3D Object Classification Using Scale Invariant Heat Kernels with Collaborative Classification

Mostafa Abdelrahman[1], Moumen El-Melegy[1,2], and Aly Farag[1]

[1] Computer Vision and Image Processing Laboratory,
University of Louisville, Louisville, KY 40292, USA
[2] Electrical Engineering Department, Assiut University, Assiut 71516, Egypt
{mostafa.abdelrahman,moumen.elmelegy,aly.farag}@louisville.edu
http://www.cvip.uofl.edu/

Abstract. One of the major goals of computer vision is the development of flexible and efficient methods for shape representation. This paper proposes an approach for shape matching and retrieval based on scale-invariant heat kernel (HK). The approach uses a novel descriptor based on the histograms of the scale-invariant HK for a number of critical points on the shape at different time scales. We propose an improved method to introduce scale-invariance of HK to avoid noise-sensitive operations in the original method. A collaborative classification (CC) scheme is then employed for object classification. For comparison we compare our approach to well-known approaches on a standard benchmark dataset: the SHREC 2011. The results have indeed confirmed the high performance of the proposed approach on the shape retrieval problem.

Keywords: Heat kernels, shape retrieval, collaborative classification, 3D shape descriptors.

1 Introduction

Recently, using 3D objects data has become more important in the area of computer vision, as recognition based on 3D models is less sensitive, or may be invariant, to lighting conditions and pose variations as compared to 2D models. The emergence of laser/lidar sensors, reliable multi-view stereo techniques and more recently consumer depth cameras have made the acquisition of 3D models easier than before. The domain of the presented work is the classification of these 3D objects into a set of pre-defined classes. One of the main challenges in that regard is the development of flexible and efficient methods for shape representation or the creation of a shape descriptor or signature for shape matching. The descriptor captures the properties of the shape that distinguish it from shapes belonging to other classes.

Shape descriptor should have as many of the following properties as possible: 1) Isometry invariant: isometric shapes should have the same descriptor independently of the objects given representation and location. 2) Scale invariant: For some applications, it is necessary that the descriptor is independent of the

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part I, LNCS 7583, pp. 22–31, 2012.

objects size, therefore the descriptor should optionally be scale invariant. 3) Similarity: Similarly shaped objects should have similar descriptors. 4) Efficiency: The time and space needed to compute those descriptor should be reasonable. 5) Completeness or shape-awareness: descriptor should give a complete characterization of the shape, thus representing the shape uniquely.

1.1 Review of Related Work

Although global shape descriptors (e.g., [2]) have shown good performance on many data sets, they have an underlying assumption that shapes are rigidly transformed. Other approaches have used local feature detection and local descriptor to describe 3D shapes, such as spin images [10], local patches [11], and conformal factor [12]. But these methods cannot deal with the non-rigid shape deformation, and cannot cover the properties of the desired shape descriptor.

The problem of non-rigid shape deformation needs more work to compensate for the degrees of freedom resulting from local deformations. Early work by Elad and Kimmel [3] proposed modeling shapes as metric spaces with the geodesic distances as an intrinsic metric, which are invariant to inelastic deformations. Bronstein et al [4] used this framework with a metric defined by internal distances in 2D shapes. Reuter et al. [5] used the Laplacian spectra as intrinsic shape descriptors, and they employed the Laplace-Beltrami spectra as 'shape-DNA' or a numerical fingerprint of any 2D or 3D manifold (surface or solid). They proved that 'shape-DNA' is an isometry-invariant shape descriptor.

Recently Sun et al. [8] proposed heat kernel signatures (HKS) as a deformation-invariant descriptors based on diffusion of multi-scale heat kernels. HKS is a point based signature satisfying all of the good descriptor properties except for scale invariance. It characterizes each vertex on the meshed surface using a vector. However, the authors did not demonstrate how to retrieve shapes using HKS, although they pointed out the future potentials in shape retrieval applications. Fang et al [9] defined the temperature distribution (TD) of the heat mean signature (HMS) as a shape descriptor for shape matching. Their TD is a global shape descriptor and they used $L2$ norm which is a very basic matching method to compute the distance between two TD descriptors. Bronstein et al [7] solved the HKS scale problem through a series of transformations. The same research group has recently introduced the Shape Google approach [17] based on the scaled-invariant HKS. The idea is to use HKS at all points of a shape, or alternatively at some shape feature points, to represent the shape by a Bag of Features (BoF) vector. Sparsity in the time domain is enforced by preselecting some values of the time.

1.2 Paper Contribution

In this paper, we present an approach for shape matching and retrieval based on scale-invariant heat kernel (HK). Several aspects are novel in our approach. We use the first non-trivial Laplace-Beltrami eigenfunction to detect a small number of sparse critical points on the shape surface. These points are robust

to the shape class, and their number can in itself be used as one of the discriminatory features among the various classes. Then we calculate the HK for the detected critical points at different time scales. Then scale invariance is achieved using an improved method to Bronstein et al's approach [17]. A concatenation of the histograms of the significant components of the scale-invariant HK for all the points is used as a feature vector for classification. The resulting descriptor captures the local as well as global shape information since it uses the temperature distribution at the critical points at several time samples. For the sake of comparison we compare our approach to the Shape Google approach [17], the shape-DNA [6] and the TD approach of [9] on the SHREC 2011-Shape Retrieval Contest of Non-rigid 3D Watertight Meshes [13]. In particular, we demonstrate that our approach can perform partial matching, when there are missing data, and more robust performance against noise. In [18] similar approach to this work but the scale invariance achieved by different way, also the classification is done by different technique.

2 Heat Kernel Basics

In this section we start with the basics of diffusion on Riemannian manifolds that are necessary to define the proposed heat kernel signature. We will model the shape as a Riemannian manifold, possibly with boundary. The heat kernel quantitatively encodes the heat flow across a manifold $\mathbf{M}$ and is uniquely defined for any two vertices i, j on the manifold. The heat diffusion propagation over $\mathbf{M}$ is governed by the heat equation

$$\triangle_{\mathbf{M}} u(x,t) = -\frac{\partial}{\partial t} u(x,t), \tag{1}$$

where $\triangle_{\mathbf{M}}$ denotes the positive semi-definite Laplace- Beltrami operator of $\mathbf{M}$, which is Riemannian equivalent of the Laplacian. The solution $u(x,t)$ of the heat equation with initial condition $u(x,0) = u_0(x)$ describes the amount of heat on the surface at point x in time t. $u(x,t)$ is required to satisfy the Dirichlet boundary condition $u(x,t) = 0$ for all $x \in \partial M$ and all t. Given an initial heat distribution $f : M \to \mathbb{R}$, let $H_t(f)$ denote the heat distribution at time t, namely $H_t(f)$ satisfies the heat equation for all t, and $\lim_{t\to 0} H_t(f) = f$. H_t is called the heat operator. Both $\triangle_{\mathbf{M}}$ and H_t are operators that map one real valued function defined on $\mathbf{M}$ to another such function. It is easy to verify that they satisfy the following relation $H_t = e^{-t\triangle_{\mathbf{M}}}$. Thus both operators share the same eigenfunctions and if λ is an eigenvalue of $\triangle_{\mathbf{M}}$, then $e^{-\lambda t}$ is an eigenvalue of H_t corresponding to the same eigenfunction. The solution of (1) is called heat kernel and can be thought of as the amount of heat that is transferred from x to y in time t given a unit heat source at x. In other words, $H_t(x,.) = H_t(\delta_x)$, where δ_x is the Dirac delta function at $x : \delta_x(z) = 0$ for any $z \neq x$, and $\int_{\mathbf{M}} \delta_x(z)dz = 1$.

If $\mathbf{M}$ is compact then the heat kernel has the following eigen decomposition

$$H_t(x,y) = \sum_{k=1}^{\infty} e^{-\lambda_i t} \phi_i(x)\phi_i(y), \tag{2}$$

where λ_i and ϕ_i are the i^{th} eigenvalue and the i^{th} eigenfunction of the Laplace-Beltrami operator respectively, and x and y denote two vertices. The $H_t(x, y)$ is defined as the heat affinity $H_{af}(x, y)$ between a pair of vertices which is a measure of heat transfered between node x and y after time t.

Properties of the Heat Kernel: The heat kernel $H_t(x, y)$ has many good properties [8]. It is symmetric: or $H_t(x, y) = H_t(y, x)$. It is invariant under isometric deformations: which is a direct consequence of the invariance of the Laplace-Beltrami operator. It is informative: by only considering its restriction to the temporal domain we can obtain a concise and informative signature. It is multi-scale: for different values of t the heat kernel reflects local properties of the shape around x at small t and the global structure of M from the point of view of x at large values of t. And it is stable under perturbations of the underlying manifold.

3 Proposed Approach

In this paper, we propose to construct the shape descriptor as follows: HKs are calculated at some critical points detected on the surface (see below) at various time samples (about 150). Then scale-invariance is introduced in the computed HK as explained in the following subsection. Since the complexity of using the heat kernel as a signature is extremely high, and it would be difficult to compare descriptors of two different points, we use histograms to overcome the descriptor alignment problem and to reduce the descriptor size. At each time sample, (as described in the Scale Invariance sec. 3.1) after taking the logarithmic transformation, and the amplitude of the Fourier transform, a histogram of 100 pins is calculated for the low-frequency components. Then all the histograms from all detected critical points are concatenated to build a long feature vector. Then the normalized eigenvalues of the Laplace-Beltrami operator are appended to this vector. This vector, dubbed Critical Points-based Heat Kernel (CP-HK), can be used for classification using some well-known classifiers. However, for the latter part, we use collaborative classification [21]. In order to construct the HK at a given vertex x based on formula (2), we use a finite number eigenfunctions and eigenvalues of the Laplace-Beltrami operator which is replaced by its cotangent formula for triangular meshes [14]:

$$(\triangle_{\mathbf{M}} u)_i = -\frac{1}{A_i} \sum_{j \in N_{ei(i)}} (\cot \alpha_{ij} + \cot \beta_{ij})(u_i - u_j), \tag{3}$$

where $(\triangle_{\mathbf{M}} u)_i$ for a mesh function u denotes its discrete Laplacian evaluated at vertex i (for $i = 1; 2; ....; N$, N number of vertices); A_i is the Voronoi area at the i^{th} mesh vertex [14]; ans α_{ij}, β_{ij} are the two angles supporting the edge connecting vertices i and j. This discretization preserves many important properties of the continuous Laplace-Beltrami operator, such as positive semi-definiteness, symmetry, and locality, and it is numerically consistent [15]. In a matrix form we can write

$$(\triangle_{\mathbf{M}} u)_i = A^{-1} L u, \tag{4}$$

where $A = diag(A_i)$, $L = diag(\sum_{l \neq i} w_{il}) - w_{ij}$, and $w_{ij} = (\cot \alpha_{ij} + \cot \beta_{ij})$. The first k smallest eigenvalues and eigenfunctions of the Laplace-Beltrami operator discretized according to (4) are computed by solving the generalized eigendecomposition problem $W\phi = \lambda A\phi$, where $L = \phi \Lambda \phi^T$, Λ is a diagonal matrix of eigenvalues,and ϕ is a $NX(k+1)$ matrix whose columns correspond to the right eigenvectors of L.

3.1 Scale Invariance

Scale invariance can be achieved by four different methods: (1) trying to detect scale, as done in most feature descriptors in image analysis (e.g. SIFT). However, 3D shapes are usually poorer in features and scale detection can be done reliably only at a sparse set of feature points. (2) through the normalization of Laplace-Beltrami eigenvalues, but this method may suffer if the object has missing parts [7]. In such case, the scale invariance must be introduced locally rather than globally. (3) Using a series of transformations applied to the HKS [7] in order to avoid scale detection. This allows creating a dense descriptor. This method is considered local, thus can work with objects with missing parts. (4) the local equi-affine invariant Laplace-Beltrami operator proposed by Raviv et al [1].

In this paper, we propose an improved variant of the third method to achieve scale invariance. It was shown [7] that scaling a shape by a factor β results in changing $H_t(x, y)$ to $\beta^2 \; H_{\beta^2 t}(x, y)$. Thus a series of transformations are applied to HK as follows. Starting from each critical point x, the HK is sampled at every surface point y logarithmically in time ($t = \alpha^\tau$) and the function $h_\tau = H_{\alpha^\tau}(x, y)$ is formed. Scaling the shape by β results in a time shift $s = 2 \log_\alpha \beta$ and amplitude scaling by β^2. That is, $h'_\tau = \beta^2 h_{\tau+s}$. The logarithmic transformation $\log \; h'_\tau$ decouples the multiplicative constant from $h_{\tau+s}$. Bronstein et al [7] proposed to take the derivative afterwards to remove the effect of the resulting additive β^2 term and then taking the amplitude of the Fourier transform (FT) of the derivative to remove the effect of the time shift s. Since the derivative operator is sensitive to noise, in a departure from [7], we propose to apply the Fourier transform directly to $\log \; h'_\tau$. The effect of the multiplicative constant β^2 is eliminated by dropping the DC (zero frequency) component of the FT, and then the amplitude of the remaining significant FT components (we normally use 6) are attained. This eliminates the scale effect without having to use the noise-sensitive derivative operation.

3.2 Critical Points

For a piecewise linear real-valued function ϕ given by the values at the vertices (ϕ_i) of a triangle mesh, we define a critical point as a vertex i whose function ϕ_i is a maximum or minimum over its neighborhood (in two rings). These points are detected using the local maxima/minima of the first non-trivial Laplace-Beltrami eigenfunction [19]. Critical point detected near the boundary are discarded. Figure 1 shows the critical points detected from the first non-trivial eigenfunction for sample shapes. The figure gives the total number of critical points for each shape.

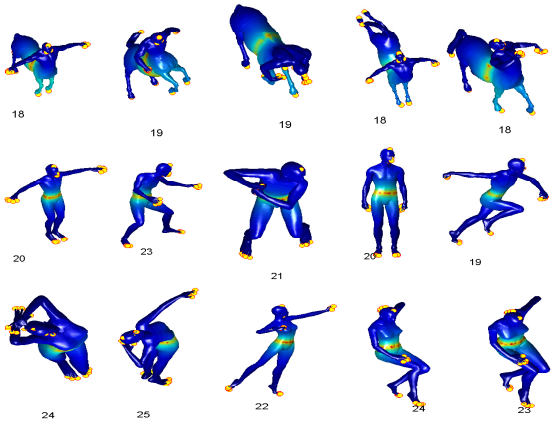

Fig. 1. Critical points detected for several shapes. Number below each shape represents the total number of shape critical points. Colors visualize the average temperature induced from these critical points throughout all shape vertices.

It is interesting to observe that shapes belonging to the same class consistently have almost the same number of critical points, whereas these numbers differ from one class to another. As such, this number can be used as one of the discriminatory features between the different classes, in addition to the HK descriptor.

3.3 Collaborative Classification

Recently, collaborative representation has also been used in pattern classification. Zhang [20] proposed a new classification scheme, namely collaborative representation (CR) based classification with regularized least square (CRC-RLS), which has significantly less complexity than the sparse representation based classification (SRC) but leads to very competitive classification results. Then [21] propose a relaxed collaborative representation (RCR) model, which considers both the similarity and distinctiveness of different features in coding and classification stages. Zhang [21] showed that RCR is simple, and very competitive with state-of-the-art image classification methods. We use the RCR approach for the coding and classification of our proposed descriptors. More details about the algorithm in [21].

4 Experimental Results

To test the performance of the proposed approach we use the SHREC 2011 - Shape Retrieval data set [13]. This is a large-scale database which consists of 600 non-rigid 3D objects that are derived from 30 original models. For the sake of comparison, we show the results of the Shape-DNA approach [6], describing shapes by the vector of the first eigenvalues of the Laplace-Beltrami operator. We

used first 15 eigenvalues to construct the Shape-DNA descriptors. Eigenvalues were computed using the same cotangent weight discretization. We also compare our results to the method in [9] that uses the (TD) as a shape descriptor and the Shape Google approach [17]. Figure 2 shows sample shape retrieval results of the CP-HK descriptor on the SHREC 2011 dataset. The figure shows the first 15 matches for each query ranked according to the distance measure of the RCR classifier. Afterwards, the objects ranked from 30-35 for each query are shown on the right. Several of these objects are also similar in shape to the query object.

Fig. 2. Shape retrieval results of SHERC 2011 dataset. Left: queries. Middle: First 15 matches using the HK descriptor. Right: matches from 30 to 35. The color represents the first component HK at each point. The detected critical points shown in yellow.

Table 1. Results on SHERC11 dataset. Note the results for the Shape Google method is from our implementation as described [17]

Feature	SI	Classifier	NN	1-Tier	2-Tier	e-Measure	DCG
TD	–	NN	0.6483	0.3704	0.4768	0.3369	0.6684
Shape-DNA	–	NN	0.9900	0.8588	0.9295	0.6797	0.9649
SI-HKS	[17]	BOF	0.9567	0.6225	0.7288	0.5245	0.8718
CP-HK	Our method	RCR	**0.9733**	**0.7798**	**0.8823**	**0.6443**	**0.9364**

For the sake of quantitative assessment of the approach performance with all the tested classifiers, we record the following standard five evaluation measures (see [16] for detailed definitions): Nearest Neighbor (NN) where N = 1, First Tier (FT), Second Tier (ST), E-measure (E), and Discounted Cumulative Gain (DCG). Table 1 shows the performance on the SHERC 2011 dataset. The table compares the proposed descriptor against TD, Shape-DNA and Shape Google. The proposed approach significantly outperforms the TD approach and performs higher than the Shape Google. Although the Shape-DNA shows the

Table 2. Performance versus noise, shot-noise, and scale in three severity levels of the CP-HK descriptor using the RCR classifier compaerd to the results of Shape Google [17]. (1.00 mean 100%)

Noise L.	NN our	NN [17]	Shot N. L.	NN our	NN [17]	Sclae L.	NN our	NN [17]
1	1.0000	1.0000	1	0.9333	0.9333	1	1.0000	0.8000
2	1.0000	0.9000	2	0.8666	0.8666	2	1.0000	0.4666
3	0.9333	0.1333	3	0.8000	0.5333	3	1.0000	0.2333

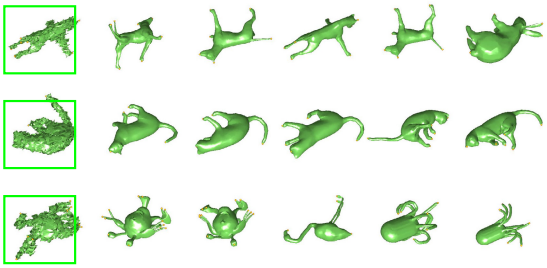

Fig. 3. Some shape retrieval results for different shapes with different noise levels

best performance in this experiment, it severely suffers when there are missing parts in the objects (i.e., on partial shape matching, different scale and noise). This was clearly demonstrated in [17] compared with the Shape Google approached.

Another experiment is carried out to assess the approach performance under several distorted data scenarios. Here we compare the performance of the proposed approach with that of the Shape Google approaches. We have formed a query set consisting of 30 shapes taken from the SHERC11 data set, after applying several distortions: a Gaussian white noise, shot-noise, scaling, and missing parts. The performance versus white noise in three different levels of severity is shown in Table 2. Figure 3,and 4 illustrates sample shapes corrupted with these three different noise, and shot noise levels. Clearly the proposed approach shows a more robust noise performance.

The average performance of the two approaches in the case of shot noise is summarized, and lists the overall performance versus query objects with several scales are also shown in Table 2. The proposed approach has retrieved the shapes with different scales with 100% accuracy. The Shape Google presented considerably lower performance. The performance against missing parts is demonstrated in Figure 5. The results show that the proposed approach is better to handle partial data. Note that we used data set consists of 30 class and we don't ignore similar-class positive shapes (males and females, centaur, horse, and human shapes) as in Shape Google experiment [17]. This justifies why the results of Shape Google in Table 2 are lower compared to the results reported in [17].

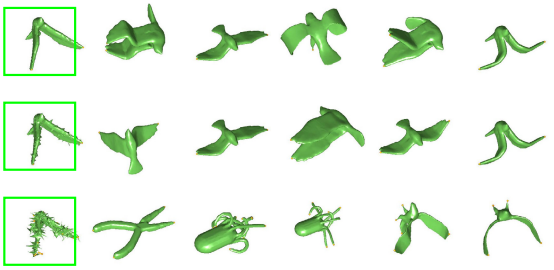

Fig. 4. Some shape retrieval results for different shapes with different shot noise levels

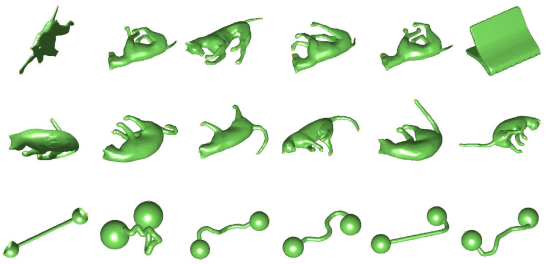

Fig. 5. Some shape retrieval results for different shapes with different missing parts. Left: queries. Right: the first 5 matched shapes.

5 Conclusion and Future Work

In this paper, we have presented an approach for shape matching and retrieval based on scale-invariant heat kernel (HK). An improved method to introduce scale-invariance has been also proposed to avoid noise-sensitive operations in the original transformation method. We have also proposed to use the first non-trivial Laplace-Beltrami eigenfunction to detect a small number of sparse critical points on the surface of the shape. These points were shown to be robust to the shape class, and their number can in itself be used as one of the discriminatory features among the various classes. We have utilized a collaborative classification scheme for object matching and retrieval. Our experimental results have shown that the proposed descriptor can achieve high performance on a public, well-known benchmark dataset. An important observation from our experiments is that the proposed approach is more able to handle data under several distortion scenarios (noise, shot-noise, scale, under missing parts) than the well-known Shape Google approach. Therefore, the proposed approach is more suitable for partial shape matching and retrieval from databases. Our current research is directed towards using the proposed approach to address dense correspondence between non-rigid shapes.

References

1. Raviv, D., et al.: Affine-invariant diffusion geometry for the analysis of deformable 3D shapes. In: Proc. Computer Vision and Pattern Recognition, CVPR (2011)
2. Kazhdan, M., et al.: Rotation invariant spherical harmonic representation of 3D shape descriptors. In: Proc. SGP, pp. 156–164 (2003)
3. Elad, A., Kimmel, R.: Bending invariant representations for surfaces. In: Proc. CVPR, pp. 168–174 (2001)
4. Bronstein, A.M., Bronstein, M.M., Bruckstein, A.M., Kimmel, R.: Analysis of two-dimensional non-rigid shapes. IJCV (2008)
5. Reuter, M., et al.: Discrete Laplace-Beltrami operators for shape analysis and segmentation. Computers and Graphics 33(3), 381–390 (2009)
6. Reuter, M., Wolter, F.E., Peinecke, N.: Laplace-Beltrami spectra as Shape-DNA of surfaces and solids. Computer-Aided Design 38(4), 342–366 (2006)
7. Bronstein, M., Kokkinos, I.: Scale-invariant heat kernel signatures for non-rigid shape recognition. In: IEEE Computer vision and pattern recognition (CVPR), pp. 1704–1711 (2010)
8. Sun, J., Ovsjanikov, M., Guibas, L.: A concise and provably informative multi-scale signature based on heat diffusion. In: SGP 2009: Proceedings of the Symposium on Geometry Processing, pp. 1383–1392 (2009)
9. Fang, Y., Sun, M., Ramani, K.: Temperature Distribution Descriptor for Robust 3D Shape Retrieval. In, NORDIA 2011 (2011) (CVPRW)
10. Johnson, A., Hebert, M.: Using spin images for efficient object recognition in cluttered 3 d scenes. Trans. PAMI 21(5), 433–449 (1999)
11. Toldo, R., Castellani, U., Fusiello, A.: Visual vocabulary signature for 3D object retrieval and partial matching. In: Proc. Eurographics Workshop on 3D Object Retrieval (2009)
12. Ben-Chen, M., Weber, O., Gotsman, C.: Characterizing shape using conformal factors. In: Proc. Eurographics Workshop on Shape Retrieval (2008)
13. Lian, Z., et al.: SHREC 2011 Track: Shape Retrieval on Non-rigid 3D Watertight Meshes. In: Proceedings of the Eurographics/ACM SIGGRAPH Symposium on 3D Object Retrieval (2011)
14. Grinspun, E., Gingold, Y., Reisman, J., Zorin, D.: Computing discrete shape operators on general meshes. [Eurographics (2006) Best Paper, 3rd Place]. Eurographics (Computer Graphics Forum) 25(3), 547–556 (2006)
15. Wardetzky, M., Mathur, S., Kalberer, F., Grinspun, E.: Discrete Laplace operators: no free lunch. In: Conf. Comp. Grap. and Interactive Techniques (2008)
16. Shilane, P., Min, P., Kazhdan, M., Funkhouser, T.: The princeton shape benchmark. In: Proc. SMI 2004, pp. 167–178 (2004)
17. Bronstein, A.M., Bronstein, M.M., Ovsjanikov, M., Guibas, L.J.: Shape Google: geometric words and expressions for invariant shape retrieval. ACM Trans. Graphics (TOG) 30(1), 1–20 (2011)
18. Abdelrahman, M., El-Melegy, M.T., Farag, A.A.: Heat Kernels for Non-Rigid Shape Retrieval: Sparse Representation and Efficient Classification. In: CRV, pp. 153–160 (2012)
19. Reuter, M.: Hierarchical shape segmentation and registration via topological features of laplace-beltrami eigenfunctions. Proc. IJCV 89(2), 287–308 (2010)
20. Zhang, L., Yang, M., Feng, X.: Sparse Representation or Collaborative Representation: Which Helps Face Recognition? In: ICCV (2011)
21. Yang, M., Zhang, L., Zhang, D., Wang, S.: Relaxed Collaborative Representation for Pattern Classification. In: CVPR (2012)