Int. J. Open Problems Compt. Math., Vol. , No. , December, 2013
ISSN 1998-6262; Copyright © ICSRS Publication, 2013
www.i-csrs.org

3D models retrieval using Earth Mover's Distance

O. Ait Zemzami, H. Aksasse, M. Ouanan, and B. Aksasse

Université Moulay Ismaïl, Faculté des Sciences et Techniques
Département d'informatique, Equipe ASIA
BP 509 Boutalamine 52000 Errachidia, Maroc
omarzemzami@yahoo.fr
haksasse@yahoo.fr
ouanan_mohammed@yahoo.fr
baksasse@yahoo.com

(*Communicated by Lahcen Oukhtite*)

Abstract

Thanks to the recent and advanced development in three-dimensional (3D) computer imaging technology, 3D objects begin to compete with traditional media (2D images, sounds, videos). These objects are usually represented by triangular meshes and are used in various fields. Thus, 3D object databases with size increasingly important become available. To effectively manage and manipulate this database, an indexing and retrieval system for 3D models is required. In this work, we present a method for 3D models retrieval based on the similarity measure between 2D slices. The main idea is to represent the 3D triangular mesh model, as a series of slices along a direction so that the similarity between 3D models is transformed into a measure of similarity between the 2D slices. In this process, three problems are involved: preprocessing of 3D models, the method of getting slices and the similarity measure between slices.

Keywords: *3D Object, Object retrieval, 2D slices, Earth Mover's Distance (EMD), 3D triangular mesh model.*

1 Introduction

Currently 3D objects play a very important role in many computer applications. They are thus increasingly present in many areas, that they are either playful field (games, multimedia) or scientific (medical, industrial, cultural heritage, etc...).

Browsing in these databases to find information remains difficult. This because the majority of 3D models contain geometrical information only in the form of facets, and they do not contain any semantic information to facilitate indexing and retrieval. This has led many researchers to find efficient methods for the description of these three-dimensional objects by their content with low-level descriptors such as the shape. These descriptors must then be invariant under to certain geometric transformations such as translation, rotation and scaling. At this level, the problem of similarity between the indexes of these 3D objects leads to define one or more metrics that should express a distance between two visual objects.

3D model retrieval by similarity consists in extracting, from a database, the most similar models for a given query. Many methods and approaches have been developed in this way; among the main approaches are the statistical methods, structural, by processed or visual recognition.

Among the statistical methods, we quote that of Assfalg et al. [1,2] who adopted a method which consists in comparing the maps of curvature as classical 2D images. This method is based on the idea that the shape of a 3D object can be described by a map of the curvatures of the surface. It is worth mentioning that the results obtained by this approach are better than the ones obtained by a method which uses histograms of curvature calculated directly on 3D objects.

In another hand, the structural methods, we can cite the ones of Attali et al.[3] and Ogniewicz with Ilg [7] who proposed approaches that calculate the skeleton from the Voronoi diagram (VD) of a set of points which is a discrete sampling of the contour of the object. The Voronoi diagrams and skeletons established via these approaches converge to the ideal case as far as we increase the sampling density. So to implement this idea in practice, one need to find a discrete sampling that guaranteed a better approximation to the shape of the object [3,7].

The third category consists method by processed. They try to calculate the 3D Zernike descriptors based on 3D Zernike moments is developed by Canterakis [4] and used by Klein and Novotni [6] on discrete and normalized objects.

Finally, characterizing a 3D object by ten sets containing ten orthographic views taken in the first faces of a dodecahedron centered on the object is the method proposed by Chen et al. [5] entitled the Light Field Descriptor (LFD).

In this work, we deal with this problem using a new idea. The general idea of our approach is based on the fact that two similar 3D models have similar cuts, and vice versa. Thus, we propose to represent a 3D model by a series of 2D slices in a certain direction so that the problem of similarity between the 3D models is converted into a measure of similarity between the series of 2D slices. This is guaranteed by using Earth Mover's Distance (EMD), which is a distance measure between two vectors of any dimension. It is based on the metaphor of the minimum work that should be done by a roadman to transform a pile of sand (represented by a vector) into a hole.

The outline of the paper is as follows. The next section contains a description of the proposed approach. Experimental results are presented and discussed in Section 3. The conclusion is presented in Section 4. Finally, Section 5 concludes the paper.

2 Proposed Method

A 3D model can be obtained in any position Fig. 1 (a). However, most of the indexing methods are not invariant to changes in scale, translation and rotation. To overcome this problem, a normalization preprocessing step is needed in order to represent a 3D model in a canonical coordinate system. The aim of the normalization step is to guarantee that the same feature representation can be properly extracted from the same 3D object with any different scale, position and orientation.

The preprocessing step consists in centering, scaling and alignment.

2.1 Centering

Centering can provide invariance to translations. It consists in setting the center of the 3D mesh at the location (0,0,0). The components of the center of mass of a 3D model are given by the formula:

$$G\left(\frac{\sum_{i=0}^{N} x_i}{N};\frac{\sum_{i=0}^{N} y_i}{N};\frac{\sum_{i=0}^{N} z_i}{N}\right)$$

where N is the number of points in the mesh model.
Thus, for a given 3D model P, we obtain a centered one P_c as follows:
$P_c = P - G = \{q / q = p - G, p \in P\}$ (see Fig. 1 (b)).

2.2 Scaling

Scaling process allows us to provide an indexing method invariance to scale changes. It consists in defining for each object a scaling factor given by the formula:

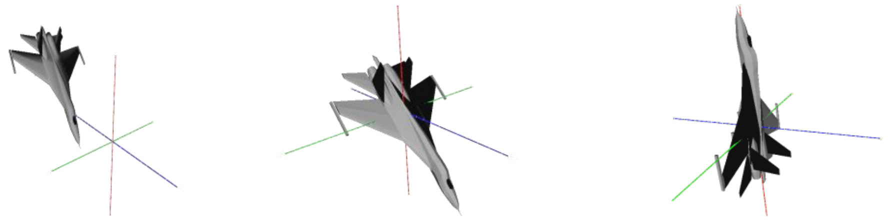
$$s = \max_{1 \le i \le N} \|G - p_i\|$$

where $(p_i)_{1 \le i \le N}$ is the set of the vertex of the object and G its center of mass.

The scale invariance is then obtained by multiplying all vertex $(p_i)_{1 \le i \le N}$ by $\frac{1}{s}$.

2.3 Alignment

The alignment can make the indexing method robust to rotation. The principle of alignment is to determinate the principal axes of these objects and to align them with those of the system axes. Principal component analysis (PCA) is the most used technique to align a 3D object. However, it is sensitive to changes in mesh resolution. For our work we use the Continuous Principal Component Analysis (CPCA) [11] (see Fig. 1 (c)), which is more stable and more accurate.



(a) (b) (c)

Fig. 1 Shape normalization using CPCA

2.4 Feature vector extraction

After centering, scaling and aligning the 3D model, it is subjected to a series of 100 cuts. The method of obtaining slices consists at first in obtaining a set of points which are the intersection of the plane perpendicular to the X axis and the 3D triangular mesh model. Then, we connect these points according to their intrinsic neighborhood. This allows describing slices using polygons.

Construction of the descriptor vector is based on determining both polygons air and the Steiner points. The Steiner point is the centroid of the polygon vertices weighted by the angle vertex.

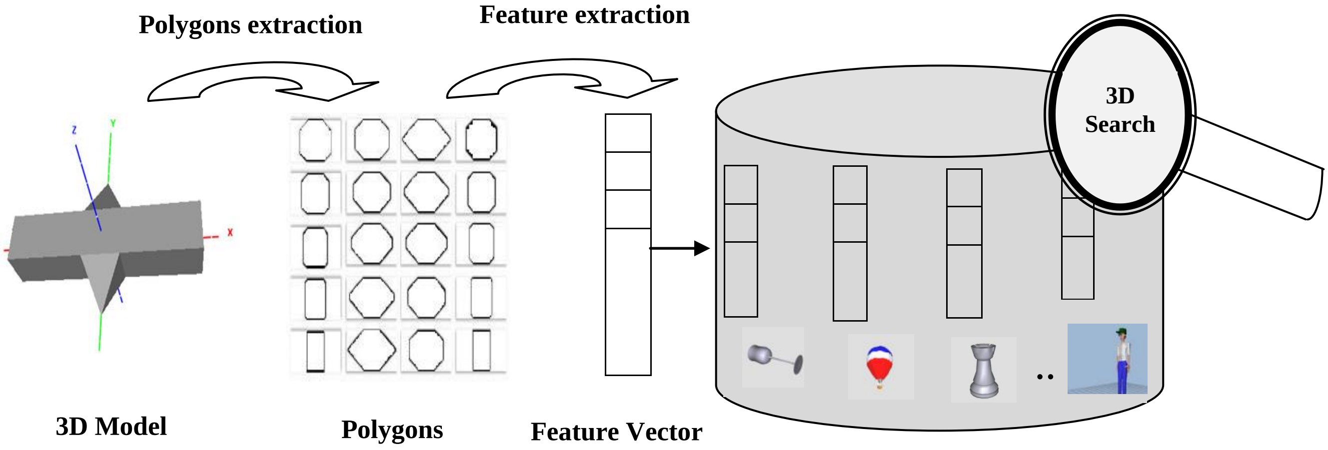


Fig. 2 The process of 3D indexing

2.5 Earth Mover's Distance (EMD)

We use the Earth Mover's Distance in order to express the similarity between two signatures from two different distributions efficiently and to retrieve the most similar models to the query model.

Earth Mover's Distance is based on the famous transport problem, which is studied at first by the military engineer and mathematician Gaspard Monge in the late eighteenth century [12]. His name was introduced by Solfti [10]. It was Peleg et al. [8] who have introduced it initially as a measure of distance for monochrome images and Rubner et al. [9] in a second time to the color and textured images. The EMD is based on the metaphor of the minimum work required for a roadman to provide to transform a pile of dirt in a hole as shown in Fig. 3. This can be formulated as the following linear programming problem:

Let $P = \{(p_1, \omega_{p_1}),...,(p_m, \omega_{p_m})\}$ be the first signature with m clusters, where p_i is the cluster representative and ω_{p_i} is the weight of the cluster.
Let $Q = \{(q_1, \omega_{q_1}),...,(q_n, \omega_{q_n})\}$ be the second signature with n clusters. Let $C = [c_{ij}]$ be the ground-distance matrix, where c_{ij} is the ground distance between cluster p_i and q_j.
We assume that the signature P denotes "piles of dirt," the signature Q denotes "holes," and the ground distance c_{ij} denotes the transportation cost between each pile of dirt and hole. Our aim is to find a flow $F = [f_{ij}]$ such that f_{ij} lies between p_i and q_j, and it minimizes the following overall cost: $Work(P,Q) = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij} f_{ij}$
subject the following constraints:

1) $f_{ij} \geq 0 \qquad 1 \leq i \leq m \quad 1 \leq j \leq n$

2) $\sum_{j=1}^{n} f_{ij} \leq \omega_{p_i} \qquad 1 \leq i \leq m$

3) $\sum_{i=1}^{m} f_{ij} \leq \omega_{q_j} \qquad 1 \leq j \leq n$

4) $\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = \min(\sum_{i=1}^{m} \omega_{p_i}, \sum_{j=1}^{n} \omega_{q_j})$

Once the transportation problem is solved and the optimal flow F is found, the EMD is defined as the resulting work normalized by the total flow:

$$\mathrm{EMD}(P,Q)=\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n}f_{ij}}$$

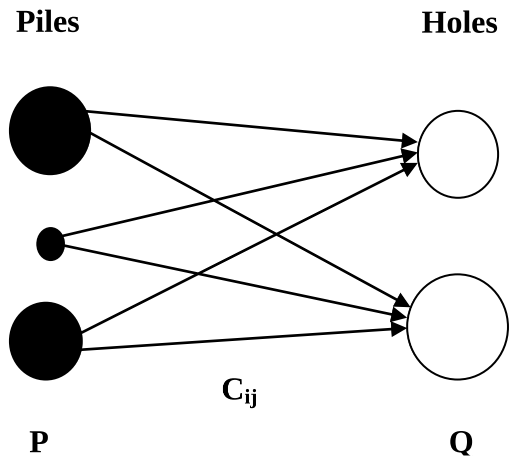


Fig. 3 Principle of Earth Mover's Distance

3 Experimental Results

In order to test our method we used 366 3D models classified into 12 classes (Fig.4). These models are extracted from the database NTU (National Taiwan University) [13] publicly available and well known. This database, which contains 10,911 3D models, is provided for research purposes, among them: 3D model retrieval, matching, recognition, classification, clustering and analysis...

All 3D models are converted into Wavefront file format (.obj) in the database. Thumbnail images of each 3D model are also included in the database.

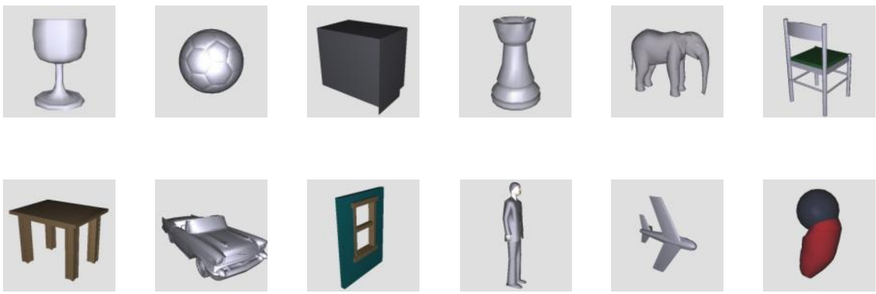


Fig. 4 Thumbnail images of the representatives classes of the test database

For a given query, the top 10 3D models selected by the method as having the most similar shape objects are displayed. We have developed a graphic user interface that allows you to make a subjective quality assessment (see Fig. 5 and Fig. 6). The query model is located in the upper left and the most similar objects are ordered from top to bottom and left to right.

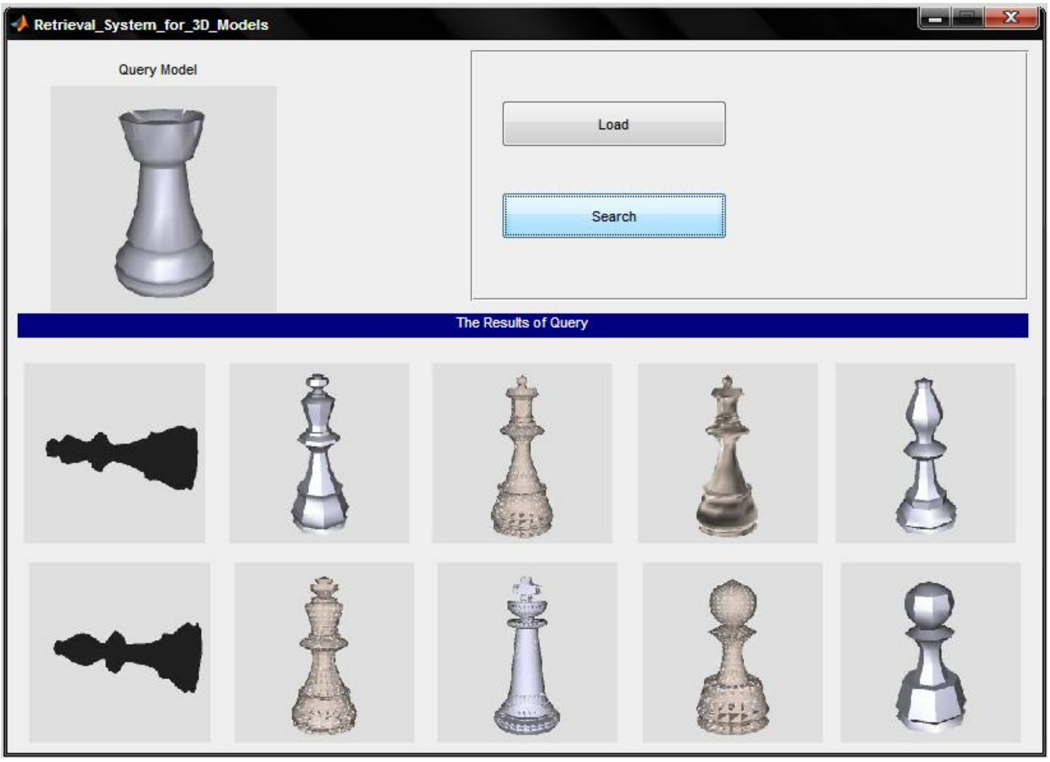


Fig. 5 Retrieval results by our approach. (Example 1)

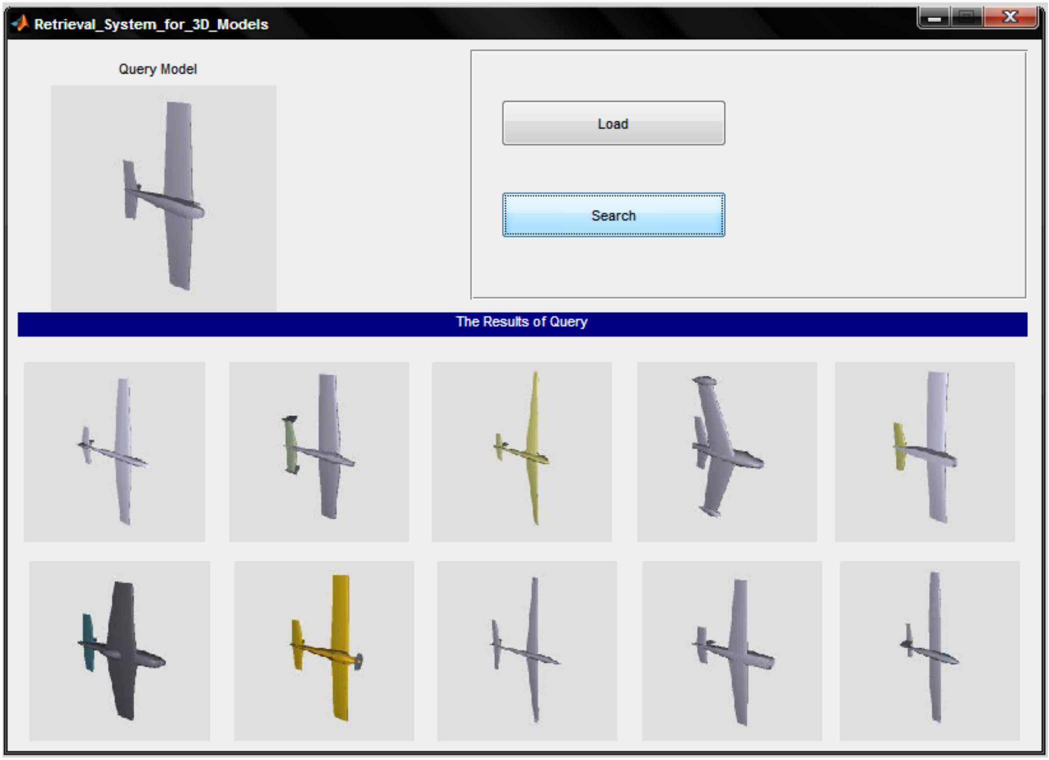


Fig. 6 Retrieval results by our approach. (Example 2)

To well evaluate the performance of the method we use the diagram of recall and precision. This is a traditional and a common way of evaluating performance in documental and visual information retrieval. The recall measures the ability of the system to retrieve all models that are relevant; and the precision measures the ability of the system to retrieve only relevant models. They are defined as follows:

$$\mathrm{Re}\,call = \frac{relevant\,correctly\ retrieved}{\text{all relevant}}$$

$$\mathrm{Pr}\,ecision = \frac{relevant\,correctly\ retrieved}{\text{all retrieved}}$$

To well show the effectiveness of the method, we compared our results with the 3D Zernike moments method [6] (Fig. 7).

Calculating the 3D Zernike descriptor is done by combining linearly the geometrical moments m_{pqr} for all $p,q,r \geq 0$ and $p+q+r \leq n$, with $n \in [0,N]$ and $l \in [0,n]$, such as $(n-l)$ is an even number, and $m \in [-l,l]$.

The 3D Zernike descriptor is the norm of the vector Ω_{nl}: $F_{nl} = \left\|\Omega_{nl}^{m}\right\|$

where $\Omega_{nl}^{m} = \frac{3}{4\Pi} \sum_{p+q+r \leq n} \overline{X_{nlm}^{pqr}} m_{pqr}$

and

$$X_{nlm}^{pqr} = c_l^m 2^{-m} \sum_{v=0}^{k} q_{kl}^{v} \sum_{\alpha=0}^{v} \binom{v}{\alpha} \sum_{\beta=0}^{v-\alpha} \binom{\alpha-v}{\beta} \sum_{u=0}^{m} (-1)^{m-u} \binom{m}{u} i^u \sum_{\mu=0}^{\left[\frac{l-m}{2}\right]} (-1)^{\mu} 2^{-2\mu} \binom{l}{\mu} \binom{l-\mu}{m+\mu} \sum_{v=0}^{\mu} \binom{\mu}{v}$$

Where $2k = n-1$, c_l^m is the normalization factor and the coefficients q_{kl}^{v} are to guarantee the orthonormality of the functions within the unit sphere:

$$c_l^m = c_l^{-m} = \frac{\sqrt{(2l+1)(l+m)!(l-m)!}}{l!}$$

$$q_{kl}^{v} = \frac{(-1)^k}{2^{2k}} \sqrt{\frac{2l+4k+3}{3}} \binom{2k}{k} (-1)^v \frac{\binom{k}{v}\binom{2(k+l+v)+1}{2k}}{\binom{k+l+v}{k}}$$

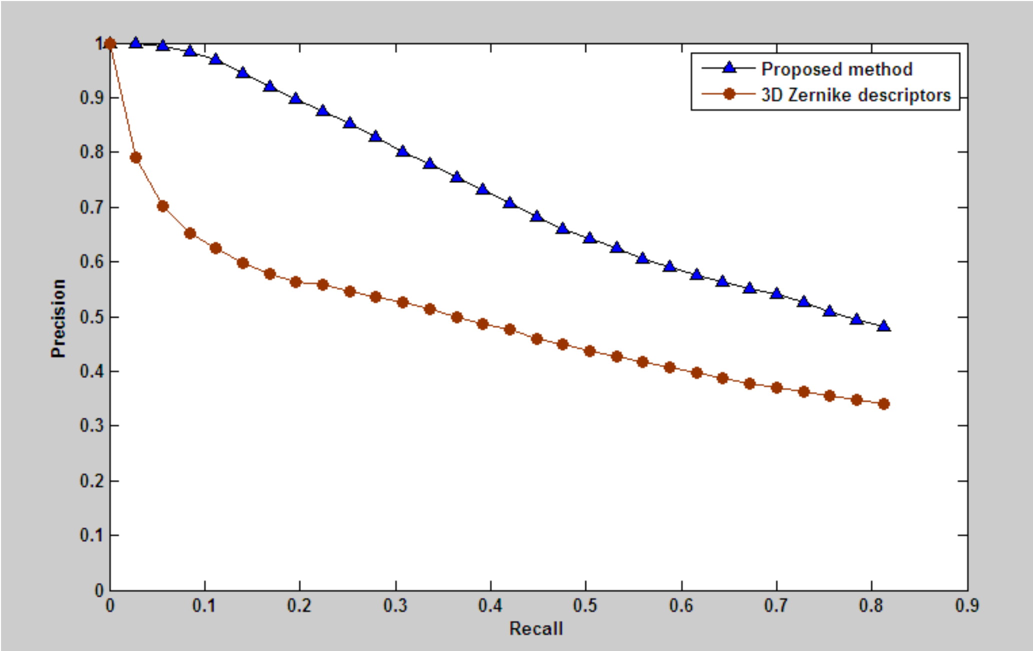


Fig. 7 Overall precision-recall graph for two descriptors

We can see that the proposed shape descriptor is powerful than the 3D Zernike descriptor.

4 Conclusion

In this article, we have proposed a method of 3D model retrieval, for 3D triangular mesh models, based on extracting feature vectors from polygons obtained by cutting each 3D model by plans along a given direction. The similarity between models is calculated by using Earth Mover's Distance (EMD), which allows the measurement between two vectors of dissimilar size. The obtained results show that the proposed descriptor is robust, and the comparison with the will known 3D Zernike moments method shows good performance.

5 Open Problem

We conclude our paper with following open question:

Does the number of cuts can be unsupervised depending on the complexity of the 3D object?

References

[1] Assfalg, J., Del Bimbo, A. et Pala, P. , Curvature Maps for 3D CBR, International Conference on Multimedia and Expo (ICME '03), Baltimore, MD, Etat-Unis,volume 1, pages 301–304.

[2] Assfalg, J., Del Bimbo, A. et Pala, P., Retrieval of 3D Objects using Curvature Maps and Weighted Walkthroughs, International Conference on Image Analysis and Processing (ICIAP '03), Mantova, Italie,pages 348–353.

[3] Attali, D., Bertolino, P. et Montanvert, A., Using polyballs to approximate shapes and skeletons, International Conference on Pattern Recognition (ICPR '94), Jerusalem,pages 626–628.

[4] Canterakis, N., 3D Zernike Moments and Zernike Affine Invariants for 3D Image Analysis and Recognition, International Conference on Image Analysis (ICIA '99). Kangerlussuaq, Groenland, 1999, pages 47, 48, 49, 110, 144.

[5] D. Y . Chen, X. P . Tian, Y . T . Shen, and M. Ouhyoung, On visual similarity based 3D model retrieval, In Eurographics, volume 22, 2003, pages 223–232.

[6] Novotni, M. et Klein, R., 3D Zernike descriptors for content based shape retrieval, The eighth ACM Symposium on Solid Modeling and Applications (SM '03), 2003, pages 216–225.

[7] Ogniewicz, R. et Ilg, M., Voronoi skeletons: Theory and applications, Computer Vision and Pattern Recognition, 1992, pages 63–69.

[8] S. Peleg; M. Werman, and H. Rom, A unified approach to the change of resolution: Space and gray-level, IEEE Transactions on Pattern Analysis and Machine Intelligence11: 739–742, 1989.

[9] Yossi Rubner; Carlo Tomasi, Leonidas J. Guibas, A Metric for Distributions with Applications to Image Databases, Proceedings ICCV, 1998, 59–66.

[10] J. Stolfi, personal communication to L. J. Guibas, 1994.

[11] D.Vranic, D.Saupe, and J.Richter, Tools for 3D-object retrieval: Karhune Loeve transform and spherical harmonics, 2001 Workshop Multimedia Signal Processing, Cannes , France, October 2001.

[12] Mémoire sur la théorie des déblais et des remblais, Histoire de l'Académie Royale des Science, Année 1781, avec les Mémoires de Mathématique et de Physique.

[13] National Taiwan University. NTU Database. http://3d.csie.ntu.edu.tw