CEDRIC Research Report n° CEDRIC-14-2906

A new description for scalable 3D partial object retrieval

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

[1] CEDRIC/CNAM - 292, rue Saint-Martin - F75141 Paris Cedex 03, France
[2] Paris-Est University, IGN/SR, MATIS lab - 73 avenue de Paris - F94160 Saint-Mandé, France
nguyenvu.hoang@cnam.fr , valerie.gouet@ign.fr

Abstract

This paper presents an approach for 3D object retrieval, dedicated to partial shape retrieval in large datasets. A Bag-of-Words representation is employed, based on the extraction of 3D Harris points and on a local description involving local Fourier descriptors. 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. A dedicated disk-based indexing structure is also employed to make the whole description effective on large-scale datasets. We have evaluated it in terms of quality of retrieval and scalability, facing several state-of-the-art methods and on five public 3D benchmarks involving different contents and degrees of complexity.

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, a 2D image query or a formulation, 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. According to [FMK$^+$03, SWS$^+$00], a simple solution is to annotate the entities with keywords. However, due to the inherent complexity of 3D data, this is incomplete, insufficient and impractical. 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 (see section 2), yet knowing that the indexing step should influence the system in terms of computational efficiency and then of effectiveness.

In this work, we propose an approach that combines a standard Bag-of-Word (BoW) descriptor with a spatial representation of the words, and a dedicated index structure, for the 3D shape retrieval problem on objects represented with polygonal meshes. Our contributions are triple. Firstly, the description, associated with 3D interest points, is computed from the local Fourier spectrum over a large neighboring area of the feature, then it is very discriminative and moreover quite robust to noise or connectivity changes. No information about the object's structure is considered, making the approach invariant to isometric deformations or topological changes. Secondly, we consider the geometry of the words by embedding triangular spatial relationships between them into the description. The strengths of this approach are its invariance to several geometrical transformations like translation, rotation, scale, 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. Additionally, we associate a dedicated access method to this representation, aiming at performing 3DOR efficiently in large-scale databases of 3D objects.

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 also describe both the feature point detector and its descriptor based on manifold harmonics transformation. Section 4 presents how to embed geometrical relationships information into the local description. Finally, we experiment and evaluate our approach facing state of the art in section 5, before concluding in section 6.

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]), local features distribution ([Lav11, TCF03, RABT13, OBMMB09, RBB+11]), which propose to index the distribution of descriptors under mathematic forms which characterize the 3D object shape. It can be the analysis of the global 3D shape, or local features such as curvatures, surface points, angle measurements, elementary volumes, etc.

— Structural approaches resting upon graph-based models ([CR01, ZTS02, TVD07, APP+09]) 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. However, by sampling the space only in radial direction the latter descriptors do not capture fine object coherence in this direction.

— 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. In addition, they are not adapted to partial similarity retrieval, useful when local deformation, cropping, clutter or occlusion is present, and are far to deal with topological changes. 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]. The matching problem in 3D scenes shares many aspects with CBIR : the common goal is to find the relation between a model and its transformed instance in the scene. 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. [TCF03] 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. 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 ; however, the complexity of the solution makes it clearly not applicable to large datasets. Another problem comes from the key points themselves, which may be not very robust regarding connectivity and topological changes (as the segmentation used in [TCF03]). In case of dense set of points, it is necessary to insure their uniform distribution over the surface, even with irregular connectivity. In [Lav11] a connectivity independent uniform sampling of the feature points is employed to face this problem, based on Lloyd relaxation. However, one of the limits of [Lav11] is the requirement of the normalization of the 3D shape, making it less robust for partial retrieval with

complex scenes. Finally, another problem is that most of the encountered approaches are not scalable, or were not evaluated for this purpose.

3 Main concepts of 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] revisited in section 3.1. 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 section 3.2. On each neighboring area, we compute an improved descriptor, presented in section 3.3 and based on the Laplace-Beltrami operator. Then these 3D point descriptions are quantified into words 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. Note that in section 4, we will enrich this representation by embedding some spatial information between words and by indexing them into a dedicated access method.

3.1 Selection of interest points with the 3D Harris detector

In [HS88], the Harris detector was proposed for the detection of interest points in images ; this technique has been popular due to its good repeatability facing rotation, illumination variation and image noise. The Harris detector is based on the local autocorrelation function of a signal. The authors proposed to analyze the eigenvalues matrix (E) of each image pixel, by assigning to each pixel the following value : $h(x, y) = det(E) - k(tr(E))^2$ with k constant.

To adapt the detector to the 3D problem, [Glo09] and [SB11] suggested some approaches, and in this work we consider them as robust vertex detector on 3D objects modeled with polygonal meshes. To calculate the partial derivatives involved in E, according to [SB11], we can fit a quadratic surface to the set of points. For vertex $v(x, y, z)$, we find a paraboloid of the form :

$$z = f(x, y) = \frac{p_1}{2}x^2 + p_2xy + \frac{p_3}{2}y^2 + p_4x + p_5y + p_6. \quad (1)$$

Then, at v the derivatives are : $f_x = \frac{\delta f(x,y)}{\delta x}$, $f_y = \frac{\delta f(x,y)}{\delta y}$

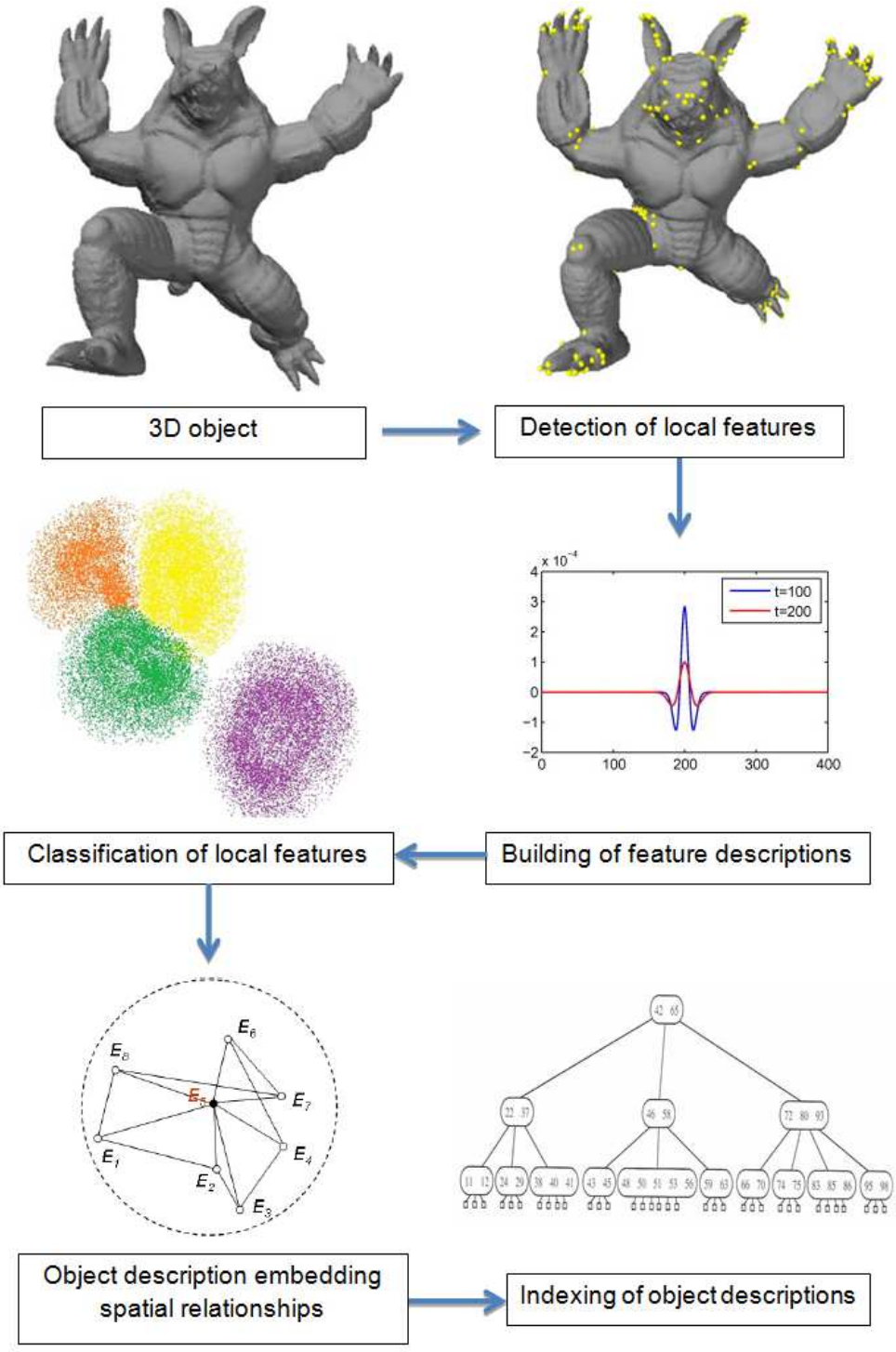

FIGURE 1 – Pipeline of our approach.

We can formulate matrix E of Harris operator at v by $E = \begin{pmatrix} A & C \\ C & B \end{pmatrix}$ knowing that :

$$A = \frac{1}{\sqrt{2\pi\sigma}} \int_{R^2} e^{\frac{s}{2\delta^2}} . f_x^2 dxdy = p_4^2 + 2p_1^2 + 2p_2^2 \quad (2)$$

$$B = \frac{1}{\sqrt{2\pi\sigma}} \int_{R^2} e^{\frac{s}{2\delta^2}} . f_y^2 dxdy = p_5^2 + 2p_2^2 + 2p_3^2 \quad (3)$$

$$C = \frac{1}{\sqrt{2\pi\sigma}} \int_{R^2} e^{\frac{s}{2\delta^2}} . f_x(x,y)^2 . f_x^2 dxdy \quad (4)$$

$$= p_4 p_5 + 2p_1 p_2 + 2p_2 p_3 \quad (5)$$

where δ is a constant that defines the support of Gaussian function and $s = -(x^2 + y^2)$. For computing the derivatives, the neighborhood selection is done as explained in equation 3.2 of section 3.2. Then vertex v is selected if its Harris operator value $h(v)$ is a local maximum in $ring_1(v)$. Finally, we select the vertices having the highest Harris responses, in order to maintain a constant fraction of keypoints depending on the application. In the experiments of section 5, this fraction is at least 10% of the total number of vertices of the 3D object mesh.

3.2 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.

3.3 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 strong noise addition. This is the reason why the wavelet filtering of [VL08] can provide an efficient representation of the shapes. Similarly as in [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 :

$$\bar{\Delta}_{ij} = -\frac{cot\beta_{ij} + cot\beta'_{ij}}{\sqrt{|v_i^*||v_j^*|}} \quad \text{and} \quad \bar{\Delta}_{ii} = -\sum_j \bar{\Delta}_{ij} \tag{6}$$

   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 Laplace-Beltrami operator Δ is the counterpart of the Laplace operator in the Euclidian space. It is defined as the divergence of the gradient for functions defined over manifolds. 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.

Then MHT is the transformation of each coordinate in the initial geometry into frequency space by using MHB. The new coordinates are also called the spectral coefficients. 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 :

$$\tilde{x}_k = \sum_{i=1}^{m} x_i |v_i^*| H_i^k \tag{7}$$

The k^{th} spectral coefficient amplitude is then defined as $c_k = \sqrt{(\tilde{x}_k)^2 + (\tilde{y}_k)^2 + (\tilde{z}_k)^2}$. 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. In [Lav11], 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 interesting 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, the MHB are kept unchanged under isometric transformations, therefore, a rotation in the Euclidean domain yields the same rotation in the spectral domain $(\tilde{x}, \tilde{y}, \tilde{z})$, without any influence on the coefficient amplitudes c_k if the object is normalized. However, 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 scaling, hence, the 3D objects has to be normalized to unit before processing, to ensure invariance to scale. In [Lav11] the 3D objects are normalized to unit before processing, to ensure invariance to scale, then the high computation cost added. To avoid the normalization of 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 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 :

$$c_k = \sum_{i=1}^{m} |v_i^*|^2 H_i^k \tag{8}$$

Each 3D Harris point v_i is associated with a local area A_i for which we compute description c_i (see section 3.2). 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 like [Lav11]) , 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, then the robustness of the descriptor can be conserved.

At the end of this step, we quantify the descriptions into words, as in a classical BoW representation, 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 description is called Harris_MHB.

4 Scalable 3D object retrieval including spatial relationship information

This section presents our second and third contributions : in sections 4.1 and 4.2, we detail the description of the local geometry between 3D words, based on spatial relationships between triplets of words, and the similarity measure associated, while section 4.3 focuses on the description of a dedicated access method allowing scalable 3DOR.

4.1 Spatial relationship description

Traditionally, the BoW representation does not encapsulate any information about the spatial layout of the words. We propose to describe the spatial relationships between 3D words by extending the Δ-TSR [HGBRM10], originally designed for CBIR, to 3D objects. By extension, each 3D object O of the database 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 :

$$\begin{aligned} \Delta\text{-TSR}(O) = \quad & \{S(E_i, E_j, E_k)/E_i, E_j, E_k \in O; \\ & i, j, k \in [1, N_O];\ L_i \geq L_j \geq L_k\} \end{aligned} \tag{9}$$

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 encompass several kinds of information, such as the geometrical relationships between the points. We choose to keep information on the angles of the triangle formed by (E_i,E_j,E_k). It is also possible to consider an orientation 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 :

$$S^o(E_i, E_j, E_k) = (K_1, K_2, K_3, K_4, K_5) \tag{10}$$

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

n_w is the size of the dictionary. K_1 is the unique coding of word's labels from the vocabulary, it can represent almost 600 billions of triangles, easily manageable with *long* type. 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 can 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]).

For notation, if we consider only the relationships based on angles and not orientations, we define the descriptions S^a and Δ-TSR$_{3D}(O)$ such as :

$$\begin{aligned} \Delta\text{-TSR}_{3D}(O) = \quad & \{S^a(E_i, E_j, E_k) = \{K_1, K_2, K_3\}; \\ & i, j, k \in [1, N_O]\} \end{aligned} \tag{11}$$

Δ-TSR may involve triplets of points located far away from each other in the 3D object. Such a representation may be adequate for global 3D object description, but not for local description and then partial retrieval. Similarly as in [HGBRM10], we employ a strategy for triangles selection, by keeping here only the smallest triangles : for each interest point p, we build only the triangles between p and other interest points that are the neighbors of p (see section 3.2).

4.2 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 $S^u(T_Q)$ and $S^u(T_O)$ $(u \in \{a, o\})$ of a query triangle T_Q and a database triangle T_O. The associated similarity measure between two triangle descriptions, called sim, is similar to the one originally proposed in [HGBRM10] :

$$sim = \begin{cases} sim^u(S^u(T_Q), S^u(T_O))\ \forall u \in \{a, o\} \\ \quad \text{if } \Delta(K_1(T_Q), K_1(T_O)) = 1 \\ \quad \text{and } S^u(T_O) \text{ validates the tolerance intervals } \delta_{\{a,o\}} \\ 0 \text{ otherwise} \end{cases} \tag{12}$$

where :

$$\begin{aligned} sim^a(S^u(T_Q), S^u(T_O)) &= f(T_Q, T_O, 2, \delta_a) \\ sim^o(S^u(T_Q), S^u(T_O)) &= \tfrac{1}{2}\left[f(T_Q, T_O, 2, \delta_a) + f(T_Q, T_O, 4, \delta_o)\right] \end{aligned} \tag{13}$$

$\Delta(.,.)$ is the Kronecker's function and $f(.,.,.,.)$ such as :

$$f(T, T', i, \delta) = \begin{cases} 1 \text{ if } \delta = 0 \\ \frac{1}{2}\sum_{t=i}^{i+1}(1 - \frac{|K_t(T) - K_t(T')|}{\delta}) \text{ if } \delta \neq 0 \end{cases} \tag{14}$$

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.

4.3 Associated access method for 3DOR retrieval

Classically, the retrieval process requires the comparison of the description(s) of the query with all the ones stored in the database. To accelerate it in order to reach scalability, we propose to use a structure indexing all triangle descriptions. To find the triangle description similar to a description $S^u(T_Q)$, $u \in \{a, o\}$, the retrieval follows these steps :

1. Search for all the descriptions having a key K_1 equal to $K_1(T_Q)$, T_Q being a triangle in query Q ;
2. Select descriptions $S^a(T_O)$ (resp. $S^o(T_O)$), found in step 1, that validate tolerance interval δ_a (resp. δ_o) ;
3. Compute $sim(S^u(T_Q), S^u(T_O))$ $(\forall u \in \{a, o\})$.

Here, an interesting property is that the triangle descriptions can be ordered, such as : $S^u(T_O) > S^u(T_Q)$ if and only if $\exists i \ / \ K_i(T_O) > K_i(T_Q) \wedge \ \forall j < i \ K_j(T_O) = K_j(T_Q)$. Then the retrieval process becomes the search of the set of descriptions $S^u(T_O)$ in the interval $[BI_i, BI_f]$ where $BI_i = (K_1, K_2 - \delta_a, K_3 - \delta_a, K_4 - \delta_o, K_5 - \delta_o)$ and $BI_f = (K_1, K_2 + \delta_a, K_3 + \delta_a, K_4 + \delta_o, K_5 + \delta_o)$ when $u = o$; $\delta_{\{a,o\}}$ are the tolerance thresholds. Consequently, it is optimal to use the classical searching structure B-tree, by considering composite keys, to index all the composite descriptions $S^u(T)$. The searching time is $O(N_{AT} log_b N_T)$ where N_{AT} is the average number of triangles in the query, N_T is the number of triangles in the database and b is the B-tree order. For large scale 3DOR retrieval, this disk-based indexing structure is efficient, as it will be experimented in section 5. Our approach of description rests upon the BoW technique, however, by embedding spatial information between triplets of entities, the matching problem is not a classical comparison of two histograms, therefore, the use of inverted files, as indexing structure usually associated with BoW, would be not efficient here. The complexity of the inverted files would be largely superior to the one of the composite B-tree : $O({n_w}^3 {n_u}^4)$ where n_w is the size of the BoW dictionary, and n_u the average number of quantized values for components $K_2, .., K_5$.

5 Experiments and evaluation

Here, we evaluate the relevance of the local description Harris_MHB proposed in section 3.3 and then of Δ-TSR, the complete description including spatial relationships presented in section 4. In section 5.1, we evaluate the influence of the main parameters in the 3D point description Harris_MHB, before comparing it with state of the art in section 5.2. Sections 5.3, 5.4 and 5.5 consider the spatial relationship description and compare Δ-TSR with state of the art on different benchmarks, while 5.6 discusses the volume of the features generated and the scalability of the whole proposal.

We consider the five following public 3D benchmarks, having different contents and sizes :

— The collection SHREC 2007 [1], called "Shape Retrieval Contest of 3D Face Models" : it consists of 1500 different instances of 3D face models (see examples in the first row of Fig.2).

1. http ://ensor.labs.cs.uu.nl/shrec/shrec2007

— The Gun Database [2] : it is composed of 150 synthetic scenes, captured with a (perspective) virtual camera, and each scene contains 3 to 5 objects (see examples in the second row of Fig.2). The model set is composed of 20 different objects, taken from different sources and then processed in order to obtain comparably smooth surfaces of almost uniform 100 - 350k triangles.
— The collection SHREC 2013 [3], called "Large-Scale Partial Shape Retrieval using Simulated Range Images" [SMB+13] : it consists of 20 object classes with 18 models per class and 7200 queries (see examples in the third row of Fig.2).
— The McGill Database [4] : it contains more than 450 objects divided into ten classes (Ant, Crabs, Hands, Humans, Octopuses, Pliers, Snakes, Spectacles, Spiders and Teddy) ; the intra-class variations concern non-rigid transforms of the models (see examples in the fourth row of Fig.2).
— The NTU Database [5] : it contains more than 10K publicly available 3D models collected from WWW pages. This database is not a ground truth benchmark ; we use it only for the evaluation at large scale of section 5.6.

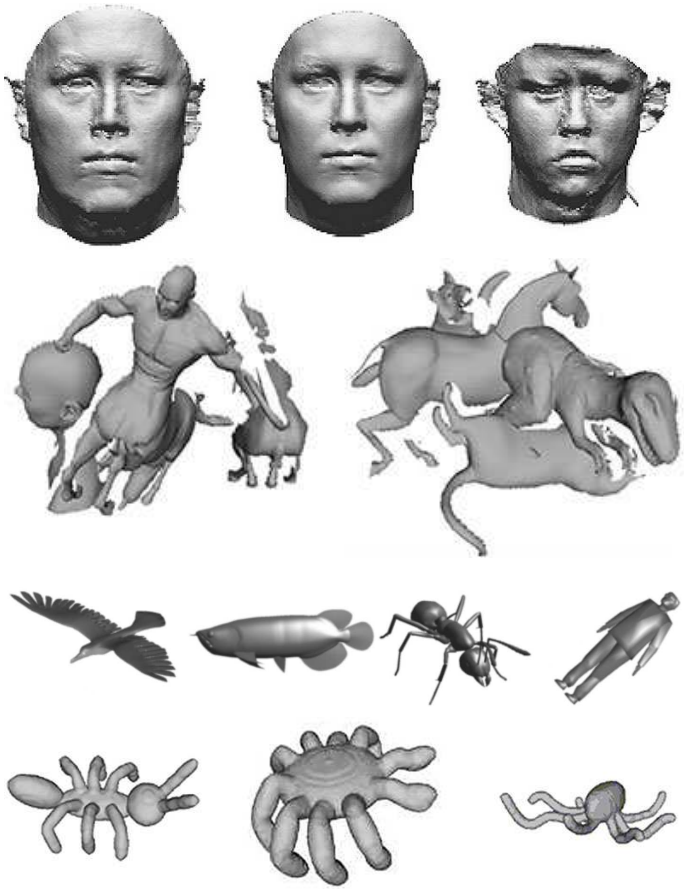

FIGURE 2 – Some examples from SHREC 2007, Gun, SHREC 2013 and McGill benchmarks (one benchmark per row).

2. http ://www.dsi.unive.it/∼rodola/data.html
3. http ://dataset.dcc.uchile.cl
4. http ://www.cim.mcgill.ca/∼shape/benchMark
5. http ://3d.csie.ntu.edu.tw

5.1 Influence of the parameters

Harris_MHB is based on two principal parameters : the word dictionary size n_w and the number of coefficients n_c of the spectral descriptor, noting that the neighborhood of each local feature is selected as explained in section 3.2. According to [Lav11], when the number of spectral coefficients n_c increases, the discriminative power of the Fourier descriptor can increase, hence the performances are better. However, when n_c is too high, it can go down because including too high frequencies in the spectral descriptor that remove a part of its robustness. Similarly, n_w can impact the description exactly in the same way.

Firstly, to have an overview of the impact of these two parameters, we select n_w in $\{200, 2000, 10000\}$ and n_c in $\{20, 40, 60\}$ for Harris_MHB and also for our implementation of approach [Lav11], which is recent and similar to Harris_MHB. Let recall the main differences between the two methods : Harris_MHB uses 3D Harris detector to find features and the projection of surfaces as descriptor, while [Lav11] detects features via Voronoi distribution and uses the projection of points as descriptor. The number of points detected per object with Harris_MHB varies in 3K-6K while [Lav11] fixes them to 200 according to the best parameters found by the authors. Note that, at this point, we do not consider the spatial relationships information both in the two approaches.

Table 1 shows the MAP (Mean Average Precision) obtained on SHREC 2007 and Gun benchmarks, which have very different contents. On the one hand, $n_c = 40$ is clearly the best parameter for the two approaches on the two benchmarks. On the other hand, Harris_MHB is improved with $n_w = 2000$, which is the best configuration for the two approaches, except for [Lav11] on SHREC.

TABLE 1 – MAP for Harris_MHB and [Lav11] for different parameter settings, on SHREC 2007 and Gun benchmarks.

SHREC 2007 Benchmark		$n_w = 200$	2000	10000
$n_c = 20$	Harris_MHB	0.583	0.655	0.649
	[Lav11]	0.572	0.536	0.511
$n_c = 40$	Harris_MHB	0.707	**0.801**	0.790
	[Lav11]	**0.682**	0.670	0.671
$n_c = 60$	Harris_MHB	0.635	0.736	0.692
	[Lav11]	0.629	0.617	0.594
Gun Benchmark		$n_w = 200$	2000	10000
$n_c = 20$	Harris_MHB	0.281	0.365	0.349
	[Lav11]	0.172	0.216	0.201
$n_c = 40$	Harris_MHB	0.394	**0.461**	0.412
	[Lav11]	0.262	**0.298**	0.273
$n_c = 60$	Harris_MHB	0.401	0.456	0.442
	[Lav11]	0.229	0.277	0.267

These preliminary results show that Harris_MHB obtain the best results facing [Lav11] for these two datasets. It seems that the interest point detection based on Voronoi distribution used in [Lav11] provides a lower discrimination between features. Certainly, on the SHREC 2007 benchmark, there are many features which have similar descriptions on flat areas like face cheeks, foreheads. Moreover, to our point of view, the number of features in [Lav11] is quite low (200) to be very discriminative. We

increased it to 1000 to see the performance of [Lav11] with more features, noting that, with $n_c = 40$, each patch has at least 40 vertices, hence the number of features cannot exceed 1500 while the number of meshes varies in 60K-80K. On SHREC 2007, approach [Lav11] obtained MAP = $\{0.651; 0.678; 0.669\}$ respectively for $n_w = \{200; 2000; 10K\}$. The best MAP is not reached, because with smaller patches, we think that there are fewer choices in coefficients n_c of the spectral description, making the feature description probably less rich and robust.

5.2 Comparison of Harris_MHB with different descriptions

Now, we examine the performance of several descriptions on the 3D database SHREC 2007 : Harris_MHB, again the approach of Lavoué [Lav11] (in its version without spatial relationships), the one of Toldo et al. [TCF03] (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 the spatial relationship information in these approaches. The size of the word dictionary for all BoW approaches is $n_w = 200$. Fig. 3 presents the Precision/Recall curves obtained. The two methods of [Lav11] and [TCF03] present quite comparable performances, however [Lav11] is slightly better, MAP of [TCF03] being 0.615 and MAP of [Lav11] being 0.682. 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, except for very large recalls where [Lav11] becomes better. Its MAP is 0.707.

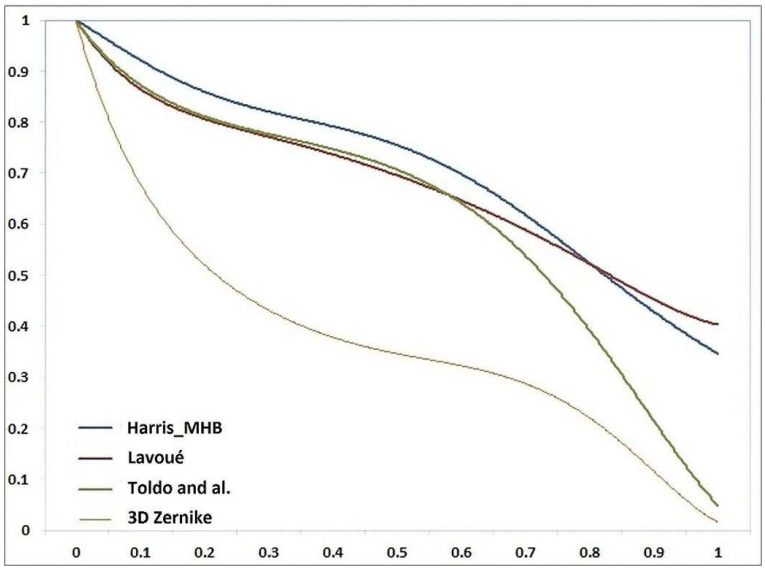

FIGURE 3 – Precision/Recall curves for different descriptions on SHREC 2007.

5.3 Including spatial relationships

On SHREC 2007 benchmark, where global shapes are very similar, the embedding of spatial relationship in BoW approaches improves slightly the quality of retrieval. Indeed, Fig. 4 presents the

Precision/Recall curves obtained from Harris_MHB+Δ-TSR that includes spatial relationships with the two versions Δ-TSR$_{3D}$ and Δ-TSR$_{5D}$ of Δ-TSR and from the full approach of Lavoué [Lav11] that includes spatial relationships based on co-occurrences (called "Lavoué with CR"). From this experiment, we see that the associated MAP for "Lavoué with CR" is 0.691 vs. 0.682 without the use of spatial relationships, and the ones for Δ-TSR$_{5D}$ and Δ-TSR$_{3D}$ are 0.721 and 0.715 respectively, while it is 0.707 for Harris_MHB. These results confirm that exploiting spatial relationships improves the object description, as well as the use of orientations in Δ-TSR$_{5D}$.

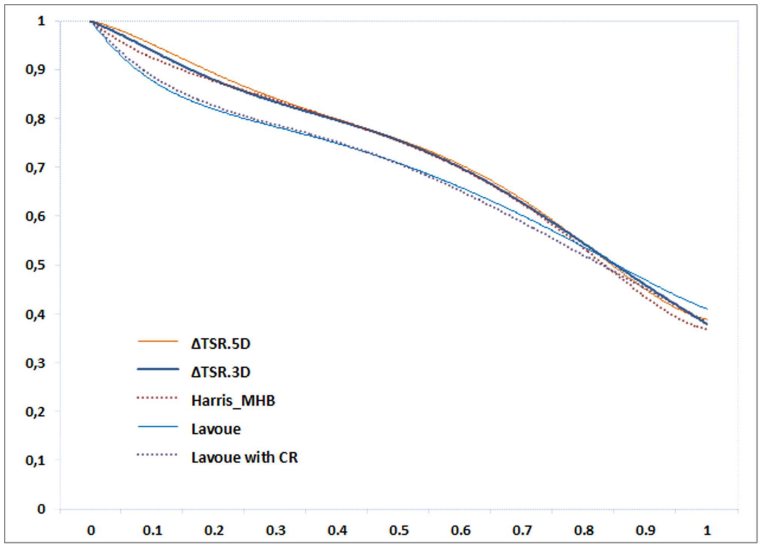

FIGURE 4 – Precision/Recall curves for different descriptions with spatial relationship on SHREC 2007.

The Gun dataset is a difficult benchmark for 3DOR, particularly relevant for partial retrieval : every 3D object is partly present in the scene. This is why we choose it to examine the performance of our approach facing [Lav11] which obtained the second best results in experiment of section 5.2 and which is also dedicated to partial shape retrieval. All the experiments are done with $n_w = 2000$ and $n_c = 40$ which are the best configurations both for the two approaches on this dataset.

Here, we compare the two versions of Δ-TSR with the full version of [Lav11], and also with Reeb Pattern approach of [TVD09] that is dedicated to partial 3D shape retrieval. Fig. 5 shows the Precision/Recall curves obtained for all the versions of the approaches, on the Gun benchmark.

Because method [Lav11] requires the a priori normalization of 3D objects and the point detection rests upon Voronoi distribution on 3D surface, this approach is clearly penalized on this particular 3D dataset. Normalized objects may not have the same size when included in a normalized scene containing other objects. Furthermore, the choice of a Voronoi distribution is clearly questionable when the 3D objects are present partially with other ones in the scene. Another remark is that here, the co-occurrence relationship version improves slightly the results ; associated MAP is 0.315 vs. 0.298 without. Concerning the Reeb Pattern approach of [TVD09], because it is a topology-based approach, the decomposition and matching processes are quite sensitive to topology variations within a same class of objects. Consequently, its associated MAP is only 0.275. With Δ-TSR, not surprisingly, the information on triangular spatial relationships and on orientation of 3D points demonstrates its power of description facing Harris_MHB, its MAP is 0.684 with Δ-TSR$_{5D}$ (and 0.553 with Δ-TSR$_{3D}$). With

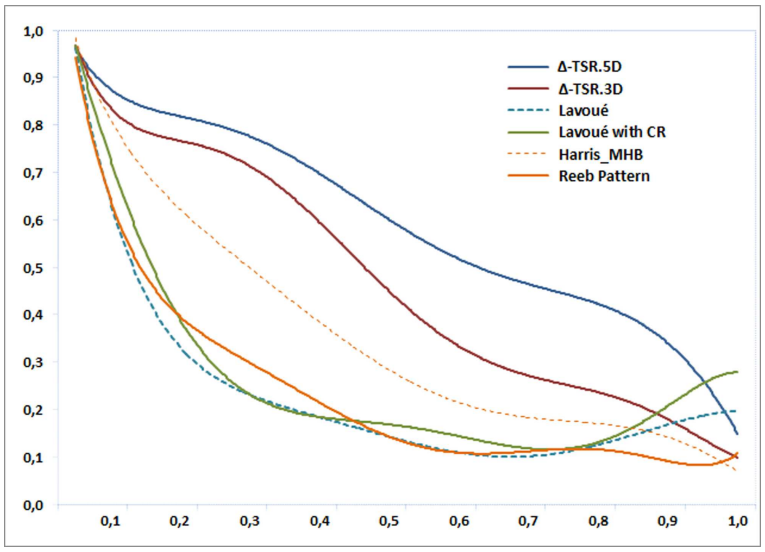

FIGURE 5 – Precision/Recall curves for two versions of Δ-TSR, two versions of Lavoué approach [Lav11] and Reeb Pattern approach [TVD09], on Gun benchmark.

this benchmark, the results obtained in this experiment prove the efficiency of Δ-TSR for the partial 3D shape similarity problem. In the following experiments, we use the version Δ-TSR$_{5D}$ of Δ-TSR, which provides the best results.

5.4 Comparative evaluation on SHREC 2013

In this section, we compare Harris_MHB+Δ-TSR facing two recent methods submitted in the SHREC 2013 track :

— Range Scan-Based 3D Model Retrieval by Incorporating 2D-3D Alignment ([LLJ12, BBF12]). For abbreviation, this method is called Li-Lu-Johan.

— Partial Shape Retrieval with Spin Images and Signature Quadratic Form Distance ([SB12]). For abbreviation, this method is called Sipiran-Bustos.

Fig. 6 gives the Precision/Recall curves obtained on this benchmark with these approaches, and Tab.2 summarizes the results with other performance metrics used in the SHREC 2013 track : Nearest Neighbor (NN), First Tier (FT) and Second Tier (ST) [SMB+13].

From the Precision/Recall plots, we observe the superiority of our method. This can be also evidenced by the results of performance measures of Table 2.

TABLE 2 – Performance measures on SHREC 2013 benchmark.

	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**

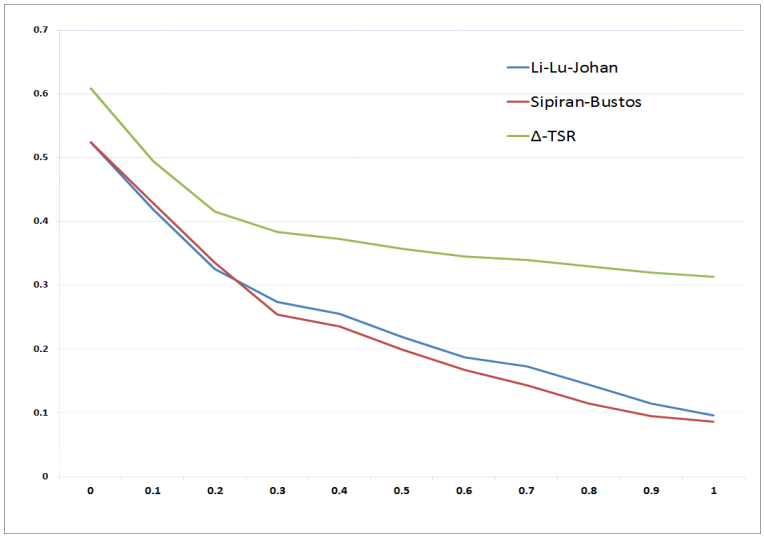

FIGURE 6 – Precision/Recall curves for different approaches on SHREC 2013 benchmark.

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 the 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 key points 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 the 3D object. The distribution of interest points is almost homogeneous. Moreover, with Δ-TSR, the information on triangular spatial relationships and on orientation of 3D points reinforces the object description for partial query image. It demonstrates its power of description facing other approaches ; its MAP is 0.3434.

To have more insight in the performance, a class-by-class evaluation of our approach is shown in Tab.3. The detail of class-by-class evaluation for the two other approaches can be found 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.

5.5 Comparative evaluation on McGill

We continue to evaluate Δ-TSR by comparing it to two other methods, which were publicly evaluated on the McGill benchmark [SMKF04] : the graph-based approach from [APP+09] and the hybrid 2D/3D approach from [PPT08]. Table 4 presents this comparison, including also [Lav11] in its complete version. The measure used is the First Tier (FT), obtained with the tools from [SMKF04]. The first remark is that the graph-based algorithm [APP+09] provides the best results for four classes among ten ; this result is logical since this benchmark only considers skeletal articulation deformations without topology changes, hence it is particularly suited for graph-based representations. However, on some 3D objects having a complex surface, we notice that Δ-TSR, whereas considering almost no structural information, provides quite good results, almost always better than [PPT08] and [Lav11]. Globally, Δ-TSR obtains the best results for five classes among ten.

TABLE 3 – Performance measures of Δ-TSR by class on SHREC 2013 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

TABLE 4 – Retrieval statistics on four approaches for the McGill benchmark.

Method	Class	FT	Class	FT
[Lav11]	Ant	57.7	Crabs	60.7
[APP$^+$09]		74.1		54.9
[PPT08]		55.7		**73.6**
Δ-TSR		**76.5**		61.3
[Lav11]	Hands	51.5	Humans	70.8
[APP$^+$09]		**83.9**		**93.5**
[PPT08]		43.4		47.0
Δ-TSR		66.8		85.3
[Lav11]	Octopuses	26.8	Pliers	88.7
[APP$^+$09]		58.8		**100**
[PPT08]		29.5		71.6
Δ-TSR		**79.7**		85.5
[Lav11]	Snakes	24.7	Spectacles	86.7
[APP$^+$09]		**43.2**		70.2
[PPT08]		23.7		53.5
Δ-TSR		31.2		**90.5**
[Lav11]	Spiders	77.7	Teddy	96.1
[APP$^+$09]		87.2		45.3
[PPT08]		71.5		90.3
Δ-TSR		**92.7**		**98.4**

5.6 About volumes of features and scalability

In this section, we evaluate the performance of Δ-TSR in terms of volume of the descriptions produced and of retrieval times. We measure the CPU time as well as IO time necessary to access the data through our disk-based index, considering also, to be realistic, the steps of point extraction and of description for the query. The measures are averaged over all the considered models, used as queries. The experiments are done on an Intel i7 hex core pc going with 32GB of RAM, with the best configuration of Δ-TSR. On benchmarks SHREC 2007, Gun, McGill, NTU and their concatenation (more than 2000 models), table 5 shows statistics on the amount of descriptions produced and manipulated for each benchmark and for each query in average (two first rows). On the three last rows, we present averaged CPU/IO times retrieval : "Desc. CPU time" is the part of "Total CPU time" associated with query point extraction and description only, which is a very time consuming step.

First, we observe that retrieval time is largely dominated by the step of query description, which exceeds 3 seconds for the most complex dataset (Gun), while the rest of the processings are one order of magnitude lower. Secondly, thanks to the sub-linear behavior of the B-tree index with composite keys, we see that the database size and the amount of features produced do not increase the retrieval time. It is mainly the amount of descriptions in the query which has an impact on this measure. We think that these results, obtained on datasets of different sizes and complexities, show that Δ-TSR remains effective on larger datasets, and then is scalable.

TABLE 5 – Volumes of description and average CPU/IO times of retrieval for Δ-TSR, on five datasets.

Benchmark	SHREC 2007	Gun	McGill	NTU	All
DB description ($\times 10^6$)	37	15	6	140	198
Query description (K)	15-30	50-90	8-20	1-40	1-90
Total CPU time (s)	2.189	3.533	1.655	2.417	2.841
Description CPU time	1.754	3.013	1.295	1.595	2.249
IO time (s)	0.151	0.323	0.136	0.968	1.408

6 Conclusion

In this paper, we have proposed an efficient approach for 3D object retrieval, dedicated to partial shape retrieval and large datasets. A quantization of local features, as in BoW representations, is employed, based on the extraction of 3D Harris points and on a local description involving local Fourier descriptors both fast to compute and discriminative. By adding Δ-TSR, a triangular spatial information between triplets of words, the robustness of this representation is reinforced. Indeed, the main steps of the involved pipeline have been carefully designed by focusing on both the effectiveness and efficiency of retrieval by example. The experimental evaluations of the proposal were performed on different public 3D benchmarks (SHREC 2007, Gun, SHREC 2013, NTU and McGill benchmarks) involving different contents and degrees of complexity, facing several state-of-the-art techniques ([Lav11],[TCF03], [NK03], [APP+09], [PPT08], [TVD09], [LLJ12, BBF12], [SB12]). They have provided encouraging results in terms of quality of retrieval. Δ-TSR proves its performance for both classic object retrieval and partial matching in complex scenes. Among several variants, Δ-TSR$_{5D}$ was evaluated as the most relevant on the tested datasets. In addition, Δ-TSR was designed to reduce retrieval time notably, by exploiting

a composite B-tree as disk-based index structure, thus demonstrating its effectiveness and robustness facing scalability. 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] (see section 2) 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. As a final illustration of this proposal, Fig. 7 shows some examples of retrieval with Δ-TSR.

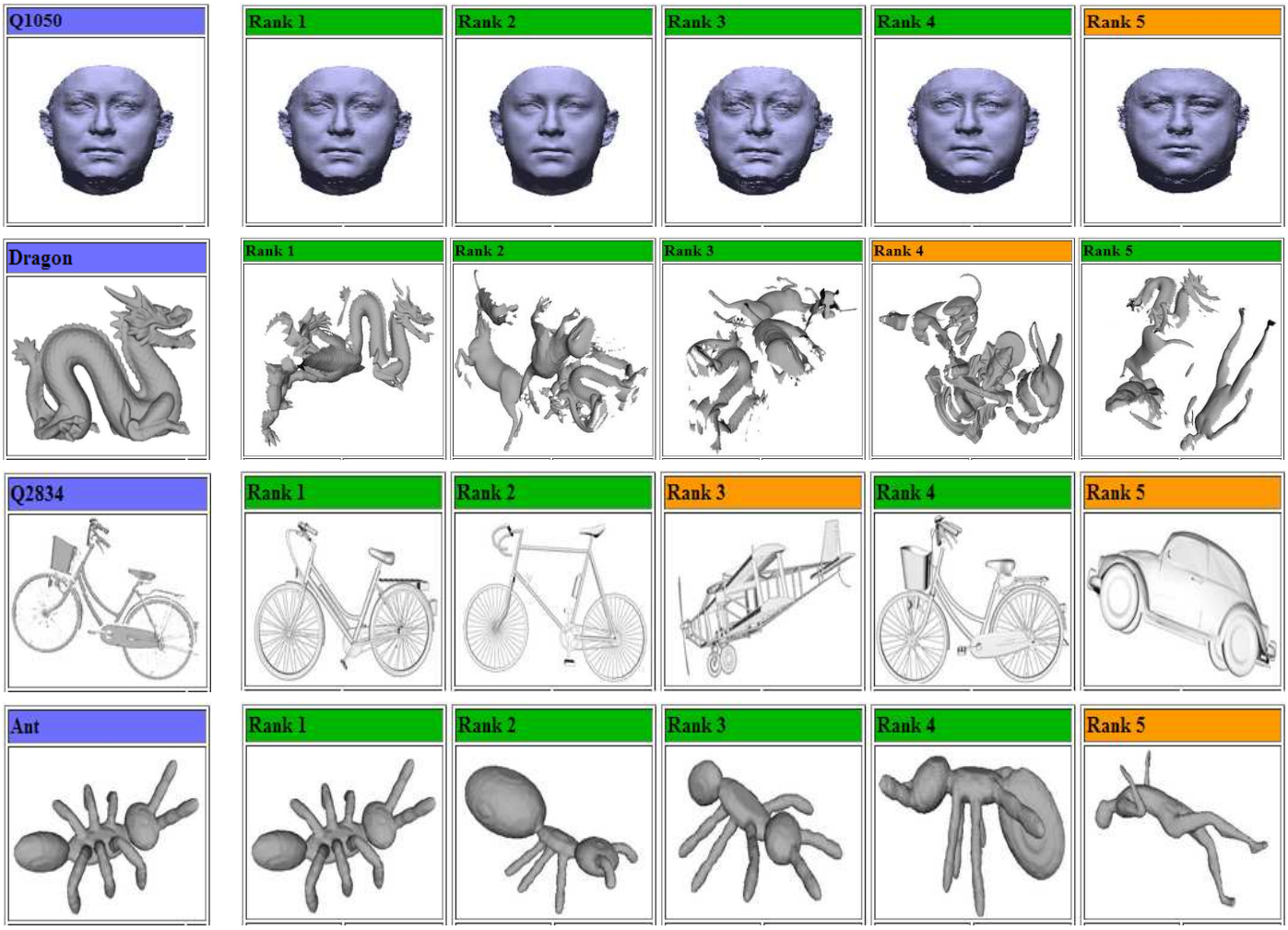

FIGURE 7 – Examples of retrieval with Δ-TSR, on SHREC 2007, Gun, SHREC 2013 and McGill benchmarks. Queries are in the first column, followed by the 5 closest responses, sorted by decreasing order of similarity. The green color indicates that the response is in the class of the query.

References

[APP+09] A. Agathos, I. Pratikakis, P. Papadakis, S. Perantonis, P. Azariadis, and N. Sapidis. Retrieval of 3d articulated objects using a graph-based representation. *Eurographics Workshop on 3D Object Retrieval*, 2009. 2, 3, 16, 17, 18

[ASWL11] Y. An, Ch. Shao, X. Wang, and Z. Li. Estimating principal curvatures and principal directions from discrete surfaces using discrete curve model. *Journal of Information & Computational Science*, page 296–311, 2011. 9

[BBF12] M. Bronstein, M. M. Bronstein, and A. M. Ferreira. Sketch-based 3d shape retrieval. *Eurographics workshop on 3D Object Retrieval*, 2012. 15, 18

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

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

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

[DCG12] H. Dutagaci, C. Chun, and A. Godil. Evaluation of 3d interest point detection techniques via human-generated ground truth. *The Visual Computer*, pages 901–917, 2012. 4

[DJLW08] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image Retrieval : Ideas, Influences, and Trends of the New Age. *ACM Computing Surveys*, pages 1–60, 2008. 3

[FMK+03] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, and D. Jacobs. A search engine for 3d models. *ACM Transactions on Graphics*, page 83–105, 2003. 1

[Glo09] P. Glomb. Detection of interest points on 3d data : Extending the harris operator. *Computer Recognition Systems*, page 103–111, 2009. 4

[HGBRM10] N.V Hoàng, V. Gouet-Brunet, M. Rukoz, and M. Manouvrier. Embedding spatial information into image content description for scene retrieval. *Pattern Recognition*, pages 3013–3024, 2010. 8, 9, 10

[HS88] C. Harris and M. Stephens. A combined corner and edge detection. *Proc. of The Fourth Alvey Vision Conference*, page 147–151, 1988. 4

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

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

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

[Lav11] G. Lavoue. Bag of words and local spectral descriptor for 3d partial shape retrieval. *Eurographics Workshop on 3D Object Retrieval*, 2011. 2, 3, 7, 8, 12, 13, 14, 15, 16, 17, 18

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

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

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

[LLJ12] B. Li, Y. Lu, and H. Johan. Sketch-based 3d model retrieval by incorporating 2d-3d alignment. *Multimedia Tools and Applications*, 2012. 15, 18

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

[NK03] M. Novotni and R. Klein. 3d zernike descriptors for content based shape retrieval. *ACM symposium on Solid modelling and applications*, pages 216 – 225, 2003. 2, 13, 18

[NS06] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. *Conference on Computer Vision and Pattern Recognition*, page 2161–2168, 2006. 4, 8

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

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

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

[PPT08] P. Papadakis, I. Pratikakis, and T. Theoharis. 3d object retrieval using an efficient and compact hybrid shape descriptor. *Eurographics Workshop on 3D Object Retrieval*, 2008. 3, 16, 17, 18

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

[RABT13] E. Rodolà, A. Albarelli, F. Bergamasco, and A. Torsello. A scale independent selection process for 3d object recognition in cluttered scenes. *International Journal of Computer Vision*, 2013. 2, 3, 19

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

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

[SB11] I. Sipiran ; and B. Bustos. Harris 3d : a robust extension of the harris operator for interest point detection on 3d meshes. *The Visual Computer*, pages pp 963–976, 2011. 3, 4

[SB12] I. Sipiran and B. Bustos. Key-component detection on 3d meshes using local features. *Eurographics Workshop on 3D Object Retrieval*, 2012. 15, 18

[SMB+13] I. Sipiran, R. Meruane, B. Bustos, T. Schreck, B. Li, Y. Lu, and H. Johan. Shrec'13 track : Large-scale partial shape retrieval using simulated range images. *Eurographics Workshop on 3D Object Retrieval*, 2013. 11, 15, 16

[SMKF04] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser. The princeton shape benchmark. *Shape Modeling International*, 2004. 16

[SOG09] J. Sun, M. Ovsjanikov, and L. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *Computer Graphics Forum*, 2009. 3

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

[SWS+00] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, page 1349–1380, 2000. 1

[TCF03] R. Toldo, U. Castellani, and A. Fusiello. Visual vocabulary signature for 3d object retrieval and partial matching. *Eurographics Workshop on 3D Object Retrieval*, 2003. 2, 3, 13, 18

[TS10] F. Tombari and S. Salti. Unique signatures of histograms for local surface description. *European Conference on Computer Vision*, 2010. 3

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

[TVD09] J. Tierny, J. Vandeborre, and M. Daoudi. Partial 3d shape retrieval by reeb pattern unfolding. *Computer Graphics Forum*, 2009. 14, 15, 18

[VL08] B. Vallet and B. Levy. Spectral geometry processing with manifold harmonics. *Eurographics Workshop on 3D Object Retrieval*, 2008. 6, 7

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