PATREC 3693
5 May 2005 Disk Used
ARTICLE IN PRESS
No. of Pages 13, DTD = 5.0.1

ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Pattern Recognition Letters xxx (2005) xxx–xxx

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Generalizations of angular radial transform for 2D and 3D shape retrieval

Julien Ricard *, David Coeurjolly, Atilla Baskurt

Laboratoire d'InfoRmatique en Image et Systèmes d'information, UMR 5205, CNRS/Université Claude Bernard Lyon 1, 43 Bd du 11 novembre 1918, Villeurbanne F-69622, France

Received 8 July 2004; received in revised form 15 March 2005

Communicated by S. Dickinson

Abstract

The angular radial transform (ART) is a moment-based image description method adopted in MPEG-7 as a 2D region-based shape descriptor. This paper proposes generalizations of the ART to describe two-dimensional images and three-dimensional models. First, we propose an 2D extension, called GART, which allows applying ART to images while insuring robustness to all possible rotations and to perspective deformations. Then, we generalize the ART to index 3D models. This new 3D shape descriptor, so called 3D ART, has the same properties that the original transform: robustness to rotation, translation, noise and scaling while keeping a compact size and a good retrieval cost. The size of the descriptor is an essential evaluation parameter on which depends the response time of a content-based retrieval system. For both generalizations, many experiments were made on large databases and have shown, that GART outperforms ART in accuracy at the cost of speed, and that 3D ART outperforms the spherical harmonics shape descriptor (Vranic, D.V., Saupe, D., 2002. Description of 3D-shape using a complex function on the sphere, in: IEEE International Conference on Multimedia and Expo (ICME 2002), Lausanne, Switzerland, 2002, pp. 177–180; Funkhouser, T., Min, P. Kazhdan, M., Chen, J., Halderman, A., Dobkin, D., Jacobs, D., 2003. A search engine for 3D models. ACM Trans. Graphics 22(1), 83–105) in speed at the cost of accuracy.

Keywords: Content-based retrieval; Shape descriptor; Angular radial transform; 3D models; Images

1. Introduction

Content-based image retrieval has been a topic of intensive research in recent years, and particu-

* Corresponding author. Tel.: +33/0 4 72 44 83 95; fax: +33/0 4 72 43 15 36.
E-mail address: jricard@liris.cnrs.fr (J. Ricard).
URL: http://liris.cnrs.fr/julien.ricard (J. Ricard).

0167-8655/$ - see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.patrec.2005.03.030

larly the development of effective shape descriptors (SD). The MPEG-7 standard committee has proposed a region base shape descriptor, the angular radial transform (ART) (Jeannin, 2001; Kim and Kim, 1999). This SD has many properties: compact size, robustness to noise and scaling, invariance to rotation, ability to describe complex objects. These properties and the evaluation made during the MPEG-7 standardization process make the ART a unanimously recognized efficient descriptor. Furthermore, an important characteristic is the small size of the ART descriptor. For a huge database, this implies fast answers during retrieval processes. In the MPEG-7 standard, the ART similarity measure is reduced to a L_1 distance between 35 floating point values.

In the same time, the technical 3D model databases grow up since the beginning of the computer-aided design. The engineering laboratories and the design offices always increase the number of 3D solid objects and the current industrial estimations point to the existence of over 30 billion of CAD models.

This huge number of models requires a content-based mining with indexing and retrieval processes. In the framework of the Semantic-3D national project and in partnership with the car manufacturer Renault, we investigate the possibilities to make a fast descriptor to index a huge technical 3D models database and to index images to insure robustness to deformation undergone by objects in natural images. In this context, we explore the possibilities to extend ART to the retrieval of images and 3D models by taking into account the specific properties of these data.

This article presents our work on the Angular Radial Transform. First, we generalize the 2D ART shape descriptor to insure robustness to perspective deformations that can disturb a planar shape in a 2D natural image. In a second time, the ART is extended for the indexation of 3D models while preserving its properties. This paper is organized as follows: Section 2 presents the ART transform, Section 3 details the generalization of the 2D ART, Section 4 presents a survey of the related work on 3D shape matching and our new 3D ART descriptor, results are presented and discussed in the last section.

2. The angular radial transform

This part presents the 2D ART proposed by the MPEG-7 normalization process. These definitions are the starting point of the proposed generalizations. Angular radial transform (ART) is a moment-based image description method adopted in MPEG-7 as a region-based shape descriptor (Bober, 2001). It gives a compact and efficient way to express pixel distribution within a 2D object region; it can describe both connected and disconnected region shapes. The ART is a complex orthogonal unitary transform defined on a unit disk based on complex orthogonal sinusoidal basis functions in polar coordinates (Jeannin, 2001; Kim and Kim, 1999). The ART coefficients, F_{nm} of order n and m, are defined by

$$F_{nm} = \int_0^{2\pi}\int_0^1 V_{nm}(\rho,\theta) f(\rho,\theta)\rho \, \mathrm{d}\rho \, \mathrm{d}\theta \tag{1}$$

where $f(\rho,\theta)$ is an image function in polar coordinates and $v_{nm}(\rho,\theta)$ is the ART basis function that is separable along the angular and radial directions:

$$V_{nm}(\rho,\theta) = A_m(\theta)R_n(\rho) \quad \text{where} \begin{cases} A_m(\theta) = \frac{1}{2\pi}\exp(jm\theta) \\ R_n(\rho) = \begin{cases} 1 & n = 0 \\ 2\cos(\pi n\rho) & n \neq 0 \end{cases} \end{cases} \tag{2}$$

In order to achieve rotation invariance, an exponential function is used for the angular basis function. The radial basis function is defined by a cosine function. Real parts of basis functions are shown in Fig. 1.

The ART descriptor is defined as a set of normalized magnitudes of the set of ART coefficients. Rotational invariance is obtained by using the magnitude of the coefficients. In MPEG-7, 12 angular and three radial functions are used ($n < 3$, $m < 12$) (Jeannin, 2001), these values will be used in the rest of the articles. For scale normalization, the ART coefficients are divided by the magnitude of the ART coefficient of order $n = 0$, $m = 0$. The distance between two shapes described by the ART descriptor is calculated using L_1 norm:

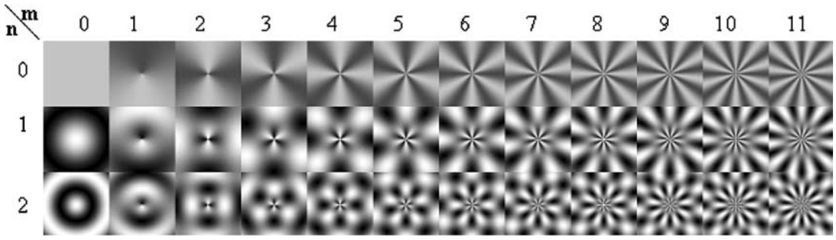

Fig. 1. Real parts of the ART basis functions.

$$d_{\mathrm{ART}}(Q,I) = \sum_{i=0}^{n \cdot m} \|\mathrm{ART}_Q[i] - \mathrm{ART}_I[i]\| \qquad (3)$$

The subscripts Q and I represent respectively the query image and an image in the database, and ART_I is the array of the normalized ART coefficients of the image I. Note that to decrease the descriptor size, quantification can be applied to each coefficient using four bits per coefficient (Jeannin, 2001). The MPEG-7 standardization process showed the efficiency of the method in the 2D indexing field. We can quote the use of ART in a multi-views 3D models retrieval (Chen and Ouhyoung, 2002), and in face detection (Fang and Qiu, 2003).

To use ART on a natural color image and to take into account the internal variations of the objects (contours, holes, texture,...), the ART descriptor can be computed on the luminance component. In that case the function $f(\rho, \theta)$ takes the values in the interval $[0, 1]$ (see Wang et al., xxxx; Laaksonen et al., xxxx; Akcay et al., 2002).

3. Generalization of ART to perspective deformations

The goal of this generalization is to make the ART robust to any rotations or perspective projections. A planar object in a natural scene can be viewed according to all orientations and can be carried by an unspecified plan. This highly probable situation will disturb the shape in the image and will prevent the identification. In Fig. 2, a plane object (a stamp) is seen with three angles of acquisition which correspond to three different shapes projected on the same image plane. To make ART descriptor robust to all possible rotations and to perspective projections it is necessary to generalize the ART transform with new basis functions. This new descriptor is called generalized ART (GART).

In order to define the transformations undergone by an object during rotations and projections onto the image plane, we consider the transformation space given by the perspective coefficient p and the normal vector to the image plane, denoted $\vec{n}$. The perspective coefficient p defines a distance

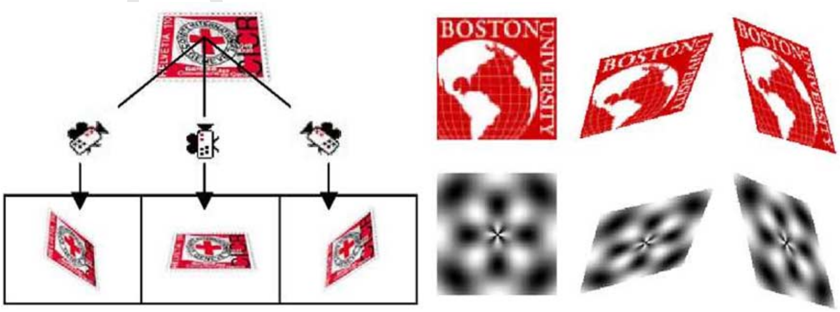

Fig. 2. Object seen according to various angles and example of projected basic functions on the object support plane.

between the original basis function and the image plane. The normal vector $\vec{n}$ is given by its Euler angles ς (radial direction) and ϕ (rotation angle) defined between the vectors $\vec{n}$ and the axis $\vec{x}$ (see the Fig. 3). The first two parameters define the orientation of the normal to the image plane and the third is the perspective coefficient which defines the perspective deformation. Fig. 3 shows these parameters.

This transformation space is sampled for each parameter according to k_ς, k_ϕ and k_p values. Hence we obtain a sampling of $K = k_\varsigma * k_\phi * k_p$ transformations. The basis functions are deformed in the same way according to the K transformations. Each object is indexed with these K sets of projected basis functions. The number of projections is limited to keep a reasonable computational cost. The values $k_\varsigma = 12$, $k_\phi = 3$ and $k_p = 3$, are chosen in our experiments presented in Section 5, because these values give the better ratio of cost to efficiency. In other words, we have $K = 108$ sets of coefficients to describe a shape. Hence we have to compute 108 similarity measures between a query object and a database object.

The complexity of the classical ART is in $\theta(n * m * N^2)$ because we compute $n * m$ basis functions values for the $N * N$ pixels of the image. The generalized ART, which creates K set of basis functions, has a complexity in $\theta(K * n * m * N^2)$.

To make the retrieval process faster, we choose to inverse the indexation and retrieval processes. Without optimization, the indexing process computes the ART descriptor between the original object and the original basis functions whereas the retrieval process computes the descriptor between the extracted object from a natural image and all the projected basis functions. Thus, the indexation process, which is offline, has a computation cost K times less than the retrieval process, which is online. Fortunately, it is possible to inverse these two processes and to index the original objects on the inverse projected basis functions, whereas an extracted object will be indexed only on the original basis functions. This increases the cost of the offline indexing process but decreases the online retrieval process without modification of the description (Table 1).

As it is shown in Fig. 4, we transform the original image F_0 to the deformed image F_1 by a transformation T and we transform the origin basis function V_0 to the inverse projected basis function V_{-1} by a transformation T^{-1}. We can see that we obtained the same descriptor values, if we index the deformed image F_1 on the original basis function V_0, or if we index the origin image F_0 on the inverse basis function V_{-1}.

Each object is described by $K = k\varsigma * k_\phi * k_p$ series of ART coefficients created from the basis functions projected on K planes of projections. The shape similarity distance is achieved by computing a set of distances $d_{\mathrm{ART}}(Q, I_j)$. For each value of j, the ART coefficients of Q, computed on the original basis functions, and those of I, computed on the jth projection of the basis functions, are compared using (3). Then the shape distance between Q and I is given by

Table 1
Number of online and offline ART descriptor computation during the original process and the optimized one

	Original process	Optimized process
Online	K	1
Offline	1	K

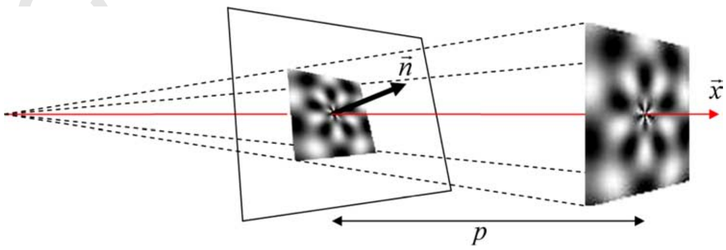

Fig. 3. The basis functions are projected on the image plane I according to ς, ϕ and p to obtain the projected basis functions.

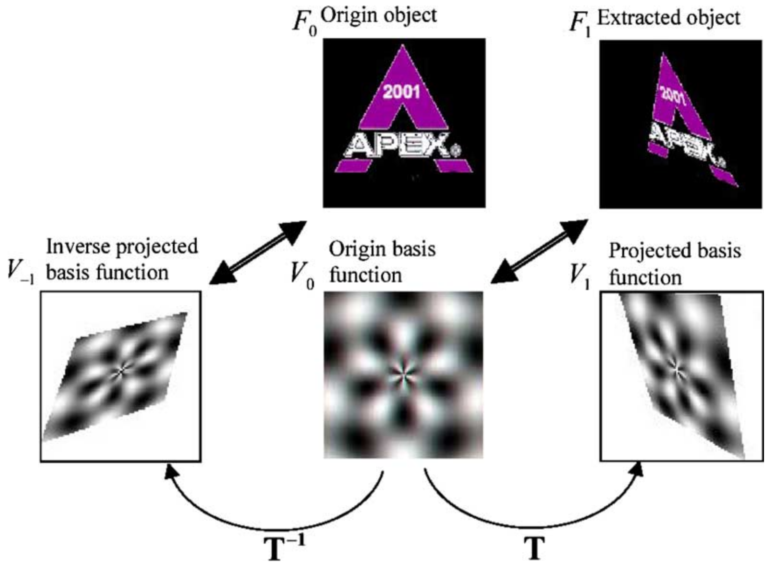

Fig. 4. Diagram of inverse indexation process.

$$d_{\text{shape}}(Q, I) = \min_{j \in K} \sum_{i=0}^{n \cdot m} \|\text{ART}_Q[i] - \text{ART}_I^j[i]\| \qquad (4)$$

where Q is the ART coefficients of the key object and I_j is the coefficients of the I object, calculated on the jth projection of the basis functions. The minimum is considered in order to take into account all the possible perspective views of the object.

4. 3D angular radial transform

In this section, we present a survey of the related works on 3D shape matching, then we generalize the MPEG-7's angular radial transform to the 3D space.

4.1. Survey of recent 3D indexing methods

3D indexing methods can be divided into two distinct groups: retrieval by an example of a three-dimensional model, and retrieval by a 2D view. In this work, we are interested in 3D model retrieval. The state of the art can be divided into two different classes of 3D shape description methods: the structural approaches and the statistical approaches.

The structural approach is a high-level one which aims to describe the shape in a more complete and intuitive manner. The principle is to split an object into sub-parts and to represent the object as the merge of these sub-parts according to adjacency relationships. A segmentation step identifies the elementary structures composing the objects satisfying given homogeneity criteria. The determined components are represented by using some specific structures such as trees or graphs. Two distinct approaches can be considered: the surface-based approaches and the volume-based approaches. A surface-based approach segments surfaces into patches. The connectivity of such patches is encoded within an adjacency graph. A similarity measure is computed between two objects by graph matching techniques (Ullmann, 1976). Dorai and Jain (1997) proposes to use a graph of maximal patches defined by functions of the principal curvature. On 3D technical models, the model signature graph (MSG) (McWherter et al., 2001) is constructed by a surface-based representation of the object. Each face is represented by valued vertices and valued edges exist if two vertices are adjacent. Hilaga et al. (2001) uses multi-resolution Reeb graph. Multi-resolution graph are constructed by computing a surface geodesic distance to define a Reed graph at various levels. Recently, the Augmented Reeb Graphs (Tung and Schmitt, accepted for publication) increases the matching process. The volume-based approaches decompose a shape using 3D elementary volumetric structures called geons (geometrics ions) based on recognition by composant theory (Biederman, 1987). The sets of 3D volumetric primitives may be: cylinders, cubes, parallelepipeds, cone (truncated or not), ellipsoids (Irani and Ware, 2000). Other interesting approaches use a set of superquadrics (Zhou and Kambhamettu, 2002) or quadratic surfaces (Park et al., 2002).

The statistical approaches characterize the 3D model shape by calculating statistical moments (Zhang and Chen, 2001) or by considering a distribution of the measurement of geometric primitives (which might be points, cords, triangles, tetrahedrons,...) (Osada et al., 2001). A geometrical normalization of the object size and position in a 3D space is used as a pre-processing step to guarantee a geometric invariance. The moment-based approaches can be defined as projections of the function defining the object onto a set of characteristic moment functions. These approaches are used in 2D pattern recognition with several 2D moments: geometrical, Legendre, Fourier–Mellin, Zernike, pseudo-Zernike moments (Teh and Chin, 1988) and ART (Jeannin, 2001; Kim and Kim, 1999). Some of these moments have been extended into 3D: 3D Fourier (Elad et al., 2001), 3D Wavelet (Paquet and Rioux, 2000), 3D Zernike (Canterakis, 1999) and the spherical harmonic (SH) decomposition, recently described by Vranic and Saupe (2002), and Funkhouser et al. (2003). The spherical harmonic analysis decomposes a 3D shape into irreducible sets of rotation independent components by sampling the three-dimensional space with concentric shells, where the shells are defined by equal radial intervals. The spherical functions are decomposed as a sum of the first 16 harmonic components (Kazhdan et al., xxxx), in an analogous way to the Fourier decomposition into different frequencies. Using the fact that rotations do not change the norm of the harmonic components, the signature of each spherical function is defined as a list of these 16 norms. Finally, these different signatures are combined to obtain a 32 ∗ 16 signature vector for each 3D model. During the retrieval step, the similarity of objects is calculated as the Euclidean distance between these vectors. In our experimentation, the proposed descriptor 3D ART is compared to SH.

4.2. 3D ART definition

First, we suppose the objects to be represented in spherical coordinates where ϕ is the azimuthal angle in the xy-plane from the x-axis, θ is the polar angle from the z-axis and ρ is the radius from a point to the origin. The 3D ART is a complex unitary transform defined on a unit sphere. The 3D ART coefficients are defined by

$$F_{nm_\theta m_\phi} = \int_0^{2\pi}\int_0^{\pi}\int_0^{1} V_{nm_\theta m_\phi}(\rho,\theta,\phi) \times f(\rho,\theta,\phi)\rho\,\mathrm{d}\rho\,\mathrm{d}\theta\,\mathrm{d}\phi \tag{5}$$

where $F_{nm_\theta m_\phi}$ is the 3D ART coefficient of orders n, m_θ and m_ϕ, $f(p,\theta,\phi)$ is a 3D object function in spherical coordinates and $V_{nm_\theta m_\phi}(\rho,\theta,\phi)$ is a 3D ART basis function (BF). The 3D BFs are separable along the angular and the two radial directions:

$$V_{nm_\theta m_\phi}(\rho,\theta,\phi) = A_{m_\theta}(\theta)A_{m_\phi}(\phi)R_n(\rho) \tag{6}$$

The radial basis function is defined by a cosine function and the angular basis functions are defined by complex exponential functions to achieve rotation invariance and continuity along both θ and ϕ values:

$$R_n(\rho) = \begin{cases} 1 & n = 0 \\ 2\cos(\pi n\rho) & n \neq 0 \end{cases} \quad \text{and} \quad \begin{array}{rcl} A_{m_\theta}(\theta) & = & \frac{1}{2\pi}\exp(2\mathrm{j}m_\theta\theta) \\ A_{m_\phi}(\phi) & = & \frac{1}{2\pi}\exp(\mathrm{j}m_\phi\phi) \end{array} \tag{7}$$

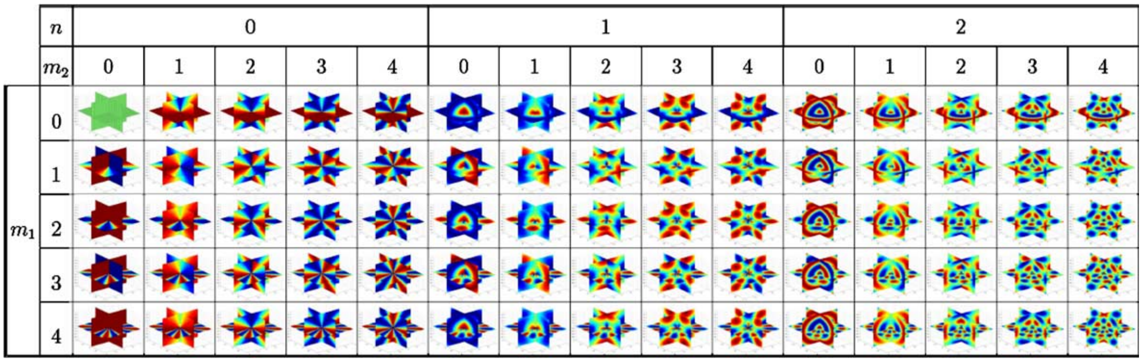

Fig. 5. Real parts of 3D ART BF.

The values of the parameters n, m_θ and m_ϕ are trade-offs between efficiency and accuracy. Choices were made by computing the Recall response for different sets of values. For the technical database presented in Section 4, we have chosen $n = 3$, $m_\theta = 5$ and $m_\phi = 5$. The real parts of the 3D ART BFs are shown in Fig. 5.

The similarity measure is computed using a L_1 distance between the 3D ART descriptors:

$$d(Q,I) = \sum_{i=1}^{n \cdot m_\theta \cdot m_\phi} \|\mathrm{ART3D}_Q[i] - \mathrm{ART3D}_I[i]\| \tag{8}$$

where Q and I represent respectively a query object and an object of the database and ART3D is the array of 3D ART descriptor values normalized by F_{000}. The choice of the L1 distance is justified by speed preoccupations but other distances could be used.

4.3. Indexing process

An important property of the 2D ART is the rotation invariance. A 2D rotation representation in polar coordinates can be express as the sum of angular components:

$$(\rho, \phi) \overset{\mathrm{Rot}_\alpha}{\rightarrow} (\rho, \phi + \alpha) \tag{9}$$

That does not modify the norm of the function $A_{m_\theta}(\theta)$ and therefore nor the ART descriptor. In 3D, unspecified rotations cannot be expressed as the sum of constant values on the angular components, and thus modify the descriptor values. However, if we consider a rotation around the z-axis, the norms of the 3D ART coefficients do not change. Hence, to have a rotation invariance, unspecified rotations must be transformed to rotations along z-axis by alignment according to the first principal direction. Thus, a principal components analysis (PCA) is applied to obtain the principal direction of the objects. PCA alignment is not really robust when the three principal directions are considered. Fortunately, here, we only need to align the first principal direction along the z-axis, therefore wrong alignments are limited. Fig. 6 shows the indexation process.

Hence, before projecting 3D models onto the BFs, the objects are pre-processed as follows: first, they are discretized in a grid in such a way to obtain interior and exterior voxels. This discretization is also used to compute the parameters of centering, scaling and alignment to the z-axis: the 3D object is centered on its gravity center and scaled up. This pre-processing step makes the 3D ART robust to translations, rotations and scaling. Finally, the discretized object is projected into the 3D ART BFs to obtain the 3D ART coefficients.

5. Experiments

This part shows the experiments that we have made to evaluate the ART generalizations. First, we present our tests on the 2D GART, then we present the 3D model databases and the experiments that we have made to illustrate the properties and the effectiveness of the 3D ART.

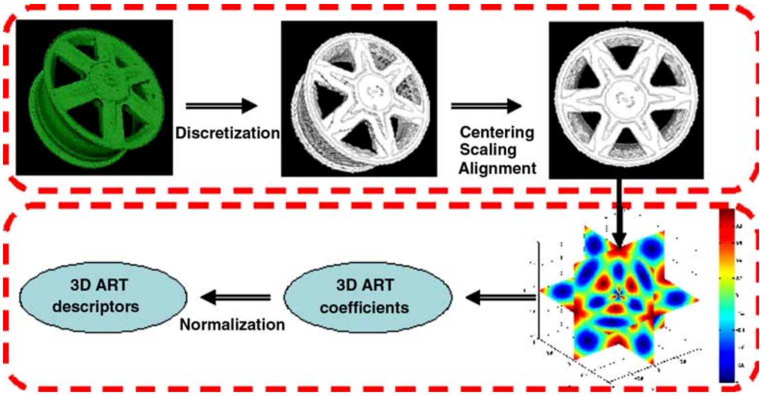

Fig. 6. Indexing process.

5.1. 2D generalized ART experiments

In these experiments, ART and GART was used on the luminance components of the described images like explained in part 2. The first test compares the ART on the luminance and our GART extension. 1813 test images were created from 37 trademark images disturbed according to 49 random perspective projections with illuminating variations and grouped in 37 classes. The original trademark images were used as query images the answer ranks of the class images were evaluated. Note that the GART descriptor is 108 times larger that the ART descriptors and the similarity measures compute have a same cost differences. Fig. 7a shows the recall/precision values. This curve shows that the best results are obtained with GART. GART is found to be more accurate but slower than ART.

The GART was defined for a detection application of a trademark in natural images. This application identifies an object extracted from an image. Its general scheme can be seen in Fig. 8 and can be split into two successive steps: the indexation and the retrieval. To evaluate the properties of the GART within the retrieval process, 50 objects were extracted from the images ranks where one finds the original trademark were considered. For this application, we have also considered a color description to take into account color properties. The color descriptor is a simple color histogram associated with a histogram similarity. The GART and the color description were mixed into a global similarity function computed as a

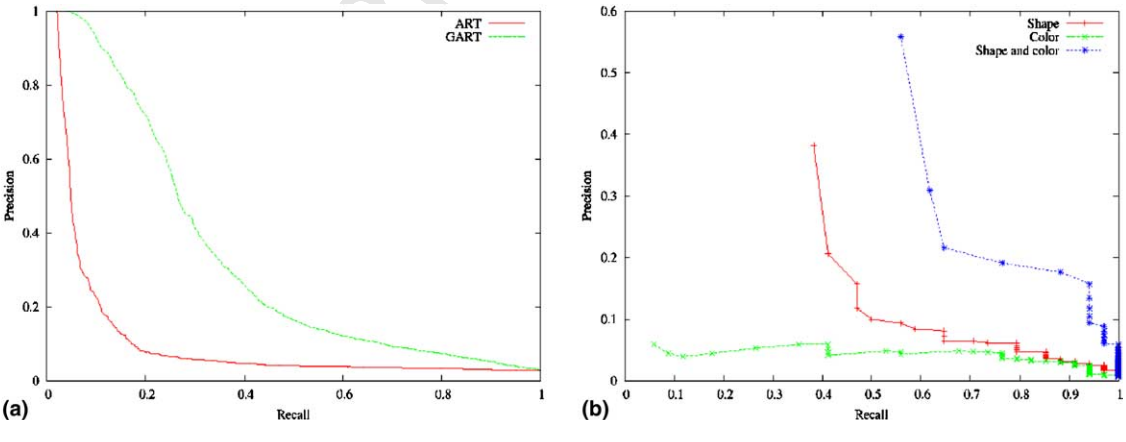

Fig. 7. Recall/precision curves: (a) ART and generalized to perspective projection ART, (b) GART, color and mixed approach.

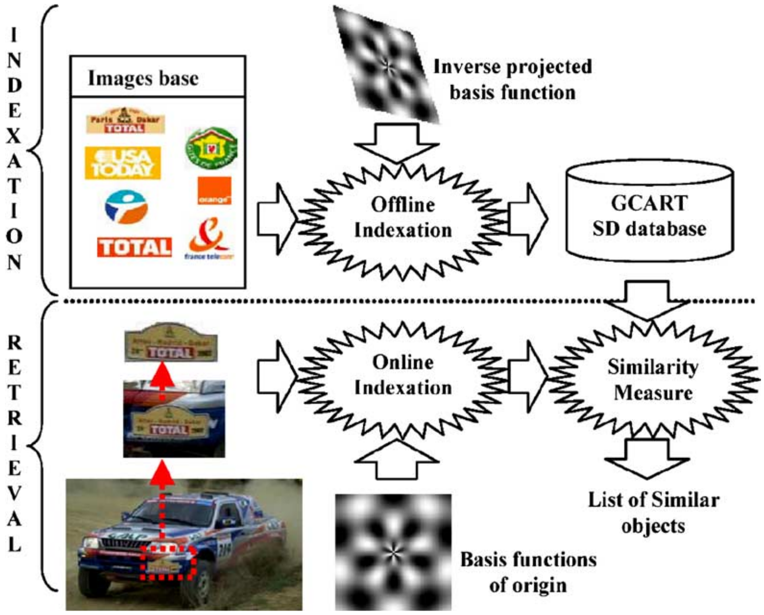

Fig. 8. General diagram of the application.

weighted sum of each distances (Idrissi et al., 2004). The Fig. 7b shows the Recall values for the GART, the color and the mixed description. The mixed description gives the original trademark at the first rank in 55%, against 38% for the GART and 6% for the color. At the rank 10, the original trademark is found in 95% of the cases, whereas the GART and the color study have found the original object, respectively in 65% and 41% of the cases.

5.2. 3D ART experiments

5.2.1. 3D model database test

The 3D experiments are made using two 3D model databases: the Princeton Shape Benchmark (Shilane et al., 2004) and a Renault database. Fig. 9 show examples of 3D models both databases.

The Princeton Shape Benchmark provides a repository of 3D models and software tools to evaluate shape-based retrieval and analysis algorithms. The motivation is to promote the use of standardized data sets and evaluation methods for research in matching, classification, clustering, and recognition of 3D models. The Princeton database contains 1814 models grouped into high-level semantic classes where the objects of a same class are heterogeneous. For example, a class of staircases contains 3D models, which represent staircases of very different shape but with the same semantic (Fig. 10). The Renault database is a technical database, which contains mechanical models. In the framework of SEMANTIC 3D and in part-

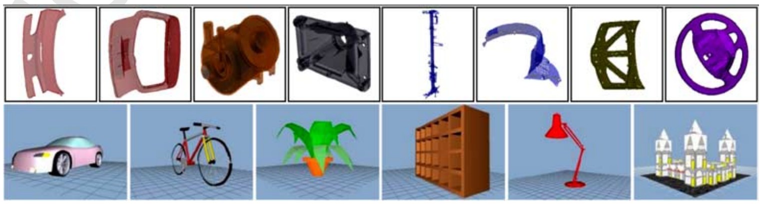

Fig. 9. Examples of Princeton Shape Benchmark 3D models and Renault 3D models.

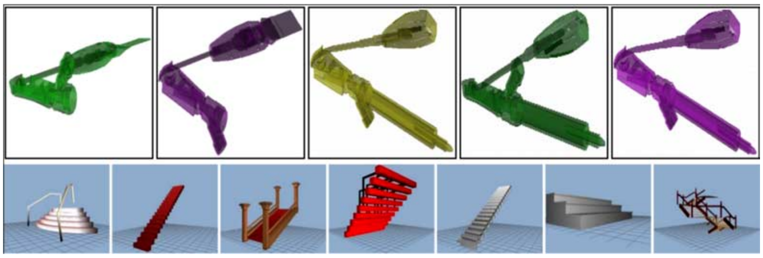

Fig. 10. Example of Princeton Shape Benchmark class: staircase and a Renault database class: seat belt part.

nership with the car manufacturer Renault, we have a huge 3D technical model database (approximately 5000 models). This database contains the pieces composing a car with all the model versions. The 5000 models were classified according to the functionalities of the different parts. 781 objects were classified in 75 classes. We can quote for example the classes: wheel, door, brake pad, disc of brake, bolt,... Not all the database objects can be classified because the database does not have enough models to guaranty a minimal number of models per class. Classes, which have a number of models less than 5, are grouped in an unspecified class. The tests were made by taking all the objects of the specified classes as request objects for the 5000 object database. The recall and precision values are the mean of the recall and precision values of all the objects of the classes. Examples of the two databases classes are shown in Fig. 10.

5.2.2. ART 3D parameters

To fix the parameter values, the recall values are compared. Twelve values of the parameters n, m_θ and m_ϕ are evaluated. Fig. 11a shows that the best results are obtained for $n = 3$ and $m_\theta = m_\phi = 5$. Fig. 11b presents the same experiment with different discretization sizes S. Better results are obtained on the technical database with the parameter value $S = 64$. Thus, we use this value in the rest of this work. This value is also suggested in (Kazhdan et al., xxxx) for the SH computation.

5.2.3. Robustness

To evaluate the robustness of the process, we distort a 3D object according to scaling, rotation, translation and noise. Table 2 shows the maximum and the mean distance obtained for these four distortions. For each distortion, we create a set of 3D objects and for all the objects, we compute the distance to the original one. The translation has no effect on the distance, because the pre-processing step centers the objects. For the same reasons,

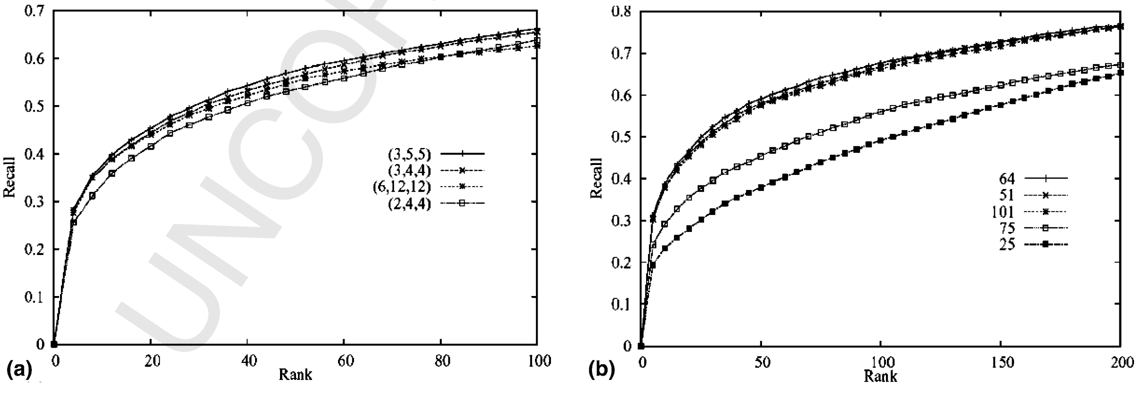

Fig. 11. The Recall values to set up parameters.

Table 2
Distance obtained for several distortions

Distort	Translation	Scale	Rotation	Noise
Max distance	0	0.016	1.272	2.217
Mean distance	0	0.003	0.750	1.012

the scale distortion has small effects due to artifacts of digitization, the maximum distance between the scaled objects are 0.016 when a mean distance between two objects of the same class is around 3. The obtained distances are smaller than intra-class distances and the classification is the same one. The rotation distortion test is a set of rotations around the three axes with random angles and gives a maximum distance of 1.272 and a mean distance of 0.75. The noise distortion is a random move of vertices of the object; each vertex is moved along a random Gaussian vector. This distance is a percentage of the object size. If this distance is higher than 10% the surface of the object is much distorted but the similarity measure is 1.6 and the object are still well classified. Fig. 12 shows distorted objects by the noise distortions.

5.2.4. Comparison

A second experiment is set up to compare the 3D ART to the Spherical Harmonic descriptor (SH). This experiment is made on the two model databases. Fig. 13a and b shows the recall values for SH and 3D ART descriptors for the two databases. On the Princeton database (Fig. 13a), the SH method gives a better description than the ART. The results on the Renault database are similar with the two methods (Fig. 13b). ART description gives better results when the objects of a same class are similar. The 3D ART goal was not to

Table 3
Size (in floating numbers) and indexing time (in seconds) comparison between 3D ART and spherical harmonic representation

	Indexing time	Descriptor size
SH	10	544
3D ART	4	74

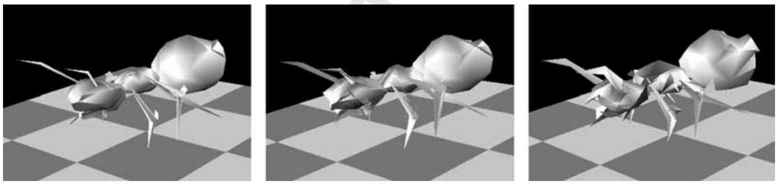

Fig. 12. Example of noise distortions for three distance values: 0%, 5% and 10%.

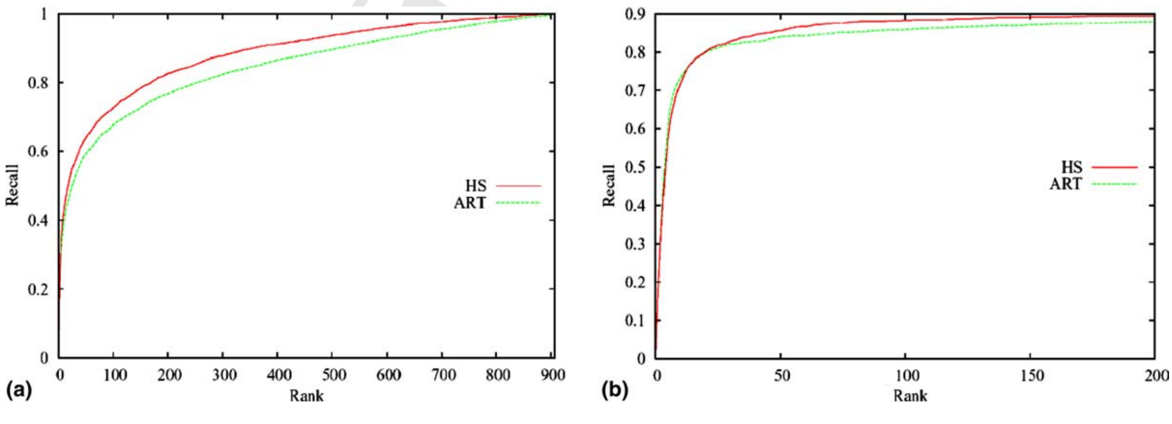

Fig. 13. Recall/precision values on (a) Princeton and (b) Renault databases.

have a the best description accuracy, but to make a small descriptor to have a fast answer.

The computational cost and the size of the descriptors are significant comparison criteria (Table 3). The 3D ART indexing computation time is 2.5 times less than a SH indexing and the descriptor size and the cost of the similarity measure is approximately 7.8 times less. These differences are due to the fact that the ART BF and the integral calculus are defined in the Euclidian space whereas the SH description is computed using complex frequency transformations. In the framework of the SEMANTIC 3D project, a huge 3D models database will be index. Thus, the cost of the retrieval must be as small as possible.

6. Conclusion

In this paper, we have presented an extension of the 2D region-based shape descriptor ART to deformed images and to 3D models. The generalization of the ART (GART), to perspective projections, increases the ART efficiency and definition domain while keeping the discriminating capacities. Moreover the optimized process makes possible to have a light online process and a quick answer for content-based image retrieval. We have shown that GART is more accurate that ART at a higher cost.

In the second part of this work, we have presented the generalization of the ART to describe 3D shape (3D ART). The proposed descriptor is robust to translations, scaling, multi-representation (remeshing, weak distortions), noises and 3D rotations. It fulfils the requirements for our CAD database indexing and retrieval application: robustness and accuracy of the indexing, and high-speed retrieval processes and similarity computation index. Moreover experiments have shown that 3D ART outperforms the spherical harmonics descriptor in speed, while keeping a close accuracy.

Acknowledgement

This work is supported by the French Research Ministry and the RNRT (Réseau National de Recherche en Télécommunications) within the framework of the Semantic-3D national project (`http://www.semantic-3d.net`).

References

Akcay, M., Baskurt, A., Sankur, B., 2002. Measuring similarity between color image regions. European Signal Processing Conference (EUSIPCO'02) 1, 115–118.
Biederman, I., 1987. Recognition by components: A theory of human image understanding. Psychol. Rev. 94, 115–147.
Bober, M., 2001. Mpeg-7 visual shape descriptors. IEEE Trans. Circ. Systems Video Technol. 1 (6), 716–719.
Canterakis, N., 1999. 3D zernike moments and zernike affine invariants for 3D image analysis and recognition. In: 11th Scandinavian International Conference on Image Analysis.
Chen, D.-Y., Ouhyoung, M., 2002. A 3D model alignment and retrieval system. In: International Computer Symposium (ICS 2002), Hualien, ROC.
Dorai, C., Jain, A.K., 1997. Shape spectrum based view grouping and matching of 3D free-form objects. IEEE Trans. Pattern Anal. Machine Intell. 19 (10), 1139–1146.
Elad, M., Tal, A., Ar, S., 2001. Content based retrieval of vrml objects—an iterative and interactive approach. In: The Sixth Eurographics Workshop in Multimedia, pp. 97–108.
Fang, J., Qiu, G., 2003. Human face detection using angular radial transform and support vector machines. In: International Conference on Image Processing (ICIP 2003), vol. I, 2003, pp. 669–672.
Funkhouser, T., Min, P., Kazhdan, M., Chen, J., Halderman, A., Dobkin, D., Jacobs, D., 2003. A search engine for 3D models. ACM Trans. Graphics 22 (1), 83–105.
Hilaga, M., Shinagawa, Y., Kohmura, T., Kunii, T.L., 2001. Topology matching for fully automatic similarity estimation of 3D shapes. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. ACM Press, pp. 203–212.
Idrissi, K., Lavoué, G., Ricard, J., Baskurt, A., 2004. Object of interest-based visual navigation, retrieval, and semantic content identification system. Comput. Vision Image Understanding 94 (1–3), 271–294.
Irani, P., Ware, C., 2000. Diagrams based on structural object perception. In: Proceedings of the Working Conference on Advanced Visual Interfaces. ACM Press, pp. 61–67.
Jeannin, S., 2001. Mpeg-7 Visual part of eXperimentation Model Version 9.0. In: ISO/IEC JTC1/SC29/WG11/N3914, 55th Mpeg Meeting, Pisa, Italy.
Kazhdan, M., Funkhouser, T., Rusinkiewicz, S., Rotation invariant spherical harmonic representation of 3D shape descriptors. In: Symposium on Geometry Processing.
Kim, W.-Y., Kim, Y.-S., 1999. A new region-based shape descriptor. In: Mpeg Meeting, TR 15-01, Pisa.
Laaksonen, J., Koskela, J., Laakso, S.P., Oja, E., PicSOM: Content-based image retrieval with self-organizing maps. Pattern Recognition Letters 21, 13–14.

McWherter, D., Peabody, M., Regli, W.C., 2001. A. Shokoufandeh, Transformation invariant shape similarity comparison of solid models. In: ASME Design Engineering Technical Conferences and 6th Design for Manufacturing Conferences (DETC 2001/DFM-21191), Pittsburgh, Pennsylvania.

Osada, R., Funkhouser, T., Chazelle, B., Dobkin, D., 2001. Matching 3D models with shape distributions. In: Proceedings of the International Conference on Shape Modeling & Applications. IEEE Computer Society, p. 154.

Paquet, E., Rioux, M., 2000. Influence of pose on 3-d shape classification. In: SAE International Conference on Digital Human Modeling for Design and Engineering, Dearborn, MI, USA.

Park, I., Kim, J., Kim, D., Yun, I., Lee, S., 2002. 3D perceptual shape descriptor: Result of exploration experiments and proposal for core experiments. In: ISO/IEC JTC1/SC29/WG11, MPEG02/M9210, Japan.

Shilane, P., Min, P., Kazhdan, M., Funkhouser, T., 2004. The princeton shape benchmark. In: Shape Modeling International, pp. 167–178.

Teh, C.-H., Chin, R.T., 1988. On image analysis by the methods of moments. IEEE Trans. Pattern Anal. Machine Intell. 10 (4), 496–513.

Tung, T., Schmitt, F., Augmented reeb graphs for content-based retrieval of 3D mesh models. In: International Conference on Shape Modeling and Applications (SMI'04), 2004. IEEE Computer Society Press, Genova, Italy, accepted for publication.

Ullmann, J.R., 1976. An algorithm for subgraph isomorphism. J. ACM 1 (23), 31–42.

Vranic, D.V., Saupe, D., 2002. Description of 3D-shape using a complex function on the sphere. In: IEEE International Conference on Multimedia and Expo (ICME 2002), Lausanne, Switzerland, 2002, pp. 177–180.

Wang, J., Li, J., Wiederhold, G., SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. IEEE Trans. Pattern Anal. Machine Intell. 23.

Zhang, C., Chen, T., 2001. Efficient feature extraction for 2D/3D objects in mesh representation. In: International Conference on Image Processing (ICIP 2001), vol. 3, Thessaloniki, Greece, pp. 935–938.

Zhou, L., Kambhamettu, C., 2002. Representing and recognizing complete set of geons using extended superquadrics. In: ICPR02, pp. 713–718.