Eurographics Workshop on 3D Object Retrieval (2014)
B. Bustos, H. Tabia, J.-P. Vandeborre, and R. Veltkamp (Editors)

Manifold harmonic transform and spatial relationships for partial 3D object retrieval

Nguyen Vu Hoang [1] & Valérie Gouet-Brunet [2]

[1]CNAM, Vertigo, CEDRIC lab, 292 rue Saint-Martin, 75141 Paris, France
[2]Paris-Est University, IGN/SR, MATIS lab, 73 avenue de Paris, 94160 Saint-Mandé, France

Abstract

This paper presents an approach for 3D object retrieval, dedicated to partial shape retrieval in large datasets. A word manipulation, i.e. quantized descriptors, as in the Bag-of-Words representation is employed, based on the extraction of 3D Harris points and on a local description involving local manifold harmonic transform. By adding Δ-TSR, a triangular spatial information between words, the richness and robustness of this representation is reinforced. The approach is invariant to different geometrical transformations of 3D shape such as translation, rotation, scale and robust to shape resolution changes. We have evaluated it in terms of quality of retrieval, facing several state-of-the-art methods and on different public 3D benchmarks involving different contents and degrees of complexity.

Categories and Subject Descriptors (according to ACM CCS): Modeling [Computer Graphics]: 3D Shape Matching—Representations, data structures, and transforms

1. Introduction

Recently, we have seen an explosion in the number of techniques for manipulating 3D objects. Many authors aim at developing 3DOR (3D Object Retrieval) systems that, given a 3D object query or a 2D image query, provide similar 3D objects. Most of the time, these objects are described in terms of 3D shapes that are often represented as a surface, in particular by polygonal meshes. We know that 3D matching is the process of determining how similar two 3D shapes are, this is often done by computing a distance or a similarity measure between two sets of features. Hence, one of major challenges in the context of 3D data retrieval is to elaborate a description of the object's shape. Serving as a key for the matching process, it decisively influences the relevance of the results. Moreover, content-based retrieval of 3D shapes necessitates the consideration of complex properties, such as the discriminative power of the shape-based description as well as its invariance/robustness under some geometric transformations. A complementary process is indexing, i.e. the process of building a data structure on the features, aiming at speeding up the search in large volumes. Then the whole retrieval process is the combination of description, matching, indexing, searching and delivering of the results from a given query, effectively and efficiently. Most of the time, 3DOR approaches mainly focus on description and matching, yet knowing that the indexing step should influence the system in terms of computational efficiency and of effectiveness.

In this work, we propose an approach that combines a quantized local description as in the Bag-of-Words (BoW) description with a spatial representation of the words for the 3D shape retrieval problem on objects represented with polygonal meshes. Our contributions are double. Firstly, the description, associated with 3D interest points quantized into 3D words, is computed from the projection of the local mesh surface in frequency space by using the local manifold harmonic transform over a large neighboring area of the feature point, differently from [Lav11] where the description is built from the transformation of points' coordinates in new space. This description is very discriminative and moreover quite robust to noise or connectivity changes. No information about the object's structure is considered, making the approach also invariant to isometric deformations or topological changes. Secondly, we consider the geometry between the 3D words by extending the 2D triangular spatial relationships approach of [HGBRM10] to 3D features. The strengths of this approach are its invariance to several geometrical transformations like translation, rotation, scale,

DOI: 10.2312/3dor.20141048

non-rigid or local deformations and cropping, making it particularly efficient for partial shape retrieval, and also to connectivity and shape resolution changes, making it robust to 3D models changes.

The paper is organized as follows: section 2 presents an overview of state of the art approaches for 3DOR, including those using BoW models. The pipeline of our proposal is introduced in section 3. Here, we describe point descriptor based on manifold harmonics transform in section 3.1, present how to embed geometrical relationships information into the local description in section 3.2 and how to compute the similarity in section 3.3. Finally, we experiment and evaluate our approach facing state of the art in section 4, before concluding in section 5.

2. Overview of 3D object retrieval methods

In this section, we revisit the existing works for 3DOR based on shape description. We focus on the analysis of low-level 3D shape features, without any high-level semantic interpretation like in [BJXX13]. We can classify the several approaches encountered in four principal groups:

- Statistic-based approaches such as shape distribution [OFCD02, ILSR02, OMT03, KBLB12, PBB*13], local features distribution [Lav11, TCF09, RABT13, OBMMB09, RBB*11], which propose to index the distribution of descriptors under mathematic forms which characterize the 3D object shape.
- Structural approaches resting upon graph-based models [CR01, ZTS02, TVD07, APP*09, FMA*10] or skeleton-based models [SSGD03, IKL*03]. These methods attempt to describe the structure of 3D objects, e.g. a graph showing how shape components are linked together.
- Transform-based approaches such as spherical harmonics [KFR03], 3D Fourier transform [LBLL11], 3D Zernike moments [NK03], etc., that are based on the transformation of the 3D shape from 3D Euclidean space to frequency space. These approaches achieve rotation invariance.
- View-based approaches such as multi-view-based approaches [CTSO03], Panorama [PPT08, PPTP10]. Here, two 3D models are similar if they look similar from all viewing angles representing projections of these objects on different plans. A natural application of this paradigm is the use of sketch-based query interfaces which allow to define the query under different views. In this case, 3DOR is similar to CBIR (Content-Based Image Retrieval).

In general, the earliest solutions introduced to tackle the problem of 3DOR were based on global descriptors that describe the form of 3D object globally. More recent invariant descriptors are based on some spectral embeddings by using eigenvalues of the Laplace-Beltrami operator or other transformations. The limit of the global descriptors of 3D objects is that they are hardly robust to rigid deformations and not adapted to partial similarity retrieval. To face these problems, some researchers turned their attention to local descriptors associated with salient feature points, following the successful CBIR approaches like SIFT [DJLW08]. In the 3D case, however, scenes can undergo a variety of non-rigid deformations such as variations in local scale, variation in the topology of the observed mesh, and even global affine deformations or warping effects due. Furthermore, the fact is that, in 3D shape of the most of cases, we do not have any information like texture, color, then existing 2D retrieval techniques are difficult to adapt to 3DOR directly. In the last years, based on the proposal of image feature detectors, different 3D feature detectors were proposed: 3D Harris point detector [SB11], multi-scale local descriptor [SOG09], SHOT descriptor [TS10], another feature detector based on an eigen decomposition of the Laplace-Beltrami operator [RPSS09], and a detector related to surface protrusions that creates and matches regions using a graph matching technique based on the Earth Mover's Distance [APP*09].

Following CBIR trends, some 3DOR techniques, resting upon the BoW models, were also published. In [LGS10], 3D models are seen as a set of 2D views which are indexed with 2D SIFT features. [LG09] and [LZQ06] propose BoW approaches based on Spin Images descriptors computed from dense feature points. [TCF09] segments 3D shapes into regions, then each region is associated with several descriptors and thus several visual words. [Lav11] considers a 3D object as an histogram of local feature points detected by using a Voronoi distribution algorithm and classified as words, knowing that each point is associated with a descriptor computed from the Fourier transformation of the local area around it. In general, all these proposals provide good retrieval results on the classical 3DOR benchmarks. However, some recurring drawbacks can be mentioned: the descriptors used are relatively poor, because encapsulating a local and low-level information only. To address this problem, it is possible to encapsulate an information about the local geometry between key points, such as in [Lav11] which considers the spatial cooccurrences of couples of words. It is also possible to improve the step of matching, such as [RABT13] which exploits the game theory to improve registration of point sets and provide very good results on the complex Gun benchmark.

3. Our approach

Our proposal can be classified as a statistic-based approach (see section 2). Its pipeline, illustrated in Fig.1, is as following: each 3D object is considered as a collection of local feature points that are detected with the 3D Harris detector [SB11]. According to results from [DCG12], 3D Harris has a global better performance facing other sparse detectors. It delivers salient points robust to different transformation like translation, rotation, scaling and resolution change. Then each detected point is associated with a local area around it in the mesh ; its construction is described in appendix A. On each neighboring area, we compute an improved descriptor, presented in section 3.1 and based on the manifold harmonic transform using the Laplace-Beltrami operator . Then these 3D point descriptions are classified as in a BoW representation, by using a scalable clustering algorithm such as hierarchical k-means [NS06]. Hence, each object is finally described by the corresponding distribution of involved words. In section 3.2, we will enrich this representation by embedding some spatial information between words and by indexing them into a dedicated access method. In section 3.3, we present how to compute the similarity.

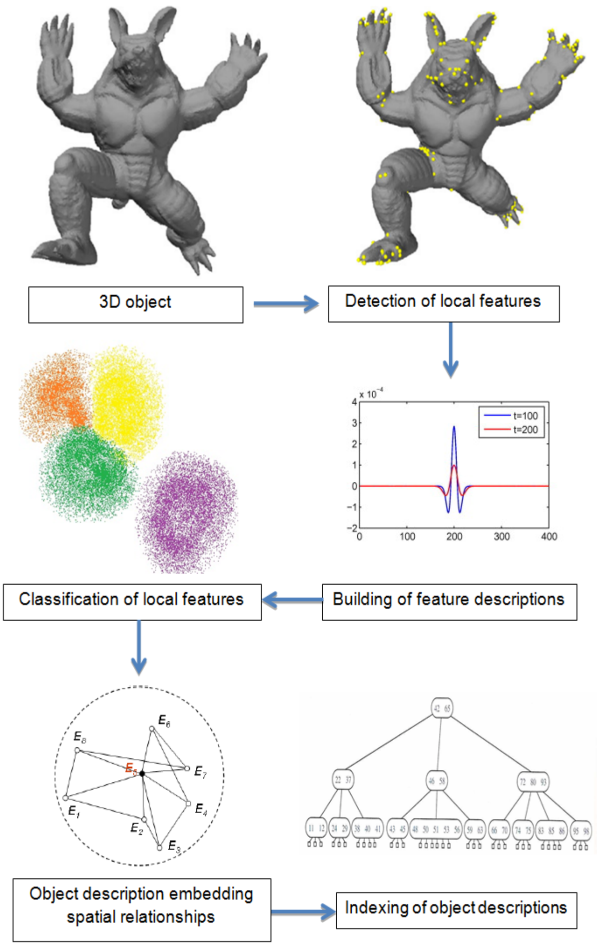

Figure 1: *Pipeline of our approach.*

3.1. Robust description of 3D interest points

Spectral methods like Discrete Cosine Transform and Discrete Fourier Transform are widely used for analyzing signals in image processing. It is well known that the eigenfunctions of the Laplace-Beltrami operator (Manifold Harmonics) define a function basis allowing for a generalization of the Fourier Transform to manifolds. In [VL08], the authors propose to use this operator in the Euclidean space for noise reduction of 3D object representation. Based on this idea, we observe that the use of the Manifold Harmonic Transform (MHT) on a 3D shape can provide a robust description for this last one. MHT is the transformation of each coordinate in the initial geometry into frequency space by using the Manifold Harmonics Basis (MHB). The new coordinates are also called the spectral coefficients. There are only small variations on the spectral amplitudes of a surface area which can be distorted under noise addition according to [Lav11]. Our idea is to exploit the local spectral amplitudes on a surface area around a given 3D interest point to describe this one. The MHB is defined with a set of eigenvectors of the discrete Laplacian-Beltrami $\bar{\Delta}$ expressed in the canonical basis:

1. Build $\bar{\Delta}$. It is a symmetric matrix, and its coefficients are given by:

$$\bar{\Delta}_{ij} = -\frac{cot\beta_{ij} + cot\beta'_{ij}}{\sqrt{|v_i^*||v_j^*|}} \quad \text{and} \quad \bar{\Delta}_{ii} = -\sum_j \bar{\Delta}_{ij} \qquad (1)$$

   where β_{ij} and β'_{ij} are the two angles opposite to the edge between vertices v_i and v_j (v_i and v_j are simply vertices on given area), $|v_i^*|$ is the surface size computed from the set of neighboring triangles around vertex v_i. The eigenfunction and eigenvalue pairs $(H_k;\lambda_k)$ of this operator satisfy the following relationships: $-\Delta H^k = \lambda_k H^k$.
2. Compute its eigenvectors H_k. The set of (H_k) vectors is called the MHB. This vector is invariant to rotation and scale of the 3D object.

The spectral coefficients $\tilde{x}_k$ (resp. $\tilde{y}_k$, $\tilde{z}_k$) are then calculated as the inner product between the initial geometry x (resp. y,z) and the sorted eigenvectors H_k:

$$\tilde{x}_k = \sum_{i=1}^{m} x_i |v_i^*| H_i^k \qquad (2)$$

The k^{th} spectral coefficient amplitude is then defined as $c^k = \sqrt{(\tilde{x}_k)^2 + (\tilde{y}_k)^2 + (\tilde{z}_k)^2}$. This coefficient c^k is used in the approach of [Lav11]. Hence, for a given area A_i around a feature point p_i having coordinates (x_i, y_i, z_i), the descriptor is the spectral amplitude vector $C_i = [c_i^1;..c_i^{nc}]$, with c_i^k, the k^{th} spectral coefficient amplitude of A_i. The descriptor for a given point is built from the n_c first spectral coefficients in order to limit the descriptor to more robust low/medium frequencies. This descriptor has some theoretical robustness properties: under a translation, only the first coefficient c^0 is modified, hence the authors of [Lav11] do not consider c^0 in their descriptor and thus obtain translation robustness. Meanwhile, a rotation in the Euclidean domain yields the same rotation in the spectral domain $(\tilde{x}, \tilde{y}, \tilde{z})$. Under a uniform scaling with a factor s, all the spectral coefficients will be scaled by s^2. Hence this descriptor is not robust to rotation and to scaling, therefore, the 3D object has to be normalized to unit before processing, to ensure invariance to scale and to rotation [Lav11].

To avoid normalizing the 3D shapes, knowing that this processing adds a computation cost and may not be adapted to partial retrieval scenarios where objects may be inserted in complex scenes, we do not use classical spectral coefficient amplitude like [Lav11] and prefer to do the projection of the mesh surface information around the points in the new frequency space. We simply define the coefficients amplitude as:

$$c^k = \sum_{i=1}^{m} |v_i^*|^2 H_i^k \tag{3}$$

Each 3D Harris point v_i is associated with a local area A_i for which we compute description C_i (see appendix A). This description is robust to different transformations: a translation or a rotation does not modify any coefficient c^k. We know that under a uniform scaling with a factor s, all the spectral coefficients are scaled by s^2. Hence, to be robust to scaling, we normalize the whole description by dividing each c^i by c^0, which has the lowest and less noisy frequency. Consequently, unlike several other approaches , the normalization of the object is not required to ensure robustness to scale change, limiting thus the processing complexity and making the description more robust to 3D deformations and partial retrieval.

At the end of this step, we quantize all the descriptions obtained, as in BoW representations, which provides a set of 3D words per object. The construction of the dictionary is made scalable by using a hierarchical k-means [NS06]. In the following, this representation of the object is called Harris_MHB.

3.2. Spatial relationship description

Traditionally, most of the BoW representations do not encapsulate any information about the spatial layout of the words. We propose to describe the spatial relationships between 3D words by extending approach Δ-TSR [HGBRM10], originally designed for CBIR, to 3D objects. By extension, each 3D object O is represented by a set Δ-TSR(O) containing the description S of all the triangular relationships between triplets of 3D points (E_i,E_j,E_k) such as:

$$\Delta\text{-TSR}(O) = \begin{array}{l}\{S(E_i,E_j,E_k)/E_i,E_j,E_k \in O;\\ i,j,k \in [1,N_O];\ L_i \geq L_j \geq L_k\}\end{array} \tag{4}$$

with N_O the number of points in O and (L_i,L_j,L_k) the Harris_MHB word's labels associated to the triplets of points. S can encapsulate several kinds of information, such as the geometrical relationships between the points. As in [HGBRM10], we keep information on the angles of the triangle formed by (E_i,E_j,E_k). In addition, we consider an orientation of the point by using the concave-convex measure at point location, based on curvature analysis. From these attributes, O can be represented by a set of 5-dimensional description called Δ-TSR$_{5D}(O)$. Each triplet description, called S^o, presents the triangular relationships of triangle (E_i,E_j,E_k) and its symmetric such as:

$$S^o(E_i,E_j,E_k) = (K_1,K_2,K_3,K_4,K_5) \tag{5}$$

$$\text{with} \begin{cases} K_1 = (L_i-1)n_w{}^2 + (L_j-1)n_w + (L_k-1) \\ K_2 = a_i; K_3 = a_j;\ K_2,K_3 \in [0°,180°] \\ K_4 = \frac{o_i}{o_k};\ o_i,o_j,o_k \in [0°,360°] \\ K_5 = \frac{o_j}{o_k}; \end{cases}$$

n_w is the size of the dictionary. K_1 is the unique coding of word's labels from the vocabulary. a_i, a_j are the angles of vertices E_i, E_j respectively. K_4 and K_5 represent the relative orientation of E_i and E_j with respect to E_k, in order to maintain invariance to rotation. We build the orientation information based on principal curvatures. Let denote o_i, o_j and o_k the orientation information of E_i, E_j and E_k knowing that $o = |\lambda_{max} - \lambda_{min}|$, λ_{max} and λ_{min} are the direction of the largest principal curvature and the smallest principal curvature respectively passing through interest point E (see [ASWL11]).

To be more robust to partial retrieval, as in [HGBRM10], only the smallest triangles in Δ-TSR are kept, which involve triplets of points located close to each other in the 3D object: for each interest point p, we build only the triangles between p and other interest points that are the neighbors of p (see appendix A). Similarly, the step of retrieval of nearest neighbors of S^o is performed optimally by using the index structure B-tree with composite keys.

3.3. Similarity measure associated with Δ-TSR

The similarity between two 3D objects can be established by the ratio of similar triangle descriptions between them. Thus, the 3DOR problem is essentially the problem of matching between descriptions $S^o(T_Q)$ and $S^o(T_O)$ of a query triangle T_Q and a database triangle T_O. The associated similarity measure between two triangle descriptions, called *sim*, is the same as the one originally proposed in [HGBRM10]:

$$sim = \begin{cases} sim^o(S^o(T_Q), S^o(T_O)) \\ \quad \text{if } K_1(T_Q) = K_1(T_O) \\ \quad \text{and } S^o(T_O) \text{ validates the tolerance} \\ \quad \text{interval } \delta_o \\ 0 \text{ otherwise} \end{cases} \tag{6}$$

where:

$$sim^o(.,.) = \frac{1}{2}[f(T_Q,T_O,2,\delta_a) + f(T_Q,T_O,4,\delta_o)] \tag{7}$$

$$f(T,T',i,\delta) = \begin{cases} 1 \text{ if } \delta = 0 \\ \frac{1}{2}\sum_{t=i}^{i+1}(1 - \frac{|K_t(T)-K_t(T')|}{\delta}) \text{ if } \delta \neq 0 \end{cases} \tag{8}$$

δ_a and δ_o are tolerance thresholds used to define the similarity between components $K_{\{2,3,4,5\}}$ in the two descriptions compared [HGBRM10]. The *sim* measure varies in interval $[0,1]$ and increases with the similarity.

4. Experiments and evaluation

In the following sections, we evaluate the relevance of the local description Harris_MHB proposed in section 3.1 as well as of the description of spatial relationships Δ-TSR presented in section 3.2.

We consider different public benchmarks to evaluate our approach facing several state-of-the-art approaches. Given the length limit of the paper, in this section we present only the results obtained on the following two public 3D benchmarks:

- The collection SHREC'07[†], called "Shape Retrieval Contest of 3D Face Models" [BV07], consists of 1500 different instances of 3D face models (see examples in the first row of Fig.2).
- The collection SHREC'13[‡], called "Large-Scale Partial Shape Retrieval using Simulated Range Images" [SMB*13], consists of 20 object classes with 18 models per class and 7200 queries (see examples in the second row of Fig.2)

Harris_MHB is compared with state of the art in section 4.1 on the benchmark SHREC'07. Sections 4.2 compare Harris_MHB including the spatial relationships to the two other approaches submitted to benchmark SHREC'13. In our technical report [HGB14], we present more experiments on 5 public benchmarks, facing other approaches.

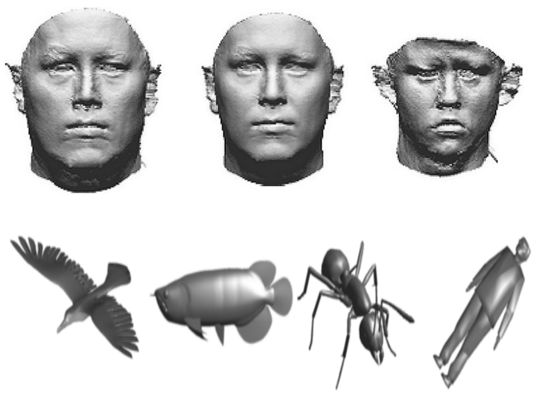

Figure 2: *Some examples from SHREC'07 and SHREC'13 benchmarks.*

Harris_MHB is based on two principal parameters: the dictionary size n_w and the number of coefficients n_c of the spectral descriptor. The descriptor is almost independent of the neighborhood size while n_c is smaller than the number of neighborhood points. Indeed, the descriptor is build from n_c lowest frequency amplitudes. In our experiments, the neighborhood size is 10% of the object size. Before comparing our approach with state of the art, we varied these two parameters on 3D benchmarks to choose the best configurations (see more detail in [HGB14]). We obtained, $n_c = 40$ and $n_w = 2000$ the best parameter on thees benchmarks. The following experiments are done with these parameters.

† http://ensor.labs.cs.uu.nl/shrec/shrec2007
‡ http://dataset.dcc.uchile.cl

4.1. Comparison of different descriptions

Now, we examine the performance of several 3D point local descriptions on the 3D database SHREC'07: Harris_MHB, our implementation of the approach of Lavoué [Lav11] (in its version with the best configurations of the author and without the description of spatial relationships), the one of Toldo et al. [TCF09] (public authors implementation), which are all three based on BoW representations, and our implementation of the global description based on 3D Zernike moments [NK03]. Note that we do not consider any spatial relationship information in these approaches. Fig. 3 presents the Precision/Recall curves obtained. The two methods of [Lav11] and [TCF09] present quite comparable performances, however [Lav11] is slightly better: its MAP (Mean Average Precision) is 0.682 and the one of [TCF09] is 0.615. On this 3D database, the global description 3D Zernike cannot show its relevance because it involves a global description not sufficiently discriminative on this benchmark of faces where global shapes are very similar; its MAP is only 0.393. Harris_MHB globally proves its efficiency, with a MAP of 0.707, except for very large recalls where [Lav11] becomes better.

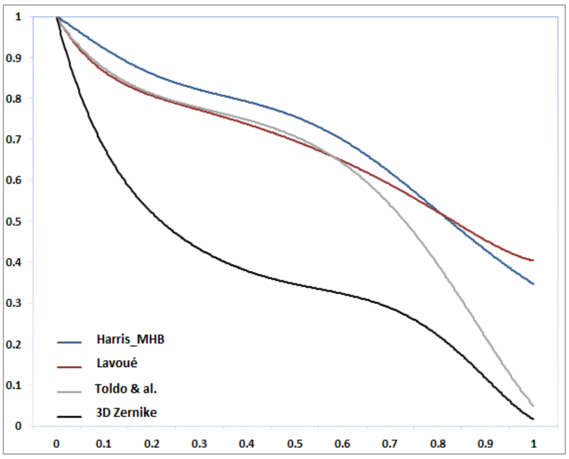

Figure 3: *Precision/Recall curves for different descriptions on SHREC'07.*

On this benchmark where the global shapes are very similar, we have observed that the use of a spatial relationship description (*i.e.* Harris_MHB+Δ-TSR and the full version of [Lav11]) improves the quality of retrieval only slightly; see the corresponding curves in [HGB14].

4.2. Other comparison

In this section, we compare our full proposal, Harris_MHB+Δ-TSR, to two recent methods submitted in the SHREC'13 Track:

- Range Scan-Based 3D Model Rettrieval by Incorporating 2D-3D Alignment ([LLJ12, BBF12]). For abbreviation, this method is called as Li-Lu-Johan.
- Partial Shape Retrieval with Spin Images and Signature Quadratic Form Distance ([SB12]). For abbreviation, this method is called as Sipiran-Bustos.

The comparison is realized on the regular measures used in [SMB*13]: MAP, NN (Nearest Neighbor), FT (First Tier) and ST (Second Tier). Fig. 4 depicts the precision-recall plot and Tab.1 summarizes the results with other performance metrics. From the precision-recall plot, we note the superior performance of our method. This can be also evidenced by the results of performance measures in Table 1.

According to [SMB*13], the performance difference of the Li-Lu-Johan method in regards to the Sipiran-Bustos method can be explained by two reasons. On one hand, the Li-Lu-Johan method obtains a set of 81 views for each model in the target set. Therefore, the probability of similarity between the partial query and a sampled view is high. On the other hand, the computation of spin images of the Sipiran-Bustos method in partial views shows some inconvenience, many keypoints are located close to the boundary of a partial query image which affects the computation of the local descriptors. Our method does not depend on the view projection of 3D object. The distribution of interest point is almost homogeneous. Movover, with Δ-TSR, the information on triangular spatial relationships and on orientation of 3D points reinforces the object description from partial query image. It demonstrates its power of description facing other approaches, its MAP is 0.3434.

To gain insight into the behavior of the proposal, a class-by-class evaluation of our approach is shown in Tab.2. The detail of class-by-class evaluation of two other approaches can be found also in [SMB*13]. In this table, we show a more detailed evaluation of our approaches from the point of view of the effectiveness in each class of the benchmark.

Table 1: *Performance measures on SHREC'13 benchmark. The best results are shown in bold type.*

	Li-Lu-Johan	Sipiran-Bustos	Δ-TSR
NN	0.3444	0.3108	**0.3501**
FT	0.2116	0.2043	**0.2976**
ST	0.1675	0.1576	**0.2994**
MAP	0.2247	0.1978	**0.3434**

5. Conclusion

In this paper, we have proposed an efficient approach for 3D object retrieval, dedicated to partial shape retrieval and large datasets. A BoW representation is employed, based on the

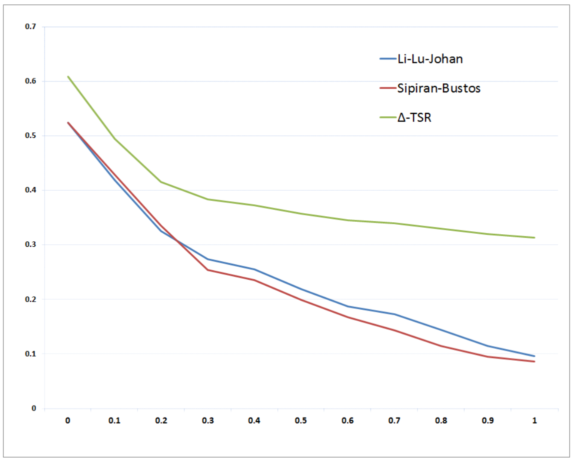

Figure 4: *Precision/Recall curves for the different approaches on SHREC'13 benchmark.*

Table 2: *Performance measures of Δ-TSR by class on SHREC'13 benchmark.*

	NN	FT	ST	MAP
Bird	0.1667	0.2974	0.2876	0.3269
Fish	0.5556	0.3137	0.2974	0.3382
Insect	0.2778	0.2712	0.2794	0.3323
Biped	0.5556	0.2876	0.2778	0.3421
Quadruped	0.4444	0.3366	0.2892	0.3470
Bottle	0.3333	0.3039	0.3186	0.3602
Cup	0.2778	0.2876	0.2761	0.3231
Mug	0.3333	0.3039	0.3023	0.3574
Floorlamp	0.3889	0.2876	0.2958	0.3320
Desklamp	0.3889	0.3137	0.3105	0.3709
Cellphone	0.2222	0.2614	0.2876	0.3197
Deskphone	0.3333	0.2810	0.2843	0.3321
Bed	0.4444	0.3235	0.3105	0.3789
Chair	0.3333	0.2843	0.2941	0.3268
Wheel Chair	0.2778	0.2745	0.2680	0.3113
Sofa	0.2222	0.2614	0.2565	0.3139
Biplane	0.2778	0.3235	0.3154	0.3490
Monoplane	0.5556	0.3072	0.3219	0.3553
Car	0.2778	0.2843	0.2876	0.3269
Bicycle	0.3333	0.3137	0.3464	0.3740

extraction of 3D Harris points and on a local description involving local Fourier descriptors both fast to compute and discriminative. By adding a triangular spatial information between words, the robustness of this representation is reinforced. The experimental evaluations performed on two public 3D benchmarks involving different contents and degrees of complexity, facing several state-of-the-art techniques (additional experiments are presented in [HGB14]), have provided encouraging results in terms of quality of retrieval. To improve the quality of retrieval even more, especially on complex datasets, it could be interesting to combine our approach with the one of [RABT13] which focuses on robust point sets matching. Here, the main challenge would be to adapt it in order to reduce its complexity and then maintain scalability.

Appendix A: Selection of neighboring points

The selection of neighboring points around a given 3D point (vertex) v in a mesh is necessary to compute derivatives as well as to provide an area for point description. There are different solutions: it is possible to select the number of rings around v if the object tessellation is uniform, this method is called k-ring selection. For a given vertex v in the set of vertices V, its k-ring neighborhood is defined as

$$ring_k(v) = \{w \in V | shortest\_path\_size(w, v) <= k\}$$

For irregular and complex meshes, an adaptive neighborhood selection may be more efficient. In this case, a classical k-ring may provide a very large or a very small area around v. It is possible to collect the neighborhood points by adding a condition based on distances between points in the mesh. The distance on surface from a point v to w is defined as: $d_s(v, w) = shortest\_path\_len(v, w)$. Finally, we prefer to choose the neighboring points around vertex v as:

$$\eta(v) = \{w \in ring_k(v) \mid k <= K \wedge d_s(v, w) <= \lambda\}$$

where K and λ are parameters; for example, λ can be a fraction of the diagonal of the object bounding rectangle.

References

[APP*09] Agathos A., Pratikakis I., Papadakis P., Perantonis S., Azariadis P., Sapidis N.: Retrieval of 3d articulated objects using a graph-based representation. *Eurographics Workshop on 3D Object Retrieval* (2009). 2

[ASWL11] An Y., Shao C., Wang X., Li Z.: Estimating principal curvatures and principal directions from discrete surfaces using discrete curve model. *Journal of Information & Computational Science* (2011), 296–311. 4

[BBF12] Bronstein M., Bronstein M. M., Ferreira A. M.: Sketch-based 3d shape retrieval. *Eurographics workshop on 3D Object Retrieval* (2012). 6

[BJXX13] Boqing G., Jianzhuang L., Xiaogang W., Xiaoou T.: Learning semantic signatures for 3d object retrieval. *IEEE Transaction on Multimedia* (2013). 2

[BV07] Blanz V., Vetter T.: Shrec'07 track: Shape retrieval contest of 3d face models. *Eurographics Workshop on 3D Object Retrieval* (2007). 5

[CR01] Cicirello V., Regli W. C.: Machining feature-based comparisons of mechanical parts. *Solid Modeling* (2001), 176–185. 2

[CTSO03] Chen D. Y., Tian X. P., Shen Y. T., Ouhyoung M.: On visual similarity based 3d model retrieval. *Computer Graphics Forum* (2003). 2

[DCG12] Dutagaci H., Chun C., Godil A.: Evaluation of 3d interest point detection techniques via human-generated ground truth. *The Visual Computer* (2012), 901–917. 2

[DJLW08] Datta R., Joshi D., Li J., Wang J. Z.: Image Retrieval: Ideas, Influences, and Trends of the New Age. *ACM Computing Surveys* (2008), 1–60. 2

[FMA*10] Ferreira A., Marini S., Attene M., Fonseca M. J., Spagnuolo M., Jorge J. A., Falcidieno B.: Thesaurus-based 3d object retrieval with part-in-whole matching. *Int. J. Comput. Vision* (2010). 2

[HGB14] Hoang N., Gouet-Brunet V.: *A new description for scalable 3D partial object retrieval*. Research report CEDRIC-14-2906, CEDRIC, CNAM, 2014. 5, 7

[HGBRM10] Hoàng N., Gouet-Brunet V., Rukoz M., Manouvrier M.: Embedding spatial information into image content description for scene retrieval. *Pattern Recognition* (2010), 3013–3024. 1, 4

[IKL*03] Iyer N., Kalyanaraman Y., Lou K., Janyanti S., Ramani K.: A reconfigurable 3d engineering shape search system. *Design Engineering Technical Conference* (2003). 2

[ILSR02] Ip C., Lapadat D., Sieger L., Regli W.: Using shape distributions to compare solid models. *Solid Modeling and Application* (2002), 273–280. 2

[KBLB12] Kokkinos I., Bronstein M. M., Litman R., Bronstein A. M.: Intrinsic shape context descriptors for deformable shapes. *CVPR* (2012). 2

[KFR03] Kazhdan; M., Funkhouser; T., Rusinkiewicz S.: Rotation invariant spherical harmonic representation of 3d shape descriptors. *Symposium on Geometry Processing* (2003). 2

[Lav11] Lavoue G.: Bag of words and local spectral descriptor for 3d partial shape retrieval. *Eurographics Workshop on 3D Object Retrieval* (2011). 1, 2, 3, 4, 5

[LBLL11] Liu Y., Bao F., Li Z., Li H.: 3d model retrieval based on 3d fractional fourier transform. *The International Arab Journal of Information Technology* (2011). 2

[LG09] Li X., Godil A.: Exploring the bag-of-words method for 3d shape retrieval. *International Conference on Image Processing* (2009). 2

[LGS10] Lian Z., Godil A., Sun X.: Visual similarity based 3d shape retrieval using bag-of-features. *Shape Modeling International* (2010). 2

[LLJ12] Li B., Lu Y., Johan H.: Sketch-based 3d model retrieval by incorporating 2d-3d alignment. *Multimedia Tools and Applications* (2012). 6

[LZQ06] Liu Y., Zha H., Qin H.: Shape topics: A compact representation and new algorithms for 3d partial shape retrieval. *Computer Vision and Pattern Recognition* (2006). 2

[NK03] Novotni M., Klein R.: 3d zernike descriptors for content based shape retrieval. *ACM symposium on Solid modelling and applications* (2003), 216 – 225. 2, 5

[NS06] Nister D., Stewenius H.: Scalable recognition with a vocabulary tree. *Conference on Computer Vision and Pattern Recognition* (2006), 2161–2168. 3, 4

[OBMMB09] Ovsjanikov M., Bronstein A. M., M. M. Bronstein a. L. J. G.: Shape google: a computer vision approach for invariant shape retrieval. *Workshop on Nonrigid Shape Analysis and Deformable Image Alignment* (2009). 2

[OFCD02] Osada R., Funkhouser T., Chazelle B., Dobkin D.: Shape distributions. *ACM Transactions on Graphics* (2002), 807–832. 2

[OMT03] Ohbuchi R., Minamitani T., Takei T.: Shape-similarity search of 3d models by using enhanced shape functions. *Theory and Practice of Computer Graphics* (2003), 97–104. 2

[PBB*13] Pokrass J., Bronstein A. M., Bronstein M. M., Sprechmann P., Sapiro G.: Sparse modeling of intrinsic correspondences. *Computer Graphics Forum* (2013). 2

[PPT08] Papadakis P., Pratikakis I., Theoharis T.: 3d object retrieval using an efficient and compact hybrid shape descriptor. *Eurographics Workshop on 3D Object Retrieval* (2008). 2

[PPTP10] Papadakis P., Pratikakis I., Theoharis T., Perantonis S.: Panorama: A 3d shape descriptor based on panoramic views for unsupervised 3d object retrieval. *International Journal of Computer Vision* (2010). 2

[RABT13] Rodolà E., Albarelli A., Bergamasco F., Torsello A.: A scale independent selection process for 3d object recognition in cluttered scenes. *International Journal of Computer Vision* (2013). 2, 7

[RBB*11] Raviv D., Bronstein A. M., Bronstein M. M., Kimmel R., Sochen N.: Affine-invariant diffusion geometry for the analysis of deformable 3d shapes. *Computer Vision and Pattern Recognition* (2011). 2

[RPSS09] Ruggeri M., Patante G., Spagnuolo M., Saupe D.: Spectral-driven isometry-invariant matching of 3d shapes. *International Journal of Computer Vision* (2009). 2

[SB11] Sipiran; I., Bustos B.: Harris 3d: a robust extension of the harris operator for interest point detection on 3d meshes. *The Visual Computer* (2011), pp 963–976. 2

[SB12] Sipiran I., Bustos B.: Key-component detection on 3d meshes using local features. *Eurographics Workshop on 3D Object Retrieval* (2012). 6

[SMB*13] Sipiran I., Meruane R., Bustos B., Schreck T., Li B., Lu Y., Johan H.: Shrec'13 track: Large-scale partial shape retrieval using simulated range images. *Eurographics Workshop on 3D Object Retrieval* (2013). 5, 6

[SOG09] Sun J., Ovsjanikov M., Guibas L.: A concise and provably informative multi-scale signature based on heat diffusion. *Computer Graphics Forum* (2009). 2

[SSGD03] Sundar H., Silver D., Gagvani N., Dickenson S.: Skeleton based shape matching and retrieval. *Shape Modeling International* (2003), 130–139. 2

[TCF09] Toldo R., Castellani U., Fusiello A.: Visual vocabulary signature for 3d object retrieval and partial matching. *Eurographics Workshop on 3D Object Retrieval* (2009). 2, 5

[TS10] Tombari F., Salti S.: Unique signatures of histograms for local surface description. *European Conference on Computer Vision* (2010). 2

[TVD07] Tierny J., Vandeborre J.-P., Daoudi M.: Reeb chart unfolding based 3d shape signatures. *Eurographics* (2007). 2

[VL08] Vallet B., Levy B.: Spectral geometry processing with manifold harmonics. *Eurographics Workshop on 3D Object Retrieval* (2008). 3

[ZTS02] Zuckerberger E., Tal A., Shlafman S.: Polyhedral surface decomposition with applications. *Computers and Graphics* (2002), 733–743. 2