

## 3D CONTENT-BASED RETRIEVAL IN ARTWORK DATABASES

David Gorisse<sup>(1)</sup>, Matthieu Cord<sup>(2)</sup>, Michel Jordan<sup>(1)</sup>, Sylvie Philipp-Foliguet<sup>(1)</sup>, Frédéric Precioso<sup>(1)</sup>

(1) ETIS – CNRS, Université de Cergy-Pontoise / ENSEA  
6 avenue du Ponceau, F95014 Cergy-Pontoise - France

(2) LIP6 – CNRS, Université P.M. Curie, Paris, France

### ABSTRACT

In this paper, we present first results obtained in the frame of the EROS-3D project, which aims at dealing with a collection of artwork 3D models, *i.e.* visualize them, classify them and compare them. Some 3D descriptors are used, in association with our active learning search engine RETIN. 3D features are described as well as our new system of classification and retrieval of objects, which we called RETIN-3D.

**Index Terms** — 3D model, indexing, retrieval, active learning, cultural heritage.

### 1 INTRODUCTION

EROS-3D project aims at dealing with the management of large 3D artwork object collections. The C2RMF (Center for Research and Restoration of the Museums of France) organizes for several years campaigns of digitalization in museums from all France. The objective of the project is to develop a software architecture to store and to manipulate (that is to display, to retrieve and to compare) these data with various levels of use.

This tool is firstly dedicated to historians and archaeologists. With such a tool, they will be able to find, display and compare artworks in a few clicks. At the moment, these persons have no tool to help their searches. They have no choice but to physically go and visit the various museums to get information. One can easily understand why such a tool will largely facilitate their works. But the database is not intended to be reserved for the professionals. In the future, museum visitors provided with their PDA for instance will have the opportunity to ask the EROS-3D database in front of a statue and thus obtain additional information.

At this time, the EROS-3D database contains 650 objects digitalized as very high definition 3D models (100.000 and 3.000.000 points). These objects mainly are figurines of the Gallo-Roman civilization (-100 - 300 after JC): Mother Divinity, Venus, moulds, vases and many fragments (fig.1).

The C2RMF pays a particular attention to two categories of objects of the base, figurines representing Mother Divinities and those representing Venus. The C2RMF has three objectives:

1) to extract Mother Divinities and Venus;

- 2) to separate the Mother Divinities from the Venus;
- 3) to separate three Mother Divinity types: those carrying two children, one child in her right arm, and one child in her left arm.



Fig. 1 Some of the objects in the EROS-3D collection: Venus, Mother Divinity, Fragment, Mould.

We present here our “RETIN-3D” search engine, which is based on a set of 3D descriptors (section 2) combined with an active learning strategy (section 3).

There are some 3D search engines available at this time. Some of them are asked by the way of textual requests or 2D/3D sketches (see [1] for a popular one), the query could also consist in a 3D model for some others search engines (see for instance [2] for a recent one). RETIN-3D belongs to this second class, and uses user interactions in order to enhance the search results.

### 2 3D FEATURES

Each 3D model of the database is represented by features which aim at giving a complete and compact description of the object, easy to compute and to deal with. In this paper, we use some widely used descriptors: cord histograms, extended Gaussian images and 3D Hough transform. Some of these descriptors need an appropriate preprocessing of the 3D models.

#### Spatial alignment preprocessing

The pose normalization aims at putting the 3D models in a canonical coordinate system, in order to be invariant to translation, rotation and scaling.

The origin of the coordinate system is set at the center of gravity of the object, and the spatial alignment is then achieved through a PCA (principal component analysis) transform. However, this transform does not provide the axis orientation, and the dissimilarity measures that will be used have to deal with these ambiguities in order to lead to rotation invariant descriptors.

## Cord Histograms

A cord is defined as the vector from the model center to a vertex [3]. Three features are defined from the cords: (i) the length of the cord; (ii) the angle between the cord and the first principal axis; (iii) the angle between the cord and the second principal axis. Two normalized histograms describing the distribution of the first two features are built, leading to an first signature named “Cord2D” (fig.2a).

## EGI (Extended Gaussian Images)

EGI (Extended Gaussian Images) were first introduced by Horn [4]. Each 3D object is projected onto a Gaussian sphere, and each point of the sphere is valued with the total area of the object faces of the same orientation (fig.2b). One obtains for each facet  $P$  of the Gaussian sphere of orientation  $n_k$ :

$$P_{n_k} = \sum_{l=1}^{N_k} A_{l,n_k} \quad (1)$$

Where  $N_k$  is the number of faces of the object in direction  $n_k$  and  $A_{l,n_k}$  is the area of the face of orientation  $n_k$ .

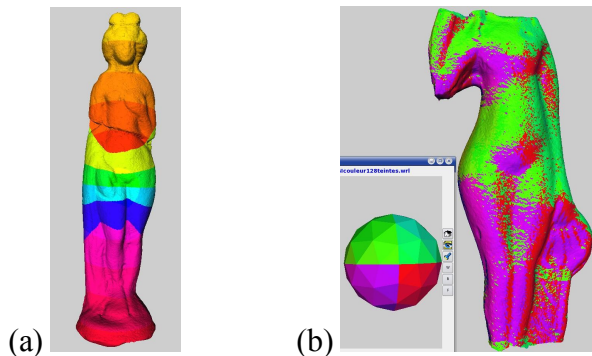


Fig. 2 Colour representation of (a) the Cord2D features and (b) the face orientations for the EGI descriptor.

This feature is invariant in translation, but is not able to differentiate convex and concave parts of the objects.

## CEGI (Complex Extended Gaussian Images)

The Complex EGI (CEGI) [5] is a variant of the EGI which allows solving the problem of discrimination between concavities and convexities. The CEGI feature describes the object thanks to two attributes: the face orientation and the distance between the face and the center of gravity of the object. For each facet of the Gaussian sphere, an accumulation of these two attributes is performed in the complex space.

With the same notations as (Eq 1):

$$P_{n_k} = \sum_{l=1}^{N_k} A_{l,n_k} e^{jd_{l,k}} \quad (2)$$

where  $d_{l,k}$  is the distance between the center of face  $l$  and the object center.

To increase the difference between concavities and convexities, this distance is signed. It is negative when the face is directed towards the object center and positive else. Finally module and phase are computed and compose the CEGI descriptor.

## 3D Hough transform

The 3D Hough feature (Hough3D) [6] is an extension of the Hough transform [7] adapted to 3D objects. It consists in accumulating the parameters of the planes defining the faces of the object. In spherical coordinates, a plane is uniquely defined by the triplet  $(s, \theta, \varphi)$ , where  $s$  is the distance of the plane to the origin,  $\theta$  the angle of azimuth and  $\varphi$  the angle of elevation. A 3D histogram is computed for the triplets (corresponding to the coordinates of the face centers), where each face contributes proportionally to its area.

This feature can be presented as an extension of the EGI. Indeed, an EGI consists in accumulating the area of faces of same orientation (*i.e.* of same rise and azimuth). Hough3D consists thus in computing an EGI on the set of faces located at distance  $s$  to the object center [6] and repeat this calculus for a set of distances  $s$ .

## 3 INTERACTIVE SEARCH

### The RETIN System

We developed in our laboratory the RETIN search engine [8] which is dedicated to interactive image classification and retrieval. Hereafter we proposed an adaptation of this engine “RETIN-3D” in order to perform interactive 3D object classification and retrieval using active learning.

The user provides an object as a query and the system has to retrieve similar objects from the database. The similarity is measured thanks to the features and the models are displayed according to their similarity to the query (fig.3). The problem is then a classification problem.

Because of the great variability in the shapes of the objects within a category (in particular because of broken parts), no set of features is able to represent the category in a univocal way. The user will thus lead the search by annotating displayed objects as belonging or not to the searched category. This gives a big flexibility to the system, since the classification is achieved online and according to the user attempts. That is to say, the results obtained by a user searching for vases with two handles won't be the same as those obtained by a user searching for any type of vase.

The problem is thus a two-class classification problem, with a semi-supervised learning, actually active learning,

since the learning set is enriched at each iteration by new examples and counter-examples provided to the system thanks to the user annotations.

The goal is to separate two classes with a function induced from available examples of both classes and thus to build a classifier that will properly work on unknown objects, *i.e.* which generalizes efficiently the classes defined from the examples. The Support Vector Machine (SVM) is a robust tool for classification in noisy and complex domains, which constructs fast classifiers for massive data. It allows a non-linear classification into two classes without requiring explicitly a non-linear algorithm thanks to the kernel theory. We have compared various kernel functions, and then adopted a Gaussian kernel, with various distances.

In order to perform active learning, the closest objects to the border of the class are displayed in a specific panel at the bottom of the interface (“active learning panel”, see fig.3).

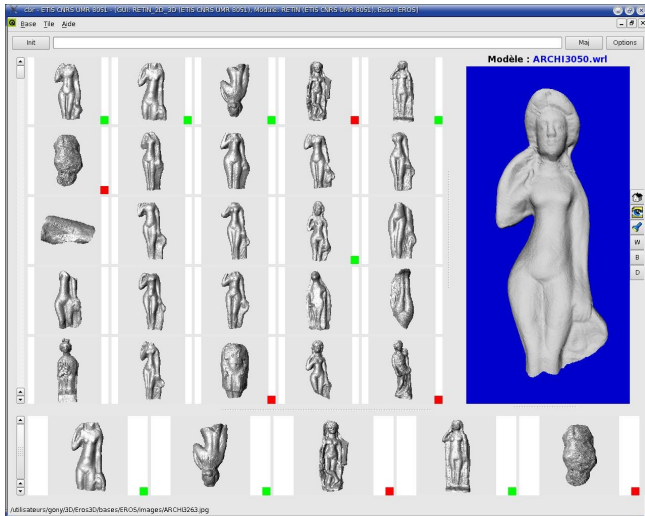


Fig. 3 User friendly interface: on the left part, 3D models ranked by their classification rate, from top to bottom, left to right. Objects are annotated with a green (resp. red) mark if they are relevant (resp. irrelevant) to the request. On the right part, one of the object of the class (Venus). At the bottom, the active learning panel.

Our aim is to compare the descriptors of section 2 and to check their capability to solve the classification problem presented in introduction. The main adaptations of the search engine RETIN reside in (i) the data representation; (ii) the kernel function that will be used. Therefore we propose here to use a Gaussian kernel function with distances  $\chi_1$  or  $\chi_2$ , which showed the best results when tried on real data.

### Tests and results

To compare the features, we manually extracted the various classes of objects presented above, in order to construct a ground-truth. For Venus and Mother Divinity categories, we

made a rather wide choice, by keeping the damaged statues (fig.3).

We present the best results obtained for each feature, through precision/recall curves.

The recall  $R$  and the precision  $Pr$  are defined as follows:

$$R = \frac{N_{ok}}{T} \text{ and } Pr = \frac{N_{ok}}{P}$$

with  $N_{ok}$  the number of objects correctly found among the  $P$  first returned objects;  $T$  the size of the class.

One obtains thus a curve  $Pr/R$  for each query. The process is repeated for each object of the category and curves are averaged over all objects of the class.

Every feature was tested with different parameters, only the best results are displayed. We tested the cord histogram in its standard version (3 histograms of 128 values) and the Cord2D version with  $12 \times 12$  or  $16 \times 16$  values. EGI and CEGI were tested with a Gaussian sphere in 128, 256 or 512 directions. Hough3D was tested with 128 directions and 4 or 8 distances.

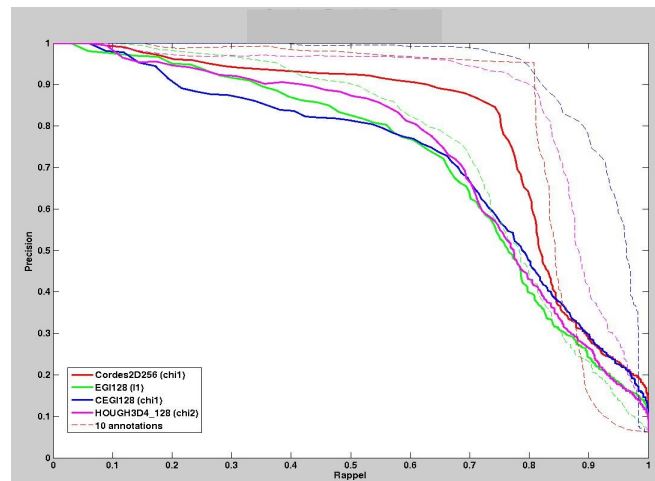


Fig. 4 - Precision/recall diagram for the Mother Divinity class.

Figures 4 and 5 display the precision/recall curves for the two classes of interest: Mother Divinities and Venus. The solid curves present the results of the initial classification (without active learning annotations, *i.e.* with only one example), and the dotted ones after 10 annotations (*i.e.* 10 objects have been annotated, either relevant or irrelevant).

The first remark is that the results for the Mother Divinity class are better than those for the Venus class, which seems to show that the Venus class is sparser than the other one. For both classes, the best results are obtained when using the Cord2D signature with 256 values, then Hough3D with 4 directions, and finally EGI and CEGI. It could be a bit surprising that EGI gives better results than CEGI, this latter supporting more information. We must suppose the phase information is not interesting for this database, so we will minor its weight in the future.

We also tried the SS3D (3D Shape Spectrum) descriptor [9] on these data. This descriptor is based on the analysis of local curvatures. The results were not satisfying, since our meshes are very accurate and most of their faces are plane. Thus a mesh preprocessing (to merge adjacent coplanar faces) should be necessary before computing the SS3D descriptor.

One can also notice that the results are much better after active learning: the best descriptors return about 80% of the classes, with less than 10% of errors after 10 annotations. These results also show that tuning the descriptors and the distances is not an easy task, but that with a bit of learning, we are able to retrieve most of the Mother Divinities, though it is more difficult for the Venus.

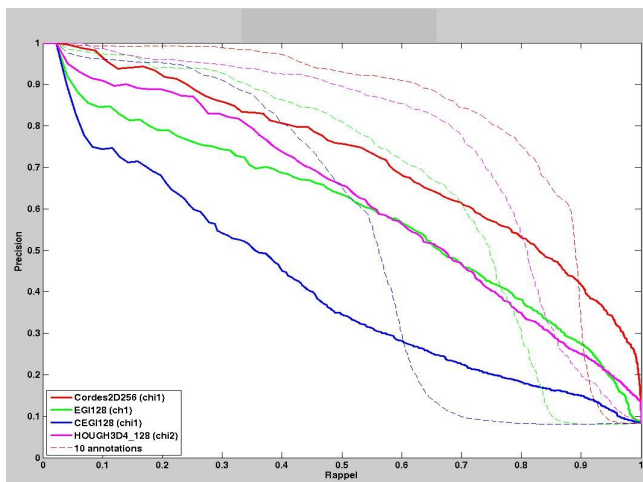


Fig. 5 - Precision/recall diagram for the Venus class.

We then tried to split the Mother Divinity class into 3 sub-classes, as explained in §1. Even if keeping only the intact statuses, the results with all descriptors were disappointing though we must be very careful because of the small number of representatives of each class. It is affordable that this kind of global descriptor cannot be helpful for discriminating local details of the 3D model, such as a child in an arm.

#### 4 CONCLUSION

We present in this paper a method for indexing and retrieving artwork 3D objects (ancient statues and fragments). This method, based on wide-used 3D descriptors and an active learning search system, shows promising results on a real but small test database. It allows extracting significant model classes with a little user interaction.

The combination of the adapted version of the RETIN search engine (RETIN-3D) and adaptations of 3D shape descriptors such as “Cord2D” has been successfully tested and present interesting performances. Another advantage of the interactive search is that it gives the user the ability to

visually evaluate at each step the classification results and stop the search as soon as the results are “good enough”. A typical classification task is performed by mean of a few user interactions (between 5 and 10 model annotations) and usually takes less than 1 minute.

Perspectives of this work include integration of new local shape descriptors and database extensions.

**Acknowledgments:** this project is conducted under french research agency (ANR) contract no. ANR-05-MMSA-0001-05. We particularly thank Pr. Christian Lahanier and the C2RMF (Louvre, Paris) for the 3D model collections that are used in this evaluation.

#### 5 REFERENCES

- [1] “The Princeton 3D Model Search Engine”, <http://shape.cs.princeton.edu/search.html>
- [2] D. Zarpalas, P. Daras, A. Axenopoulos, D. Tzovaras and M.G. Strinzis, “3D Model Search and Retrieval Using the Spherical Trace Transform”, *EURASIP Journal on Advances in Signal Processing*, 2007.
- [3] E. Paquet, M. Rioux, A. Murching, T. Naveen and A. Tabatabai, “Description of Shape Information for 2D and 3D Objects”, *Image Communication Journal*, 16, pp. 103-122, 2000.
- [4] B. K. P. Horn, “Extended Gaussian Image”, *Proc. of the IEEE*, Vol.72, pp. 1671-1686, 1984.
- [5] S. B. Kang and K. Ikeuchi, “The Complex EGI: a new representation for 3D pose determination”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.16(3), pp.249-258, 1994.
- [6] T. Zaharia and F. Prêteux, “Shape-based retrieval of 3D mesh models”, *Proc. of 2002 IEEE Int. Conf. on Multimedia and Expo (ICME'2002)*, Lausanne, Switzerland, 2002.
- [7] P.V.C. Hough, “Machine Analysis of Bubble Chamber Pictures”, *Int. Conf. on High Energy Accelerators and Instrumentation*, CERN, 1959.
- [8] P.H. Gosselin and M. Cord, “Feature based approach to semi supervised similarity learning”, *Pattern Recognition*, 39, pp.1839-1851, 2006.
- [9] T. Zaharia and F. Prêteux, “3D shape-based retrieval within the MPEG7 framework”, *Proceedings SPIE Conference on Nonlinear Image Processing and Pattern Analysis*, San Jose, California, 2001.