Eurographics Workshop on 3D Object Retrieval (2015)
I. Pratikakis, M. Spagnuolo, T. Theoharis, L. Van Gool, and R. Veltkamp (Editors)

Bag of Compact HKS-based Feature Descriptors

H. ElNaghy and S. Hamad

Scientific Computing Department, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt

Abstract

3D object retrieval has become an integral part in many today's applications attracting extensive research efforts. This paper introduces an enhanced 3D object retrieval technique using a compact and highly discriminative feature point descriptor. The key idea is based on integrating Bag of features (BoF) paradigm with Heat Kernel Signature (HKS) for feature description and detection. Initially, HKS computation phase defines HKS point signatures for each 3D model. Then, an innovative feature point detection algorithm provides a succinct set of feature points to be associated with a compact HKS-based descriptor vectors computed at local time scales. Finally, we take advantage of the BoF paradigm to encode a given 3D model with an informative feature frequency vector. The proposed approach has been evaluated on SHREC 2015 dataset of non-rigid models. The experimental results demonstrate the effective retrieval performance, invariance to different kinds of deformation and possible noise.

Categories and Subject Descriptors (according to ACM CCS): H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval; I.2.10 [Artificial Intelligence]: Vision and Scene Understanding—Shape; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism; I.3.8 [Computer Graphics]: Applications.

1. Introduction

3D Object Retrieval is one of the most significant research areas in computer graphics and multimedia. Plenty of 3D model databases are distributed freely everyday and thus designing robust methods to efficiently and effectively search interested 3D models is inevitable. A large number of competing techniques have been recently developed for matching 3D objects [EHK13].

Definitely, 3D object retrieval radically relies on the advancement of feature description and detection techniques. A promising amount of research has applied the spirit of heat diffusion for feature detection and description. Originally, Sun et al. [SOG09] propose Heat Kernel Signatures (HKS) as a deformation invariant descriptor and then Bronstein et al. [BK10] present its scale invariant version. Later, Fang et al. [FSKR11] define the Temperature Distribution (TD) of the Heat Mean Signature (HMS) as a shape descriptor. Recently, Litman et al. [LB14] have introduced a different perspective by applying machine learning techniques. They applied a bank of filters to the shape's geometric features at different frequencies generalizing the heat and wave kernel signatures. On the other hand, different approaches have introduced Bag-of-Features (BoF) paradigm, successfully applied in image analysis. Toldo et al. [TCF09] propose a part-based representation that partitions the objects into subparts, each characterized with a rigid geometric descriptor. Then, 3D visual dictionary is defined by clustering such descriptors. Ohbuchi et al. [OOFB08] also use BoF approach as a part of their introduced method. Another improved algorithm is proposed by Furuya and Ohbuchi [FO09], which extracts much larger number of local visual features by sampling each depth image densely and randomly. Moreover, the idea of combining BoF paradigm with HKS [BBGO11] also brings an innovative insight to 3D object retrieval researchers.

Choosing a compact and informative feature descriptor is a vital aspect for successfully applying the BoF paradigm in 3D object retrieval. Therefore, in this paper we introduce a 3D object retrieval technique using a bag of compact HKS-based feature point descriptors defined at concise local time scales. We thoroughly study the influence of changing the computation way of such feature descriptor until reaching an informative, easily computed 4th dimensional feature vector. Moreover, we evaluate the proposed technique on the most recent dataset of non-rigid models SHREC 2015 [Lea]. The attained results confirm that the proposed technique achieves high retrieval performance invariant against noise and different kinds of both deformations and transformations.

© The Eurographics Association 2015.

DOI: 10.2312/3dor.20151060

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

2. Proposed Approach Overview

The proposed 3D object retrieval technique proceeds through five main phases: HKS Computation, Feature Point Detection, Feature Point Description, Bag of Features and finally the Matching phase. Each of these phases is comprehensively described through the following sub sections.

2.1. HKS Computation

Basically, HKS computation phase computes HKS point signatures for each 3D model in order to be used successively by feature detection and description phases. For discrete computation of the Heat Kernel Signature, first the Laplace-Beltrami operator is estimated on triangular meshes. Second, the smallest $\boldsymbol{k}$ eigenvalues and eigenfunctions of the Laplace-Beltrami operator are calculated ($k = 300$). Finally, each point x in a given 3D model is encoded by an m-dimensional HKS feature vector $K_t(x,x)$. It describes the model's local and global geometric properties at $\boldsymbol{m}$ different time scales ($m = 100$) over the time interval $[t_{min}, t_{max}]$ with $t_{min} = 4\ln 10/\lambda_k$, $t_{max} = 4\ln 10/\lambda_2$, such that:

$$K_t(x,x) = \sum_{i=0}^{k} e^{-\lambda_i t}\phi_i(x)^2, \qquad (1)$$

where λ_i and ϕ_i are the i^{th} eigenvalue and eigenfunction of the Laplace-Beltrami operator respectively [SOG09].

2.2. Feature Point Detection

First, the feature point detection phase constitutes an initial set of feature points by capturing the HKS critical points for all models over the m times scales. For a given 3D model, a vertex x is considered a critical point, i.e. local maxima, at time scale t if $K_t(x,x) > K_t(x_i,x_i)$ for all x_i in the 2-ring neighborhood of x [SOG09].

Then, an innovative filtering technique is applied to keep the most stable and significant initially detected critical points by discarding repeated or insignificant ones. This dramatically reduces both time and space required later to construct the geometric dictionary during the Bag of Features phase.

2.3. Feature Point Description

First, the feature vector $K_t(x,x)$ is normalized into an m-dimensional feature vector $h(x)$, such that:

$$\begin{aligned} h(x) &= (h_1(x), \ldots, h_m(x)); \\ h_i(x) &= c(x)K_{t_i}(x,x), \end{aligned} \qquad (2)$$

where the constant $c(x)$ is selected in such a way that $\|h(x)\| = 1$ [BBGO11]. Then each feature point is associated with a d-dimensional feature descriptor $\boldsymbol{p(x)}$ that is: compact in size, efficient to compute, informative, discriminative and robust. Thus, $\boldsymbol{p(x)}$ is a subset of the m-dimensional HKS feature vector $h(x)$ encoding each feature point at $\boldsymbol{d}$ significant time scales such that $d \in \mathbb{R}^+$ and $d < m$. This chosen subset acts as a window on the m-dimensional feature vector $h(x)$ keeping the most discriminative and informative d HKS values, whereas ignoring redundant and non-discriminative $m - d$ HKS values.

Section 3 will show in detail our conducted experiments for choosing the most efficiently computed d-dimensional feature descriptor $p(x)$ that could achieve the best retrieval performance.

2.4. Bag of Features

Bag-of-Features (BoF) paradigm [BBGO11] is divided into two main sub phases: vocabulary construction and 3D object representation. The vocabulary construction sub phase clusters the descriptor space using K-means clustering technique in order to construct the required geometric vocabulary $P = \{p_1, \ldots, p_v\}$, such that each visual word p_i is a d-dimensional HKS based feature vector. For a 3D model X with n vertices $\{x_1, \ldots, x_n\}$, 3D object representation sub phase computes a feature distribution $\theta(x)$ for each vertex x using hard vector quantization. Such feature distribution $\theta(x) = \{\theta_1(x), \ldots, \theta_v(x)\}$ is a $v \times 1$ vector whose elements are defined as follows:

$$\begin{aligned} \theta_i(x) &= 1 \text{ if } i = \operatorname{argmin}_{i\in[1\ldots v]} \|p(x) - p_i\|; \\ \theta_i(x) &= 0 \qquad \text{otherwise}, \end{aligned} \qquad (3)$$

where v is the number of geometric words in the vocabulary ($v = 64$).

After vector Quantization, the feature distribution $\theta(x)$ is integrated over the entire 3D model, yielding a $v \times 1$ vector of feature frequency $F(x)$, which is referred as a Bag-of-Features (BoF) or Bag-of-Words (BOW). In a real discrete environment, $F(x)$ is computed as follows:

$$F(X) = \sum_{i=1}^{n} \theta(x_i), \qquad (4)$$

where n is the number of vertices for a given 3D model X.

Finally, $F(X)$ is considered a compact and significant shape descriptor that entirely reflects a 3D model's geometrical features and is highly discriminative compared to other shape descriptors.

2.5. Matching

3D models can be successfully matched by comparing their corresponding shape descriptors $F(X)$, since similar 3D models tend to have similar bags of features, whereas different 3D models tend to have different bags of features. Thereupon, given two 3D models X and Y, the L_1 distance between the two bags of features $F(X)$ and $F(Y)$ is defined as follows:

$$d_{BoF} = \|F(X) - F(Y)\|. \qquad (5)$$

© The Eurographics Association 2015.

3. Influence of HKS-based Feature Descriptor *p(x)*

The way of computing the d-dimensional feature vector $p(x)$ highly influences the retrieval performance of our proposed technique. Therefore, we conduct several experiments to select the optimum way for computing such vital feature descriptor. The first experiment is conducted by selecting the minimum, maximum, mean, median and standard deviation of $h(x)$, producing a fifth dimensional descriptor vector $p(x)$. On the other hand, the second, third and fourth experiments are conducted by sampling the HKS values using different intervals equal to 5, 10 and 20 respectively. Table 1 summarizes how the descriptor vector $p(x)$ is selected in our four conducted experiments. The third column indicates the vector's dimensionality d, while the two rightmost columns show the timing (in seconds) and the storage (in MB) required for clustering and storing the feature space respectively. Intuitively, the time and storage required for clustering the feature space decreases by decreasing the dimensionality of the feature vector $p(x)$.

Run	*p(x)*	*d*	Time (sec)	Storage (MB)
Run 1	$[min(h(x)), max(h(x)), mean(h(x)), median(h(x)), stdev(h(x))]$	5	13	2.6
Run 2	$[h_5(x), h_{10}(x), ..., h_{90}(x), h_{95}(x)]$	19	36	9.9
Run 3	$[h_{10}(x), h_{20}(x), ..., h_{80}(x), h_{90}(x)]$	9	21	5.2
Run 4	$[h_{20}(x), h_{40}(x), h_{60}(x), h_{80}(x)]$	4	11	2.1

Table 1: *Different selections of the d-dimensional feature descriptor p(x).*

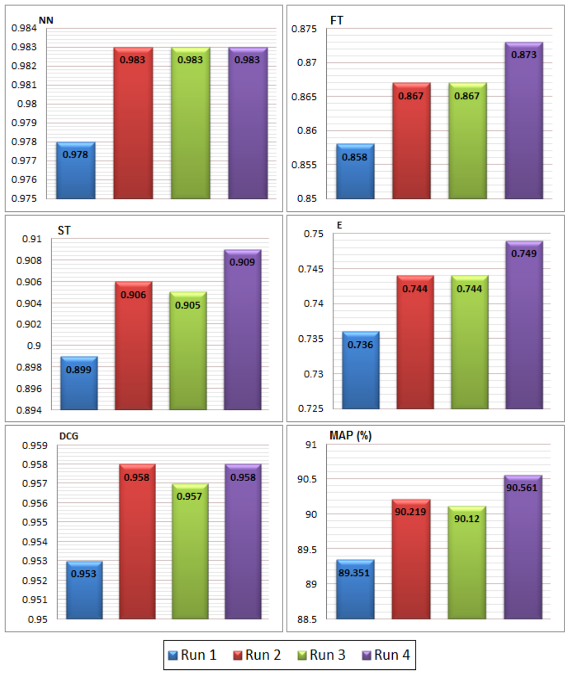

Figure 1: *Performance comparison of the four conducted experiments using different selections of the d-dimensional feature descriptor p(x)*

Figure 1 shows the effect of varying the selection of the d-dimensional feature descriptor $p(x)$ on the overall retrieval performance through our four conducted experiments. The six evaluation metrics used here are NN, FT, ST, E, DCG and MAP. All these experiments are evaluated on the whole database of SHREC 2015 [Lea].

It is obvious that fourth run outperforms all other runs, particularly upon considering the FT, ST, E and MAP evaluation metrics. Fortunately, this run is the one owning the shortest and most compact feature descriptor $p(x)$ with dimensionality equals to four. Considering the values of NN, ST, E and DCG evaluation metrics, the second and the fourth runs achieve approximate performance results. However, it is more sensible to choose the feature descriptor $p(x)$ tested by the fourth run since it is more compact than that tested by the second run. It should be pointed out that the length (dimensionality) of the feature descriptor $p(x)$ tested by the second run ($d = 19$) is almost five times the dimensionality of the feature descriptor tested by the fourth run ($d = 4$).

In conclusion, the feature vector $p(x)$ applied in Run 4 achieves the best retrieval results outperforming the other three runs. Furthermore, it requires an acceptable time (11 sec) and reasonable storage (2.1 MB) for clustering the whole feature space. Therefore, we advocate the use of such compact fourth dimensional HKS-based vector for point feature description within our proposed technique.

4. Retrieval Results

This section is devoted for visualizing samples of our retrieval results, besides showing the robustness of the proposed technique against various deformation, transformations and noise. Some examples of retrieved 3D models returned by the proposed technique are visualized in Figure 2. The leftmost column of 3D models (outlined in green) represents queries belonging to different categories.

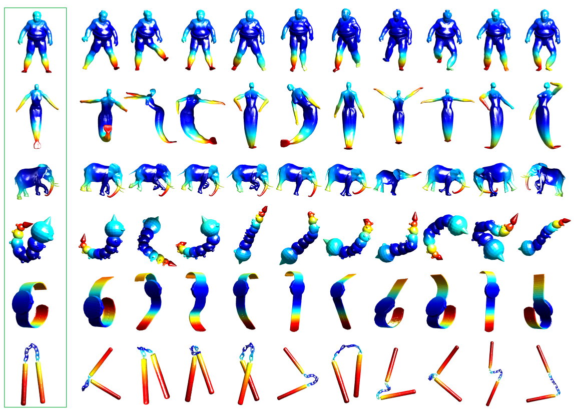

Figure 2: *Sample of 3D object retrieval results.*

© The Eurographics Association 2015.

The ten following columns represent the closest ten retrieved 3D models ordered according to their relevance. It is obvious that all the retrieved 3D models are relevant matches. It is essential to highlight that SHREC 2015 dataset contains 3D models arbitrarily scaled, rotated and translated. Therefore, such outstanding retrieval results indicate that the proposed technique is invariant to different kinds of transformations. Moreover, Figure 3 shows how the proposed descriptor is intrinsically invariant to different isometric deformations. The retrieval results of five query models with different deformations, belonging to the same category are visualized. It is obvious that retrieved 3D models of all queries are relevant matches despite of the existence of different classes of deformations applied on their corresponding query models.

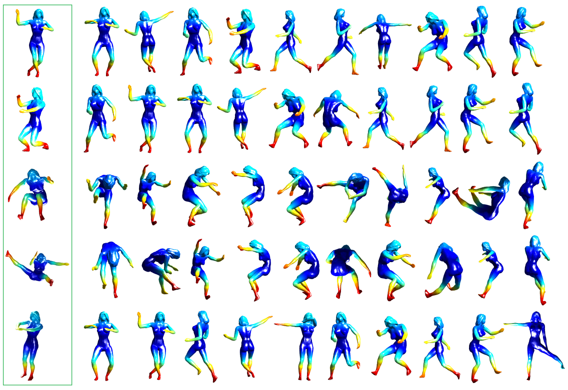

Figure 3: *Sample of 3D object retrieval results given queries with different deformations.*

In order to emphasize the robustness of the proposed technique with respect to noise, Figure 4 illustrates some of the retrieval results for different query models corrupted with a noise level equals to 0.01. Here, the noise level is analogous to the noise-to-signal ratio (NSR) between the variance of noise and the variance of the original signal (coordinates of the vertices). It is clear that our technique attains reasonable retrieval rate confirming its stability against noise.

5. Conclusion and Future Work

In this paper, an efficient and robust 3D object retrieval approach has been proposed. Applying a compact HKS-based feature point descriptor dramatically reduces time and space complexity required for the whole BoF model. The proposed technique achieves high retrieval accuracy on SHREC 2015 dataset. Moreover, the experimental results prove its invariance against different kinds of deformations, transformations and noise. In the future, we plan to test our technique on domain specific benchmarks and explore its adequacy on other applications such as partial matching, segmentation, pose estimation and dense correspondence between non-rigid 3D models.

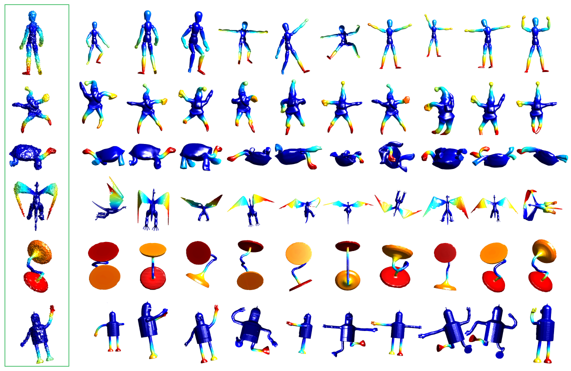

Figure 4: *Sample of retrieval results given corrupted models with noise level = 0.01.*

References

[BBGO11] BRONSTEIN A. M., BRONSTEIN M. M., GUIBAS L. J., OVSJANIKOV M.: Shape google: Geometric words and expressions for invariant shape retrieval. *ACM Trans. Graph. 30*, 1 (Feb. 2011), 1:1–1:20. 1, 2

[BK10] BRONSTEIN M., KOKKINOS I.: Scale-invariant heat kernel signatures for non-rigid shape recognition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (June 2010), pp. 1704–1711. 1

[EHK13] ELNAGHY H., HAMAD S., KHALIFA M. E.: Taxonomy for 3d content-based object retrieval methods. *International Journal of Research and Reviews in Applied Sciences (IJRRAS) 14* (2013), 412–446. 1

[FO09] FURUYA T., OHBUCHI R.: Dense sampling and fast encoding for 3d model retrieval using bag-of-visual features. In *Proceedings of the ACM International Conference on Image and Video Retrieval* (2009), CIVR '09, ACM, pp. 26:1–26:8. 1

[FSKR11] FANG Y., SUN M., KIM M., RAMANI K.: Heat-mapping: A robust approach toward perceptually consistent mesh segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (June 2011), pp. 2145–2152. 1

[LB14] LITMAN R., BRONSTEIN A.: Learning spectral descriptors for deformable shape correspondence. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 36*, 1 (Jan 2014), 171–180. 1

[Lea] LIAN Z., ET AL.: Shrec '15 track: Non-rigid 3d shape retrieval. 1, 3

[OOFB08] OHBUCHI R., OSADA K., FURUYA T., BANNO T.: Salient local visual features for shape-based 3d model retrieval. In *Shape Modeling and Applications, 2008. SMI 2008. IEEE International Conference on* (June 2008), pp. 93–102. 1

[SOG09] SUN J., OVSJANIKOV M., GUIBAS L.: A concise and provably informative multi-scale signature based on heat diffusion. In *Proceedings of the Symposium on Geometry Processing* (2009), SGP '09, Eurographics Association, pp. 1383–1392. 1, 2

[TCF09] TOLDO R., CASTELLANI U., FUSIELLO A.: Visual vocabulary signature for 3d object retrieval and partial matching. In *Proceedings of the 2Nd Eurographics Conference on 3D Object Retrieval* (2009), 3DOR '09, Eurographics Association, pp. 21–28. 1

© The Eurographics Association 2015.