d_j, where $d = \sum_{j=1}^k d_j$. Let input data $X = [x_1, \cdots, x_n] \in \mathbb{R}^{d\times n}$, label $Y = [y_1, \cdots, y_n] \in \mathbb{R}^{c\times n}$, we can define a weight matrix $\mathbf{W} = \left[\mathbf{w}_i^j\right]$, where each element $\mathbf{w}_i^j \in R^{d_j}$ denotes the contribution of all the features from j^{th} modality with respect to i^{th} class. Then we have $\mathbf{W}^\top X = Y$, which can be written as,

$$\begin{bmatrix} \mathbf{w}_1^1 & \cdots & \mathbf{w}_1^k \\ \vdots & \ddots & \vdots \\ \mathbf{w}_c^1 & \cdots & \mathbf{w}_c^k \end{bmatrix}_{c\times d} \begin{bmatrix} x_1^1 & \cdots & x_n^1 \\ \vdots & \ddots & \vdots \\ x_1^k & \cdots & x_n^k \end{bmatrix}_{d\times n} = [y_1 \cdots y_n]_{c\times n}. \tag{23}$$

To find an optimal mapping $\mathbf{W}$ for the shape signatures C_f, Sparse Multi-modal Learning (SMML) [33] is employed to integrate the multiple data. SMML selects discriminative features and leverages them by utilizing the Structured Sparsity Regularizations (SSR) to endow different importance for each data source. The general regularized objective function is expressed as,

$$\min_{\mathbf{W}} Q(\mathbf{W}) = \min_{\mathbf{W}} L(X, \mathbf{W}) + \gamma R(\mathbf{W}), \tag{24}$$

where $L(X, \mathbf{W})$ is the loss function $L(X, \mathbf{W}) = \left\|\mathbf{W}^\top X - Y\right\|_2^2$, γ is the regularization parameter, and $R(\mathbf{W})$ is the regularization term devised to distinguish different 3D shapes from multi-modality data, which is,

$$R(\mathbf{W}) = \|\mathbf{W}\|_{G_1} = \sum_{i=1}^{c}\sum_{j=1}^{k} \left\|\mathbf{w}_i^j\right\|_2. \tag{25}$$

The regularization term imposes sparsity and captures the global relationship between modalities by using G_1-norm. The weights of less discriminative features have very small values, while those of more discriminative features get larger values. In this way, the regularization term enhances the weights of features with more relevant modalities and suppresses those with less relevance. Besides, for certain categories, although most features from one modality are less discriminative, the significance of a few features remains high, and these features should be shared by all categories [21]. Thus, an additional L_{21}-norm $\|W\|_{2,1} = \sum_{i=1}^d \left\|w^i\right\|_2$ is added to (24) to ensure this, which is,

$$\min_{\mathbf{W}} Q(\mathbf{W}) = \min_{\mathbf{W}} L(X, \mathbf{W}) + \gamma_1 \|\mathbf{W}\|_{G_1} + \gamma_2 \|\mathbf{W}\|_{2,1}. \tag{26}$$

L_{21}-norm enforces sparsity between all features and non-sparsity between different classes, the larger weights to the discriminative features are assigned across the classes. The visualization and interpretation of the weight matrix is shown in Fig. 3 and the complete form of the objective function is,

$$\min_{\mathbf{W}} Q(\mathbf{W}) = \min_{\mathbf{W}} \left\|\mathbf{W}^\top X - Y\right\|_2^2 + \gamma_1 \sum_{i=1}^{c}\sum_{j=1}^{k} \left\|\mathbf{W}_i^j\right\|_2 + \gamma_2 \sum_{i=1}^{d} \left\|\mathbf{w}^i\right\|_2 . \tag{27}$$

3) Optimization and Convergence Analysis: To optimize the non-smooth regularizers in the objective function (27), an efficient algorithm is utilized to address the non-smooth problem [33]. Firstly, the value of W in (27) is equivalent to the following form,

$$\mathbf{W} = \arg\min_{\mathbf{W}} L(X, \mathbf{W}) + \gamma_1 \sum_{i=1}^{c} \mathbf{D}^i \left\|\mathbf{w}^i\right\|_2^2 + \gamma_2 \mathbf{tr}(\mathbf{W}^\top \tilde{\mathbf{D}} \mathbf{W}). \tag{28}$$

where $\mathbf{D}$ is a block diagonal matrix,

$$\mathbf{D} = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_c \end{pmatrix}, \tag{29}$$

with the j^{th} $(1 \le j \le c)$ diagonal element $d_j = \frac{1}{2\left\|\mathbf{w}_i^j\right\|_2} I_j$. Similarly, $\tilde{\mathbf{D}}$ is a diagonal matrix,

$$\tilde{\mathbf{D}} = \begin{pmatrix} \tilde{d}_1 & & \\ & \ddots & \\ & & \tilde{d}_c \end{pmatrix} \tag{30}$$

with the i^{th} $(1 \le j \le c)$ diagonal element $\tilde{d}_i = \frac{1}{2\|\mathbf{w}^i\|_2}$. By taking the derivatives of (28) with respect to $\mathbf{W}$, we get,

$$\frac{\partial L(X, \mathbf{W})}{\partial \mathbf{W}} + 2\gamma_1 \sum_{i=1}^{c} \mathbf{D}^i \mathbf{w}_i + 2\gamma_2 \mathbf{D}\mathbf{W} = 0, \tag{31}$$

which is equivalent as,

$$XX^\top \mathbf{W} - XY^\top + \gamma_1 \mathbf{D}\mathbf{W} + \gamma_2 \tilde{\mathbf{D}}\mathbf{W} = 0. \tag{32}$$

Finally, W can be obtained as,

$$\mathbf{W} = (XX^\top + \gamma_1 \mathbf{D} + \gamma_2 \tilde{\mathbf{D}})^{-1} XY^\top. \tag{33}$$

Note that $\mathbf{D}$ and $\tilde{\mathbf{D}}$ are both unknown variables and dependent on $\mathbf{W}$. An iterative algorithm is utilized to find the optimal $\mathbf{W}$, as summarized in Algorithm 1.

It can be proved that Algorithm 1 decreases the objective value of (28) by iterations, and the detailed proof can be found in Appendix I.

Algorithm 1 : W optimization

Input: $X = [x_1, \cdots, x_n] \in \mathbb{R}^{d \times n}$, $Y = [y_1, \cdots, y_n] \in \mathbb{R}^{c \times n}$.

Output: $\mathbf{W}$.

1. Let t=1. Initialize $\mathbf{W}_t \in \mathbb{R}^{d \times c}$.

2. While not converges **do**

3. Calculate the block diagonal matrix $\tilde{\mathbf{D}}_{t+1}^i$ according to (29), where the j^{th} diagonal element is $\frac{1}{2\left\|(\mathbf{w}_t)_i^j\right\|_2}\mathbf{I}_j$; Calculate the block diagonal matrix $\tilde{\mathbf{D}}_{t+1}$ according to (30), where the i^{th} diagonal element is $\frac{1}{2\left\|\mathbf{w}_t^i\right\|_2}$.

4. According to (33), for each $\mathbf{w}^i (1 \le i \le c)$, $\mathbf{w}_{t+1}^i = (XX^\top + \gamma_1 \mathbf{D}_{t+1}^i + \gamma_2 \tilde{\mathbf{D}}_{t+1})^{-1} XY^\top$.

5. $t = t + 1$.

6. End while

Return $\mathbf{W}_t \in \mathbb{R}^{d \times c}$.

IV. Experiments and Discussion

A. Experimental Setup

1) Dataset: The experiments of the proposed SSR-MM method are conducted on the SHREC'14 Human model database, which consists of two types of 3D human models, real models and synthetic models. The real model dataset is made up of scanned data from real human participants, containing 400 shapes subdivided into 40 human subjects with each subject consisting of 10 distinct poses. The synthetic dataset is built by DAZ studio, which includes 300 human models from 15 classes with each class containing 20 members. As shown in Fig. 4, the man, woman, and child models from the synthetic dataset are classified into different categories. This is the major challenge from the non-rigid 3D shapes, the categories are not naturally separated by different objects.

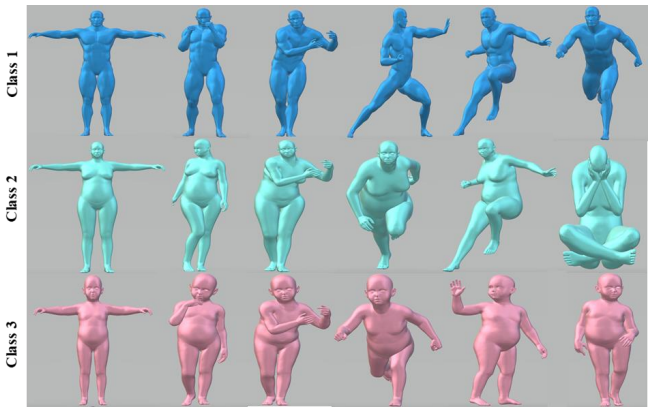


Fig. 4. SHREC'14 3D human models synthetic dataset.

2) Evaluation Metrics: To evaluate the proposed method, the same evaluation metrics from [22] are utilized. The dissimilarity between every two models is computed firstly to form a distance matrix, with the element (i, j) indicating the computed distance between models i and j. Then five quantitative measurements are employed to evaluate the retrieval performance of the proposed method by following [30].

1) The Precision-Recall curve represents the relationship between precision and recall, where precision (P) depicts the percentage of correctly retrieved models, and recall (R) indicates the percentage of retrieved relevant models.
2) The First-tier (FT) and Second-tier (ST) are the percentages of 3D models belonging to the query's group that appear in the first L matches, L relies on the scale of

TABLE I
COMPARISON OF RETRIEVAL RESULTS ON SHREC'14 3D HUMAN MODEL DATASETS.

Dataset	Methods	Shape Signature	FT(%)	ST(%)	NN(%)	E(%)	DCG(%)
SHREC'14 Real Dataset	Baseline	WKS	23.25	38.81	32.00	24.00	51.38
		HKS	61.33	81.19	72.00	40.90	81.13
		SIHKS	50.11	69.36	74.25	36.70	76.19
		WKS+HKS	52.44	72.00	64.00	37.39	75.15
		WKS+SIHKS	44.17	63.00	52.00	34.30	67.90
		WKS+HKS+SIHKS	54.15	74.17	64.75	38.68	76.65
	SSR-MM	WKS+HKS	82.67	96.31	93.75	43.40	95.14
		WKS+SIHKS	83.89	95.56	93.50	43.39	95.06
		WKS+HKS+SIHKS	**93.83**	**99.03**	**99.00**	**43.77**	**98.96**
SHREC'14 Synthetic Dataset	Baseline	WKS	73.53	94.35	94.33	66.16	93.56
		HKS	63.86	95.19	83.00	64.17	87.55
		SIHKS	71.05	90.33	84.67	63.29	91.21
		WKS+HKS	76.09	98.58	91.00	68.46	94.59
		WKS+SIHKS	77.26	96.16	90.33	67.11	94.58
		WKS+HKS+SIHKS	76.65	97.75	89.33	67.45	94.74
	SSR-MM	WKS+HKS	93.01	99.34	100	73.4	99.21
		WKS+SIHKS	95.24	99.73	99.93	73.98	99.50
		WKS+HKS+SIHKS	**96.82**	**99.76**	**99.93**	**74.13**	**99.68**

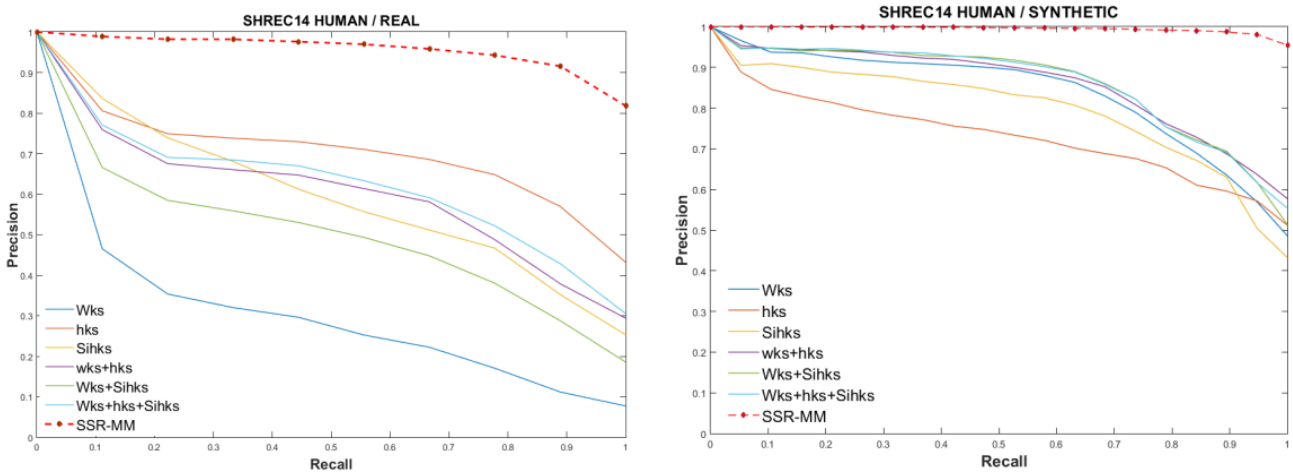


Fig. 5. Precision-Recall comparison on the SHREC'14 real dataset (left) and Synthetic dataset (right).

the query's group. Given a query's class with M models, $L = M - 1$ for FT, and $L = 2(M - 1)$ for ST.

3) The Nearest Neighbor (NN) denotes the percentage of the closest matchings that remain with the same class as the query.
4) The E-measure (E) is a composite measure of precision (P) and recall (R), which is defined as $E = \frac{2}{P^{-1}+R^{-1}}$. Higher values indicate better retrieval results.
5) The Discounted Cumulative Gain (DCG) is a measure of ranking quality reflecting the "gain" of the method based on its position in the result list, and DCG applies a discount factor to the relevance scores to devaluate late-retrieved gain [16].

Besides, the LBO for each model in the selected dataset is computed and its bases are truncated to the first 100 eigenfunctions in the experiment. The above-mentioned point descriptors HKS, SIHKS, and WKS are calculated and the dimensions are set as 56, 50, and 100, respectively. The dataset is split as 40% for training, 60% for testing, and 25% of the training data for validation. The validation set is used to search the optimal hyper-parameters for γ_1 and γ_2 in (31), which are in the range of $\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ by following the same configurations in [9] for fair comparisons. All the experiments are conducted on a Dell workstation with an Intel (R) Xeon (R) CPU (3.00GHz), 64GB RAM and one NVIDIA GPU with GeForce GTX 1080.

B. Experimental Results and Analysis

The proposed SSR-MM is compared with baselines on the SHREC'14 dataset of real models and synthetic models, respectively, as shown in Table I. According to the experiments, simply stacking shape signatures doesn't always improve retrieval performance. The descriptor combinations of WKS+HKS, WKS+SIHKS, and WKS+HKS+SIHKS sometimes get inferior results than that of individual descriptors WKS, HKS, and SIHKS in baselines. However, when the stacked descriptors are leveraged by the proposed SSR-MM

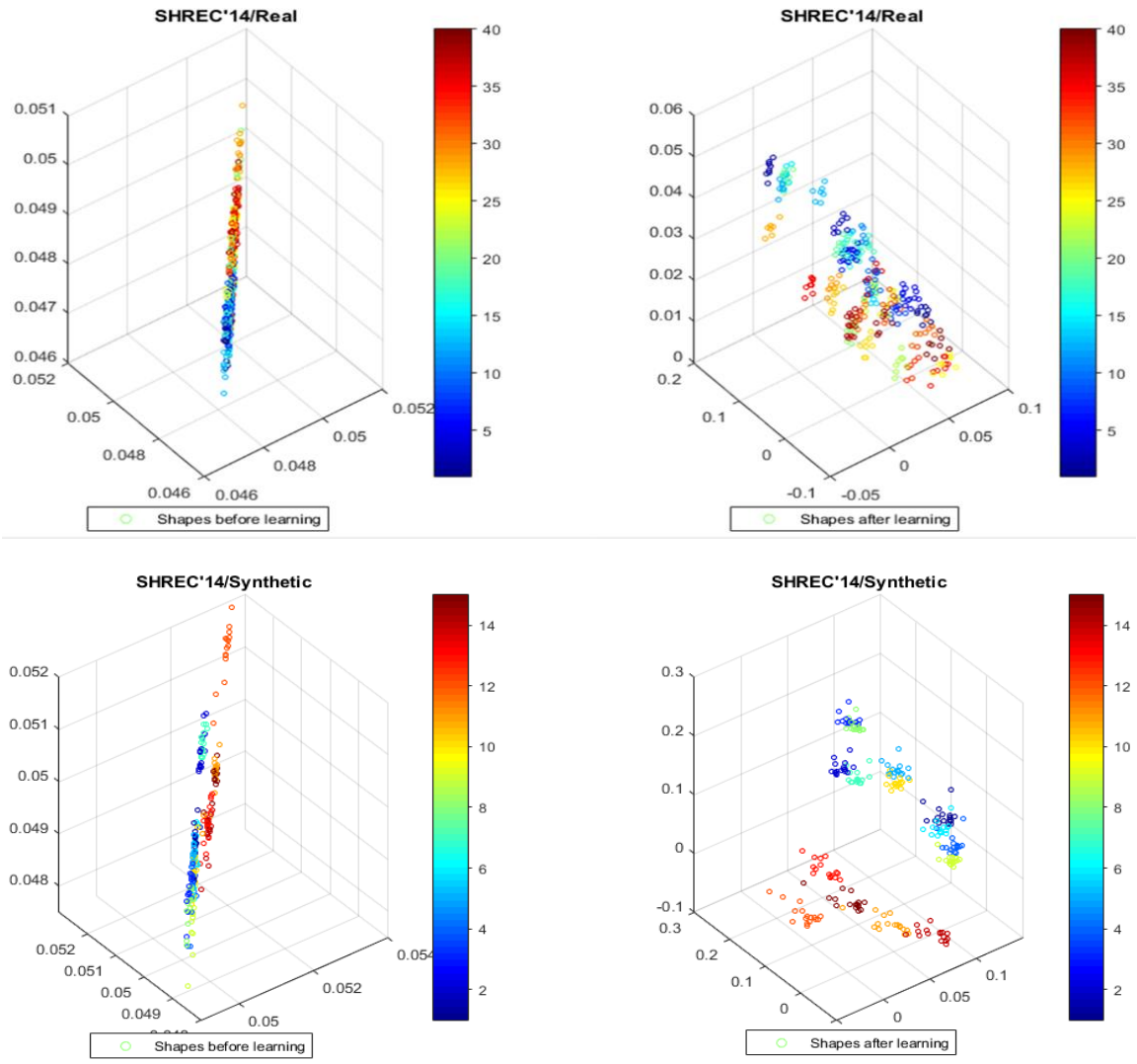


Fig. 6. Visualization of the shape signatures clustering before and after learning. The top row are 40 classes (different colors) from the SHREC'14 real dataset, and the bottom row shows 15 classes from the SHREC'14 Synthetic dataset. The proposed SSR-MM method clusters the shape signatures of the same category together and disperses the different ones apart.

TABLE II
COMPARISON OF RETRIEVAL METHODS IN TERMS OF MEAN AVERAGE PRECISION (MAP) ON SHREC'14 3D HUMAN MODELS DATASETS.

Methods	Synthetic (%)	Real (%)
ShapeGoogle (VQ) [4]	81.30	51.40
Unsupervised DL [18]	84.20	52.30
Supervised DL [18]	95.40	79.10
RMVM [11]	96.30	79.50
R-BiHDM-s [36]	64.20	64.00
CSD+LMNN [9]	99.67	97.92
SVM*	96.52	94.63
SSR-MM*	**99.94**	**98.30**

*The input shape signature is WKS+HKS+SIHKS.

method, the retrieval results surpass all the baselines by all metrics with a big margin on the very challenging SHREC'14 dataset. Among all the combinations, WKS+HKS+SIHKS obtains the best result on both real models and synthetic models. The performance of the proposed SSR-MM method confirms its capability to take advantage of additional diverse information in the stacked descriptors. Moreover, Fig. 5 illustrates the precision-recall curves of all methods on the two datasets, from which we can see that the proposed SSR-MM method (in red dashed line) outperforms all the baselines significantly. Particularly, the SHREC'14 3D Human dataset of real models is more challenging than that of the synthetic ones, please refer to the performance drop of the baselines between the synthetic models and the real ones. However, the proposed SSR-MM remains competitive on both datasets.

Table II shows the comparison of retrieval results of the proposed SSR-MM with the state-of-the-art methods, with the mean average precision (MAP) reported over five runs. SSR-MM achieves the best result comparing with the recent learning approaches. Especially, comparing SVM with the same input signature WKS+HSK+SIHKS, SSR-MM achieves an apparent better performance, which demonstrates that SSR-MM is good at exploring the efficient information contained

TABLE III
ABLATION STUDIES OF OUR PROPOSED SSR-MM BY USING WKS+HKS+SIHKS AS INPUT SHAPE SIGNATURE.

Dataset	Methods	FT(%)	ST(%)	NN(%)	E(%)	DCG(%)
SHREC'14 Real Dataset	Baseline	54.15	74.17	64.75	38.68	76.65
	loss	81.56	90.83	98.25	41.46	94.79
	G_1-norm	91.97	98.25	98.75	43.48	98.36
	L_{21}-norm	90.00	98.03	98.25	43.59	97.71
	SSR-MM	**93.83**	**99.03**	**99.00**	**43.77**	**98.96**
SHREC'14 Synthetic Dataset	Baseline	76.65	97.75	89.33	67.45	94.74
	loss	89.98	92.91	97.33	68.81	96.79
	G_1-norm	96.93	99.68	100	74.10	98.71
	L_{21}-norm	96.46	99.49	99.67	74.00	99.58
	SSR-MM	**96.82**	**99.76**	**99.93**	**74.13**	**99.68**

in multiple modalities.

To illustrate the effect of mapping W, Fig. 6 visualizes the clustering of the shape signatures, where the left column describes the signature distribution before learning, and the right column depicts the distribution after learning. The top row are 40 classes from the SHREC'14 real dataset, with each dot representing one model and one color corresponding to one category. The bottom row shows 15 classes from the SHREC'14 Synthetic dataset. The proposed SSR-MM method clusters the shape signatures of the same category together, and disperses the different ones apart. Especially for the very challenging dataset of real models, different categories are still well separated.

C. Ablation Study

To study the effect of our proposed objective function (28), which consists of the logistic regression loss, the G_1-norm, and the L_{21}-norm, an ablation study is conducted by separating each part. From Table III, we can see that the regularizes of G_1-norm and L_{21}-norm boost retrieval performance strikingly compared with the baselines of simply stacking signatures WKS+HKS+SIHKS. By combining both two regularization terms, further improvement is still obtained, which approves the necessity of both G_1-norm and L_{21}-norm in the proposed SSR-MM method.

V. CONCLUSION

In this work, we propose a novel design of Structured Sparsity Regularized Multi-Modality Method (SSR-MM) for non-rigid 3D shape retrieval. The point descriptors based on Laplace-Beltrami operator for capturing the deformation-invariant shape features are computed firstly, then the global shape signatures are constructed by a weighted average over a given shape. The multi-modality learning approach obtains a mapping for the discriminative features via joint Structured Sparsity Regularizations method for effective retrieval. Compared to state-of-the-arts, SSR-MM achieves a competitive performance and the experimental results demonstrate that the proposed SSR-MM is able to seek and select informative and effective features of non-rigid 3D shapes.

APPENDIX I

In Algorithm 1, the value of the objective function in (27) is monotonically decreased in each iteration, the detailed proof is summarized as below.

Proof: According to (27), in each iteration t, we have,

$$\mathbf{W}_{t+1} = \arg\min_{\mathbf{W}} L(X, \mathbf{W}_t) + \gamma_1 \sum_{i=1}^{c} \mathbf{D}^i \left\|\mathbf{w}_t^i\right\|_2^2 + \gamma_2 \mathbf{tr}(\mathbf{W}_t^\top \tilde{\mathbf{D}} \mathbf{W}_t). \tag{34}$$

In terms of Step 3 and 4 in Algorithm 1, we have,

$$\begin{aligned} &L(X, \mathbf{W}_{t+1}) + \gamma_1 \sum_{i=1}^{c} \mathbf{D}^i \left\|(\mathbf{w}_{t+1}^i)\right\|_2^2 + \gamma_2 \mathbf{tr}(\mathbf{W}_{t+1}^\top \tilde{\mathbf{D}} \mathbf{W}_{t+1}) \\ &\leq L(X, \mathbf{W}_t) + \gamma_1 \sum_{i=1}^{c} \mathbf{D}^i \left\|(\mathbf{w}_t^i)\right\|_2^2 + \gamma_2 \mathbf{tr}(\mathbf{W}_t^\top \tilde{\mathbf{D}} \mathbf{W}_t). \end{aligned} \tag{35}$$

Substituting $\mathbf{D}$ and $\tilde{\mathbf{D}}$ by definitions, we get:

$$\begin{aligned} &L(X, \mathbf{W}_{t+1}) + \gamma_1 \sum_{i=1}^{c} \sum_{j=1}^{k} \frac{\left\|(\mathbf{w}_{t+1})_i^j\right\|_2^2}{2\left\|(\mathbf{w}_t)_i^j\right\|_2} + \gamma_2 \sum_{i=1}^{d} \frac{\left\|\mathbf{w}_{t+1}^i\right\|_2^2}{2\left\|\mathbf{w}_t^i\right\|_2} \\ &\leq L(X, \mathbf{W}_t) + \gamma_1 \sum_{i=1}^{c} \sum_{j=1}^{k} \frac{\left\|(\mathbf{w}_t)_i^j\right\|_2^2}{2\left\|(\mathbf{w}_t)_i^j\right\|_2} + \gamma_2 \sum_{i=1}^{d} \frac{\left\|\mathbf{w}_t^i\right\|_2^2}{2\left\|\mathbf{w}_t^i\right\|_2}. \end{aligned} \tag{36}$$

For any vector A and B ($A \neq B$), we have $(\|A\|_2 - \|B\|_2)^2 \geq 0$, and the inequality can be expanded as,

$$\begin{aligned} &\|A\|_2^2 + \|B\|_2^2 - 2\|A\|_2 \|B\|_2 \geq 0 \\ &\Rightarrow 2\|A\|_2 \|B\|_2 - \|A\|_2^2 \leq \|B\|_2^2 \\ &\Rightarrow 2\|A\|_2 \|B\|_2 - \|A\|_2^2 \leq 2\|B\|_2^2 - \|B\|_2^2. \end{aligned} \tag{37}$$

Divide both sides by $2\|B\|_2$, we have the inequality:

$$\|A\|_2 - \frac{\|A\|_2^2}{2\|B\|_2} \leq \|B\|_2 - \frac{\|B\|_2^2}{2\|B\|_2}. \tag{38}$$

According to the inequality (38), by endowing $\|A\|_2 = \sum_{i=1}^{c}\sum_{j=1}^{k} \left\|(\mathbf{w}_{t+1})_i^j\right\|_2$ and $\|B\|_2 = \left\|(\mathbf{w}_t)_i^j\right\|_2$, we can derive,

$$\begin{aligned} &\sum_{i=1}^{c} \sum_{j=1}^{k} \left\|(\mathbf{w}_{t+1})_i^j\right\|_2 - \sum_{i=1}^{c} \sum_{j=1}^{k} \frac{\left\|(\mathbf{w}_{t+1})_i^j\right\|_2^2}{2\left\|(\mathbf{w}_t)_i^j\right\|_2} \\ &\leq \sum_{i=1}^{c} \sum_{j=1}^{k} \left\|(\mathbf{w}_t)_i^j\right\|_2 - \sum_{i=1}^{c} \sum_{j=1}^{k} \frac{\left\|(\mathbf{w}_t)_i^j\right\|_2^2}{2\left\|(\mathbf{w}_t)_i^j\right\|_2}, \end{aligned} \tag{39}$$

and by assigning $\|A\|_2 = \sum_{i=1}^{d} \left\|\mathbf{w}_{t+1}^i\right\|_2$, and $\|B\|_2 = \left\|\mathbf{w}_t^i\right\|_2$, we have,

$$obtainedhisPh.D.degreeinindustrialengineeringandengineeringma \tag{40}$$

By taking (36) + $\gamma_1\times$ (39) + $\gamma_2\times$ (40) on both sides, finally we get,

$$L(X, \mathbf{W}_{t+1}) + \gamma_1 \sum_{i=1}^{c} \sum_{j=1}^{k} \left\| (\mathbf{w}_{t+1})_i^j \right\|_2 + \gamma_2 \sum_{i=1}^{d} \left\| \mathbf{w}_{t+1}^i \right\|_2 \leq L(X, \mathbf{W}_t) + \gamma_1 \sum_{i=1}^{c} \sum_{j=1}^{k} \left\| (\mathbf{w}_t)_i^j \right\|_2 + \gamma_2 \sum_{i=1}^{d} \left\| \mathbf{w}_t^i \right\|_2, \tag{41}$$

which means $Q(\mathbf{W}_{t+1}) \leq Q(\mathbf{W}_t)$. The algorithm is proved to converge to the global optimum solution.

REFERENCES

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

[2] ——, "The wave kernel signature: A quantum mechanical approach to shape analysis," 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1626–1633, 2011.

[3] S. Biasotti, A. Cerri, A. Bronstein, and M. Bronstein, "Recent trends, applications, and perspectives in 3d shape similarity assessment," Computer Graphics Forum, vol. 35, no. 6, pp. 87–119, 2016.

[4] A. M. Bronstein, M. M. Bronstein, L. J. Guibas, and M. Ovsjanikov, "Shape google: Geometric words and expressions for invariant shape retrieval," ACM Trans. Graph., vol. 30, no. 1, pp. 1:1–1:20, Feb. 2011.

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

[6] M. M. Bronstein and I. Kokkinos, "Scale-invariant heat kernel signatures for non-rigid shape recognition," in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 2010, pp. 1704–1711.

[7] S. Bu, Z. Liu, J. Han, J. Wu, and R. Ji, "Learning high-level feature by deep belief networks for 3-d model retrieval and recognition," IEEE Transactions on Multimedia, vol. 16, no. 8, pp. 2154–2167, Dec 2014.

[8] ——, "Learning high-level feature by deep belief networks for 3-d model retrieval and recognition," IEEE Transactions on Multimedia, vol. 16, no. 8, pp. 2154–2167, 2014.

[9] I. Chiotellis, R. Triebel, T. Windheuser, and D. Cremers, "Non-rigid 3d shape retrieval via large margin nearest neighbor embedding," in European Conference on Computer Vision(ECCV), October 2016, pp. 327–342.

[10] A. Dubrovina and R. Kimmel, "Matching shapes by eigendecomposition of the laplace-beltrami operator," in Proc. 3DPVT, vol. 2, no. 3, 2010, p. 2.

[11] A. Gasparetto and A. Torsello, "A statistical model of riemannian metric variation for deformable shape analysis," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015, pp. 1219–1228.

[12] P. Gehler and S. Nowozin, "On feature combination for multiclass object classification," in 2009 IEEE 12th International Conference on Computer Vision. IEEE, pp. 221–228.

[13] H. Ghodrati and A. B. Hamza, "Nonrigid 3d shape retrieval using deep auto-encoders," Applied Intelligence, vol. 47, no. 1, pp. 44–61, Jul 2017.

[14] M. Gnen and E. Alpaydin, "Multiple kernel learning algorithms," Journal of Machine Learning Research, vol. 12, pp. 2211–2268, 7 2011.

[15] M. W. Hirsch, Differential Topology. Springer New York, 1976.

[16] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of ir techniques," ACM Transactions on Information Systems (TOIS), vol. 20, no. 4, pp. 422–446, 2002.

[17] H. Laga, "A survey on non-rigid 3d shape analysis," CoRR, vol. abs/1812.10111, 2018. [Online]. Available: http://arxiv.org/abs/1812.10111

[18] R. Litman, A. Bronstein, M. Bronstein, and U. Castellani, "Supervised learning of bag-of-features shape descriptors using sparse coding," Comput. Graph. Forum, vol. 33, no. 5, pp. 127–136, Aug. 2014.

[19] J. Masci, D. Boscaini, M. M. Bronstein, and P. Vandergheynst, "Shapenet: Convolutional neural networks on non-euclidean manifolds," CoRR, vol. 1501.06297, 2015.

[20] M. Masoumi and A. B. Hamza, "Shape classification using spectral graph wavelets," Applied Intelligence, vol. 47, no. 4, pp. 1256–1269, Dec 2017.

[21] G. Obozinski, B. Taskar, and M. I. Jordan, "Joint covariate selection and joint subspace selection for multiple classification problems," Statistics and Computing, vol. 20, no. 2, pp. 231–252, 2010.

[22] D. Pickup, X. Sun, P. L. Rosin, R. R. Martin, Z. Cheng, Z. Lian, M. Aono, A. B. Hamza, A. Bronstein, M. Bronstein, S. Bu, U. Castellani, S. Cheng, V. Garro, A. Giachetti, A. Godil, J. Han, H. Johan, L. Lai, B. Li, C. Li, H. Li, R. Litman, X. Liu, Z. Liu, Y. Lu, A. Tatsuma, and J. Ye, "Shape retrieval of non-rigid 3d human models," in Proceedings of the 7th Eurographics Workshop on 3D Object Retrieval, ser. 3DOR '15. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2014, pp. 101–110.

[23] M. Reuter, S. Biasotti, D. Giorgi, G. Patan, and M. Spagnuolo, "Discrete laplacebeltrami operators for shape analysis and segmentation," Computers & Graphics, vol. 33, no. 3, pp. 381 – 390, 2009, iEEE International Conference on Shape Modelling and Applications 2009.

[24] 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.

[25] E. Rodola, S. Rota Bulo, T. Windheuser, M. Vestner, and D. Cremers, "Dense non-rigid shape correspondence using random forests," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2014.

[26] S. Rosenberg, The Laplacian on a Riemannian Manifold. Cambridge University press, 1997.

[27] ——, The Laplacian on a Riemannian Manifold, ser. London Mathematical Society Student Texts. Cambridge University Press, 1997, p. 151.

[28] R. M. Rustamov, "Laplace-beltrami eigenfunctions for deformation invariant shape representation," in Proceedings of the Fifth Eurographics Symposium on Geometry Processing, ser. SGP '07. Switzerland: Eurographics Association, 2007, pp. 225–233.

[29] L. Seenivasan, F. Bai, M. Ji, X. Gu, Z. T. H. Tse, and H. Ren, "Shape tracking of flexible morphing matters from depth images," IEEE Sensors Journal, 2020.

[30] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The princeton shape benchmark," in Proceedings Shape Modeling Applications, 2004., June 2004, pp. 167–178.

[31] S. Sonnenburg, G. Rätsch, C. Schäfer, and B. Schölkopf, "Large scale multiple kernel learning," J. Mach. Learn. Res., vol. 7, pp. 1531–1565, Dec. 2006.

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

[33] H. Wang, F. Nie, H. Huang, and C. Ding, "Heterogeneous visual features fusion via sparse multimodal machine," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2013, pp. 3097–3102.

[34] M. Wardetzky, S. Mathur, F. Kaelberer, and E. Grinspun, "Discrete Laplace operators: No free lunch," in Geometry Processing. The Eurographics Association, 2007.

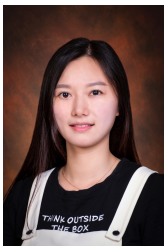
[35] J. Xie, G. Dai, and Y. Fang, "Deep multimetric learning for shape-based 3d model retrieval," IEEE Transactions on Multimedia, vol. 19, no. 11, pp. 2463–2474, 2017.

[36] J. Ye and Y. Yu, "A fast modal space transform for robust nonrigid shape retrieval," The Visual Computer, vol. 32, no. 5, pp. 553–568, May 2016.

[37] H. Zeng, Y. Liu, J. Liu, and D. Fu, "Non-rigid 3d model retrieval based on quadruplet convolutional neural networks," IEEE Access, vol. 6, pp. 76 087–76 097, 2018.

[38] W. Zhou, X. Jiang, and Y.-H. Liu, "Mvpointnet: Multi-view network for 3d object based on point cloud," IEEE Sensors Journal, vol. 19, no. 24, pp. 12 145–12 152, 2019.

Luqing Luo received the M.S. degree in Mechanical and Aerospace Engineering from North Carolina State University, Raleigh, NC, USA in 2010. She is currently pursuing the Ph.D. degree in State Key Laboratory of Internet of Things for Smart City, Department of Electromechanical Engineering, Faculty of Science and Technology, University of Macau, Macau, China. From 2008 to 2010, she was a research assistant in Department of Mechanical and Aerospace Engineering from North Carolina State University, Raleigh, NC, USA. Her current research interests include deep learning-based computer vision, as well as computer graphics, especially focus on 3D data analysis and semantic understanding.



Lulu Tang received the M.S. from Sichuan University, Chengdu, Sichuan, China, in 2012. She is currently pursuing the Ph.D. degree in State Key Laboratory of Internet of Things for Smart City, Department of Electromechanical Engineering, Faculty of Science and Technology, University of Macau, Macau, China. From 2012 to 2015, she was a Research Engineer in The Chinese Academy of Science (CAS), Chengdu, Sichuan, China. Her research interests mainly include machine learning, deep learning, computer vision, as well as computer graphics, especially focus on machine learning for 3D data.

Rui Liu has been an Assistant Professor in the College of Aeronautics and Engineering (CAE), Kent State University, Ohio, since 2019. He directed the Cognitive Robotics and AI lab (CRAI, ruiliurobotics.weebly.com), with a focus on cognitive robotics research - designing the ”Mind” for robotics and AI systems for seamless cooperation with a human. From 2018 Fall to 2019 Spring, Rui received his postdoc training at Carnegie Mellon University, Pittsburgh PA, the Robotics Institute at the School of Computer Science. In 2018 Spring, he received his Ph.D. degree from the Colorado School of Mines, Golden CO. He has finished more than 30 top-tier publications on human-Robot/Multi-Robot teaming. In the leading robotics conference IEEE ICRA 2015 and IEEE RO-MAN 2019, his works have been nominated as the best paper award.

Xiaoli Zhang received the B.S. degree in Mechanical Engineering from Xian Jiaotong University, Xian, ShanXi, China in 2003, the M.S. degree in Mechatronics Engineering from Xian Jiaotong University in 2006, the Ph.D. degree in Biomedical Engineering from the University of Nebraska Lincoln, Lincoln, USA, in 2009. Since 2013, she has been an Associate Professor in the Mechanical Engineering Department, Colorado School of Mines, Golden, CO. She is the author of more than 50 articles, and 5 inventions. Her research interests include intelligent human-robot interaction, human intention awareness, robotics system design and control, haptics, and their applications in healthcare fields.

Zhi-Xin Yang received the Ph.D. degree in industrial engineering and engineering management from the Hong Kong University of Science and Technology, Hong Kong, China, in 2000. He is currently an Associate Professor with the State Key Laboratory of Internet of Things for Smart City, the Department of Electromechanical Engineering, Faculty of Science and Technology, and the Director of Research Service and Knowledge Transfer Office, University of Macau, Macau, China. His current research interests include prognostic health management of engineering system, computer vision, robotics and intelligent safety monitoring.