

Generalizations of Angular Radial Transform for 2D and 3D Shape Retrieval

Julien Ricard, David Coeurjolly, and Atilla Baskurt

Laboratoire LIRIS, CNRS FRE 2672

43 Bd du 11 novembre 1918, Villeurbanne F-69622, France

{jricard, dcoeurjo, abskurt}@liris.cnrs.fr

Abstract

The Angular Radial transform (ART) is a moment-based image description method adopted in MPEG-7 as a 2D region-based shape descriptor. Efficiency and robustness were demonstrated on binary image. This paper proposes the generalization of the ART to describe two dimensional color images and three dimensional models.

First, the ART recommended by the MPEG-7 standard is only limited to binary images and is not robust to perspective deformations. We propose two extensions which allow applying ART to color images and to insure robustness to all possible rotations and to perspective deformations. In other words, the descriptor is not adapted to natural color images according to the shape and the color attributes.

We also generalize the ART to index 3D models. ART is a 2D complex transform in polar coordinate and can be extend to 3D data using spherical coordinates while keeping the robustness properties. The new 3D shape descriptor called 3D ART, have same properties that the original transform: robustness to rotation, translation, noise and scaling while keeping compact size and 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. Results on large 3D databases are presented and discussed.

Index Terms

Content based retrieval, Shape descriptor, Angular Radial Transform, 3D models, Color image.

I. INTRODUCTION

Content-based image retrieval has been a topic of intensive research in recent years, and particularly the development of effective shape descriptors (SD). The MPEG-7 standard comity has proposed a region base shape descriptor, the Angular Radial Transform (ART) [1], [2]. 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, it implies fast answers during retrieval processes. In the MPEG-7 standard, the ART similarity measure is reduced to a L_1 distance between the 35 floating point values.

In the same time, the technical 3D model databases grow up since the beginning of the computed-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 [3]. This huge number of models requires a content-based mining with indexing and retrieval processes. In the framework of the Semantic-3D national project¹ 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 color images by taking into account the chrominance information and 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 color images and the 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 take into account chrominance components and 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 the ART 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.

II. 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 [4]. It gives a compact and efficient way to express pixel distribution within a 2-D object region; it can describe both connected and disconnected region shapes. The ART is a complex orthogonal unitary transform defined on a unit disk that consists of the complete orthogonal sinusoidal basis functions in polar coordinates [1], [2]. The ART coefficients, F_{nm} of order n and m , are defined by:

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

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:

$$V_{nm}(\rho, \theta) = A_m(\theta) R_n(\rho). \quad (2)$$

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:

$$\begin{cases} A_m(\theta) = \frac{1}{2\pi} \exp(jm\theta) \\ R_n(\rho) = \begin{cases} 1 & n = 0 \\ 2 \cos(\pi n \rho) & n \neq 0 \end{cases} \end{cases} \quad (3)$$

¹Semantic-3D (<http://www.semantic-3d.net>) is supported by the French Research Ministry and the RNRT (Réseau National de Recherche en Télécommunications).

Real parts of basis functions are shown in Figure 1.

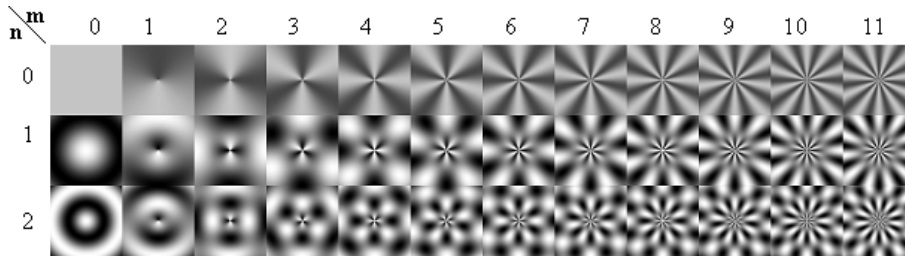


Fig. 1. Real parts of the ART basis functions.

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, twelve angular and three radial functions are used ($n < 3, m < 12$) [1], these values will be used in the rest of the paper. For scale normalization, the ART coefficients are divided by the magnitude of ART coefficient of order $n = 0, m = 0$. To decrease the descriptor size, quantification is applied to each coefficient using four bits per coefficient [1]. The distance between two shapes described by the ART descriptor is calculated using L_1 norm:

$$d_{ART}(Q, I) = \sum_{i=0}^{n \cdot m} \|ART_Q[i] - ART_I[i]\| \quad (4)$$

The subscripts Q and I represent respectively the query image and an image in the database, and ART_I is the array of ART descriptor of the image I .

Since 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 [5], and in face detection [6].

III. GENERALIZATIONS OF ART

Two generalizations of the ART on 2D color images are proposed: first, we consider all rotations and perspective deformations. Then a generalization to color images is considered.

A. Robustness to rigid deformations

The goal of this generalization is to make the ART robust to every rotation and to 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 (Fig. 2).

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 radial direction ς , the rotation angle ϕ and the perspective

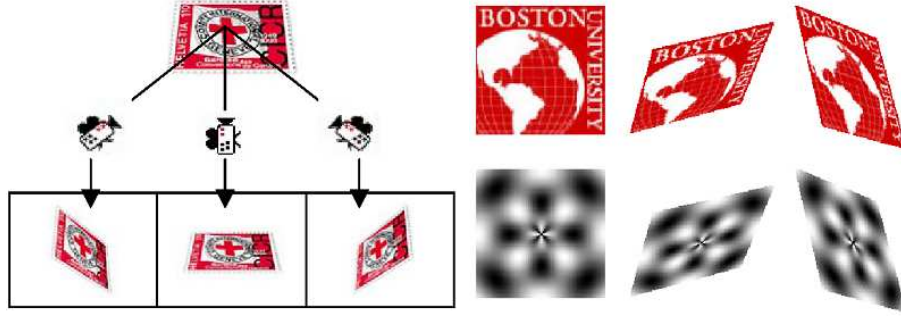


Fig. 2. Object seen according to various angles and example of projected basic functions on the support plane of the object to be identified.

coefficient p . The first two parameters define the orientation of the object plan and the third is the perspective coefficient which defined the perspective deformation. Figure 3 shows the transformation.

This transformation space is sampled for each parameter according to k_ζ , k_ϕ and k_p values. Hence we obtain a sampling of $K = k_\zeta * k_\phi * k_p$ transformations. The basis functions are deformed in 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 have a reasonable computational cost. The values, $k_\zeta = 12$, $k_\phi = 3$ 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.

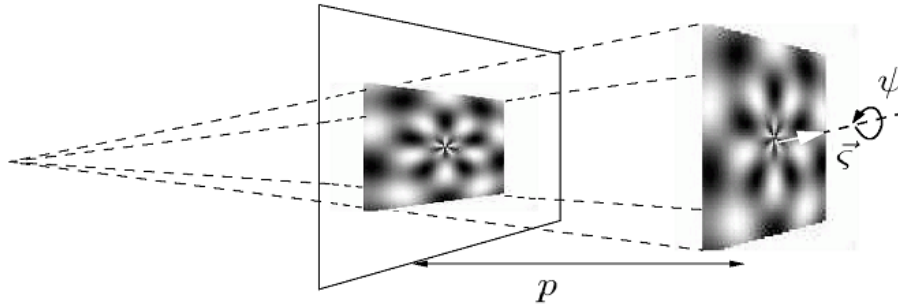


Fig. 3. The basis functions are projected on the image plan I according to k_ζ , k_ϕ and k_p to obtain the projected basis functions.

The complexity of the process is larger than that of standard ART. The complexity of 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 creates K set of basis functions with a complexity $\theta(K * n * m * N^2)$.

To make the retrieval process faster, we choose to inverse the indexation and retrieval processes. Without optimization, the indexation process computes the ART descriptor between the original object and the original basis function whereas the retrieval process computes the descriptor between the extracted object from a natural image and all the projected basis functions. In fact, the indexation process has a computation cost K times less

	Original process	Optimized process
Online	K	1
Offline	1	K

TABLE I

NUMBER OF ONLINE AND OFFLINE ART DESCRIPTOR COMPUTATION DURING THE ORIGINAL PROCESS AND THE OPTIMIZED ONE.

than the retrieval process. The retrieval is an online process and it is the longest step. It is possible to inverse these two processes and to index the object of origin on the inverse projected basis functions and an extracted object 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 (Fig. 4 and Table I). We can easily show that :

$$F_1(i', j')V_0(i', j') = F_0(i, j)V_{-1}(i, j) \quad \forall \begin{pmatrix} i' \\ j' \end{pmatrix} = T \begin{pmatrix} i \\ j \end{pmatrix} \quad (5)$$

where $F_k(i, j)$ and $V_k(i, j)$ are the image pixel and the basis function pixel (i, j) by the T^k transform. We obtain the same descriptor by indexing the object of origin on the inverse transform basis functions and an extracted object only on the origin basis functions.

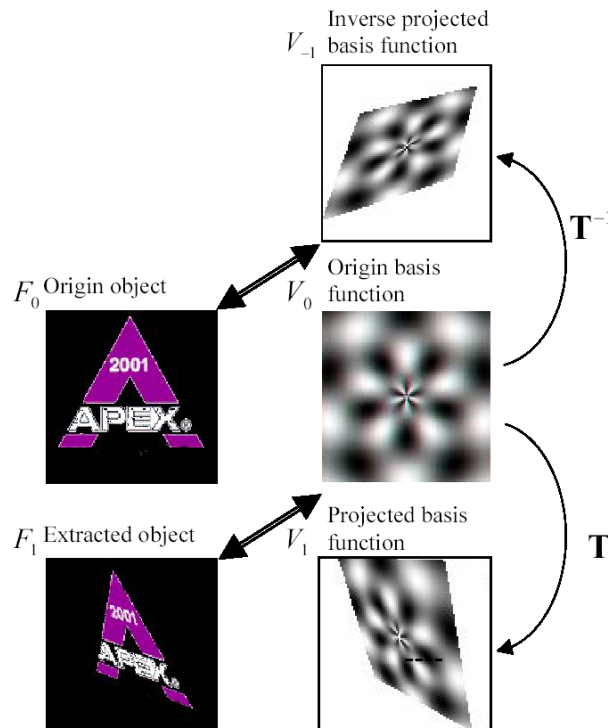


Fig. 4. Diagram of inverse indexation process.

The shape similarity distance, knowing that each object is described by $K = k_\zeta * k_\phi * k_p$ series of ART coefficients

created from the basis functions projected on K planes of projections, is achieved by computing a set of distances $d_{ART}(Q, I_j)$. For each value of j , the ART coefficients of Q , computed on the original basis functions, and those of I , computed on the j^{th} projection of the basis functions, are compared using (4). Then the shape distance between Q and I is given by:

$$d_{shape}(Q, I) = \min_{j \in K} \sum_{i=0}^{n \cdot m} \left\| ART_Q[i] - ART_I^j[i] \right\| \quad (6)$$

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

B. Color ART

In the second generalization, we make the transformation be tractable to color images. The value and the position of the dominant colors which compose the object must be taken into account in the shape retrieval system. The shape and the color of the objects are treated by two parallel studies: a study of the luminance and a study of the chrominance. We obtain two classifications of the image database which are combined to have a single one.

1) *Study of the luminance*: The basis functions of the Color ART are the same as those of the ART transform. The colored object is first represented in the perceptually uniform (L^*, a^*, b^*) color space [7]. The chrominance part of the information is not projected on the basis functions. Only the luminance component is considered to compute the ART coefficients. Note that MPEG-7 suggests the ART transform must be applied on binary objects but many systems [8], [9] used the luminance to compute the descriptor. The ART transform, applied to the luminance image of the object, gives a better result than the applied to the binary image [10]. The application of ART on the luminance component allows taking into account the internal variations of the objects (contours, holes, texture...).

2) *Study of the chrominance*: This study has the aim to classify the image database according to a color criterion. The color descriptions of the object are made using the dominant color analysis [11], [12]. The object colors are described by their dominant colors, (DC_i, p_i, σ_i) in a *Lab* color space, where DC_i is the color vector (L_i, a_i, b_i) , p_i and σ_i are the percentage and the variance corresponding to the distribution of the i^{th} dominant color. The value of DC_i are supplied directly by the segmentation process [11], [13], and the corresponding variance and percentage are computed on the object. We define a color histogram as the sum of the dominant color contributions, as follows:

$$H(x) = \sum_i \frac{p_i}{\sigma_i \sqrt{2\pi}} \exp - \frac{(x - DC_i)^2}{2\sigma_i^2} \quad (7)$$

where the value x corresponds to the bin of the histogram. The color similarity measure is computed between the histograms of all dominant colors [11], [13]. A Kullback distance is thus performed in its symmetric form [14] to measure the similarity between two generated distributions H_Q and H_I . The color distance between the query images Q and a database image I is then given by:

$$d_{color}(Q, I) = \sum_{n=1}^N \sum_{m=1}^3 (q_{nm} - i_{nm}) \log_2 \left(\frac{q_{nm} + 1}{i_{nm} + 1} \right) \quad (8)$$

where N is the number of histogram bins (256), M is the number of color components ($M = 3$ for *Lab* space), q_{nm} is the percentage of the m^{th} component of the n^{th} color in Q and i^{nm} is the percentage of the m^{th} component of the n^{th} color in I .

C. Combining features for matching

To estimate the similarity between two images, we have to evaluate the similarities between their descriptors. The color distribution and the generalized ART to projection and to rotation distribution are mixed. This transformation is called the generalized color ART (GCART). A global similarity function D is computed as a weighted sum of the similarities:

$$D = \alpha \cdot d_{color} + (1 - \alpha) d_{shape} \quad (9)$$

where α is the weight controlling the sum. It is fixed iteratively by the user according to his request or evaluated automatically by the system when the image database classes are known. Results are shown in section 5.A.

IV. 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.

A. 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 art can be divided into three different classes of methods for 3D shape description: structural approaches, multi-view approaches and 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 3D objects as the merge of these sub-parts according to adjacency relations. 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. The graphs contain sub-part information and the adjacency relations. Two distinct approaches can be considered: the surface based approaches and the 3D 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 [15]. Dorai [16] proposes to use a graph of maximal patches defined by functions of the principal curvature. On 3D technical models, the model signature graph (MSG) [17], [18] 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 [19] uses multiresolution Reeb graph. Multiresolution graph are constructed by computing a surface geodesic distance to define a Reeb graph [20] at various levels. Recently, the Augmented Reeb Graphs [21] increases the matching process. A 3D approach decomposes a

shape using 3D elementary volumetric structures called geons (geometrics ions) based on recognition by composant theory [22]. Sets of 3D volumetric primitives are used: cylinders, cubes, parallelepipeds, cone (truncated or not), ellipsoids [23]. Other interesting approaches use a set of superquadrics [24], [25] or quadratic surfaces [26].

The difficulties in the specification of a shape descriptor for 3D objects justify other categories of approaches. The multi-view approach defines a 3D shape descriptor by a set of 2D image descriptors computed on given images. This class of methods allows to describe a shape of 3D objects, to classify the objects between them, and also to retrieve 3D objects with 2D requests. To classify 3D objects between them, all the request object views are compared to the tested objects views. The keypoints of these techniques are the choice of the views and the choice of the 2D shape descriptor. A lot of 2D image shape descriptors are used to index extracted views; we can quote the CSS [1], the ART [5], ... The number of the views is a very important parameter because of the retrieval computational costs. Abbasi and Lokhtarian [27] fix the number of view to 9. Chen and Stockman [28] propose a discretization of the including sphere. This sphere is discretized into eight triangular faces. The camera is placed at the barycenter of each face to obtain the views. These methods use a principal composant analysis to align objects along the axis to reduce bad aligned views [29], [30]. Other methods create a large number of views (for example 380) and use a shape descriptor to reduce the number of views which are characteristic [16], [31].

The statistical approaches characterize the 3D model shape by calculating statistical moments [32]–[34] or by considering a distribution of the measurement of geometric primitives (which might be points, cords, triangles, tetrahedrons...) [35], [36]. A geometrical normalization of the object size and position as a 3D space is used in a preprocessing 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 [37] and ART [1], [2]. Some of these moments have been extended in 3D: 3D Fourier [38], 3D Wavelet [39], 3D Zernike [40] and a spherical harmonic (SH) decomposition, recently described by Vranic and Saupe [41], and Funkhouser et al. [42]. 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 [43], 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.

B. 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:

$$F_{nm_\theta m_\phi} = \int_0^{2\pi} \int_0^\pi \int_0^1 V_{nm_\theta m_\phi}(\rho, \theta, \phi) f(\rho, \theta, \phi) \rho d\rho d\theta d\phi \quad (10)$$

where $F_{nm_\theta m_\phi}$ is an ART coefficient of orders n , m_θ and m_ϕ , $f(\rho, \theta, \phi)$ 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:

$$V_{nm_\theta m_\phi}(\rho, \theta, \phi) = A_{m_\theta}(\theta) A_{m_\phi}(\phi) R_n(\rho) \quad (11)$$

As in 2D, the radial basis function is defined by a cosine function:

$$R_n(\rho) = \begin{cases} 1 & n = 0 \\ 2 \cos(\pi n \rho) & n \neq 0 \end{cases} \quad (12)$$

The angular basis functions are defined by complex exponential functions to achieve rotation invariance and continuity along both θ and ϕ values:

$$\begin{aligned} A_{m_\theta}(\theta) &= \frac{1}{2\pi} \exp(2jm_\theta\theta) \\ A_{m_\phi}(\phi) &= \frac{1}{2\pi} \exp(jm_\phi\phi) \end{aligned} \quad (13)$$

The values of the parameters n , m_θ and m_ϕ are trade-offs between efficiency and accuracy. The choice of the values of 3D ART parameters, n , m_θ and m_ϕ , is made by computing of the Recall values for different values of these parameters. For the technical database presented in section 4, we finally choose $n = 3$, $m_\theta = 5$ and $m_\phi = 5$. The real parts of the 3D ART BF are shown in figures 5.

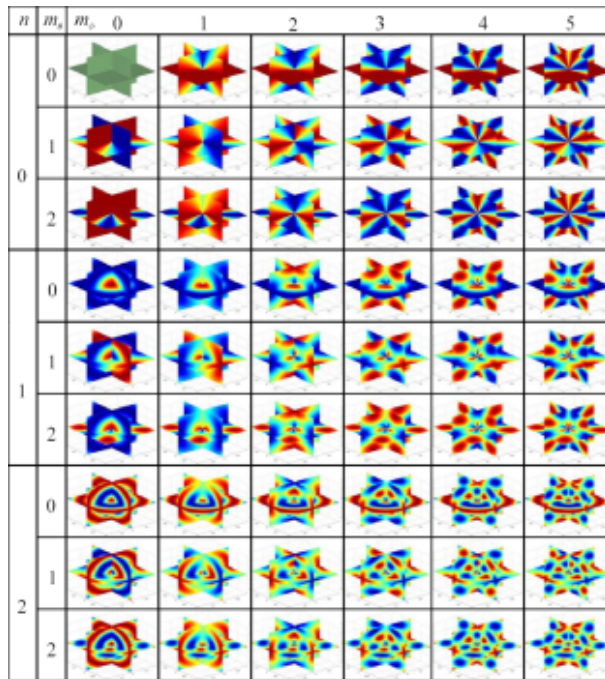


Fig. 5. Real parts of 3D ART BF.

The similarity measure is computed using a L_1 norm by the 3D ART descriptors:

$$d(Q, I) = \sum_{i=1}^{n \cdot m_\theta \cdot m_\phi} \|ART3D_Q[i] - ART3D_I[i]\| \quad (14)$$

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 of other distances could be used.

C. Indexing process

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

$$(\rho, \phi) \xrightarrow{Rot_\alpha} (\rho, \phi + \alpha) \quad (15)$$

Thus, that does not modify the norm of the function $A_{m_\theta}(\theta)$ and the ART descriptor. In 3D, unspecified rotations can not be expressed as the sum of constant values on the angular components which modify the descriptor values. However, if we consider a rotation around the z-axis, the norms of the 3D ART coefficient 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. A Principal Components Analysis (PCA) is applied to obtain the principal direction of the objects. PCA alignment does not provide a robust normalization when the three principal directions are used. Here, we only aligned the first principal direction along the z-axis and wrong alignments are limited. Fig. 6 shows the indexation process.

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

V. EXPERIMENTS

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

A. 2D generalized ART experiments

The first test compares the ART on the luminance and the generalization to projection. A test database was created, it contains 1813 images of 37 trademark images disturbed according to 49 random perspective projection with illuminating variations. The figure 8.a shows the Recall and Precision values.

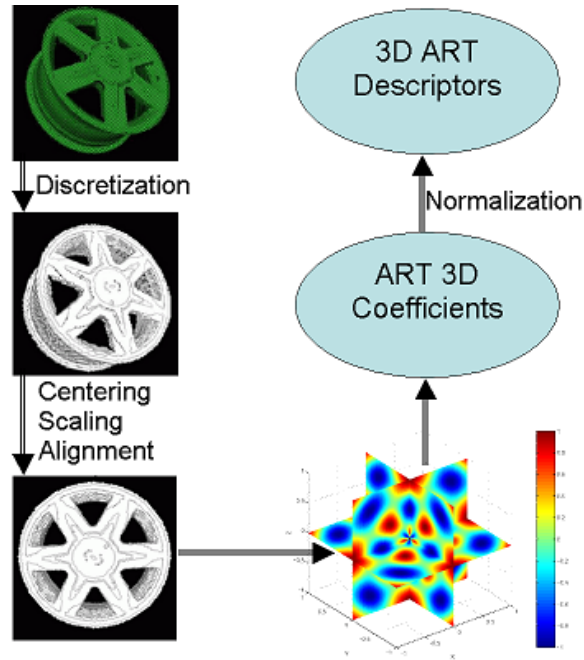


Fig. 6. Indexing process.

To evaluate the color ART, we have set up an application allowing to identify an object extracted from an image. The application can be split into two successive stages: the indexation and the retrieval steps (fig. 7).

To evaluate the properties of the GCART and the retrieval process, 50 objects were extracted from the images and we evaluate the rank where one finds the original trademark. The figure 8.b shows the Recall values for the luminance, the color and the GCART studies. The GCART study gives the original trademark at the first rank in 55%, against 38% for the luminance and 6% for the color. At the rank 10, the original trademark is found in 95% of the cases, whereas the luminance and the color study have found the original object, respectively in 65% and 41% of the cases.

B. 3D ART experiments

1) *3D model database test*: The 3D experiments are made using two 3D model databases: the Princeton Shape Benchmark [44] and a Renault database. The figures 9 and 10 show examples of 3D models from 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 shapes but with the same semantic (Fig. 11).

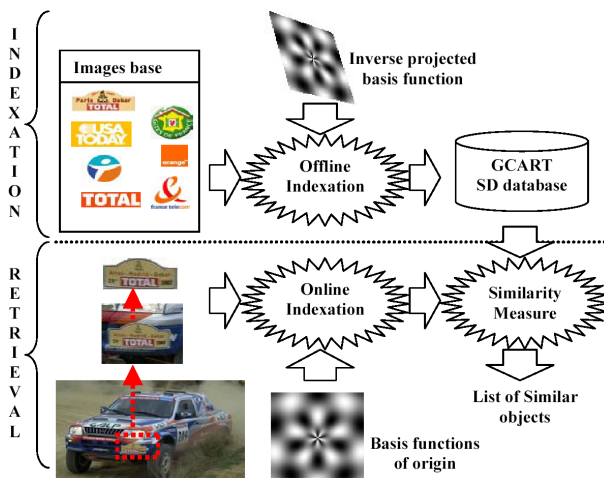


Fig. 7. General diagram of the application.

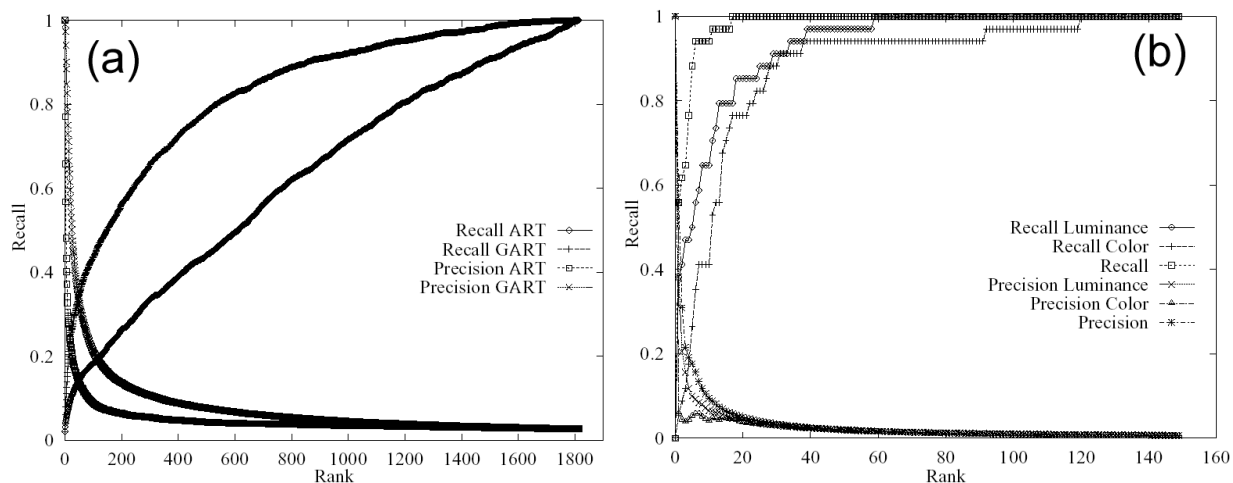


Fig. 8. Recall and precision values: (a) ART and generalized to perspective projection ART, (b) luminance, color and CART approach.

The Renault database is a technical database which contains mechanical models. In the framework of SEMANTIC 3D and in partnership 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 vehicles diversities and 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... All the database objects can not be classified because the database has not got 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 figure 11 and 12.

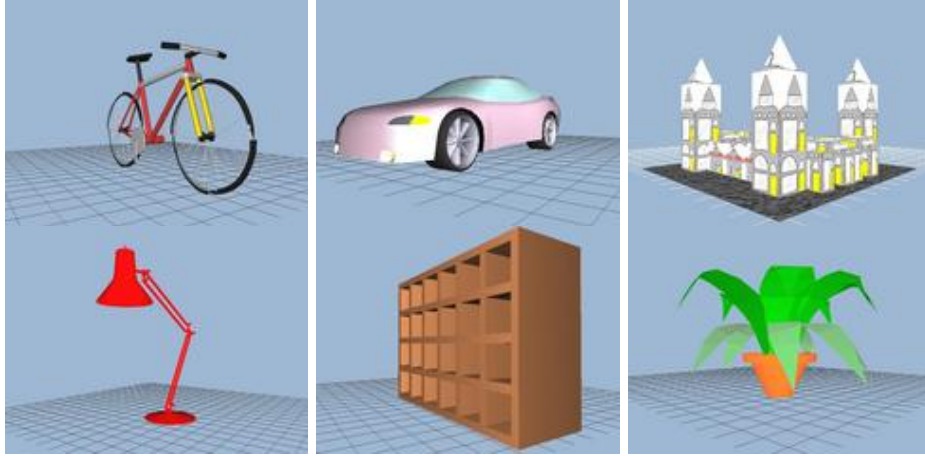


Fig. 9. Examples of Princeton Shape Benchmark 3D models.

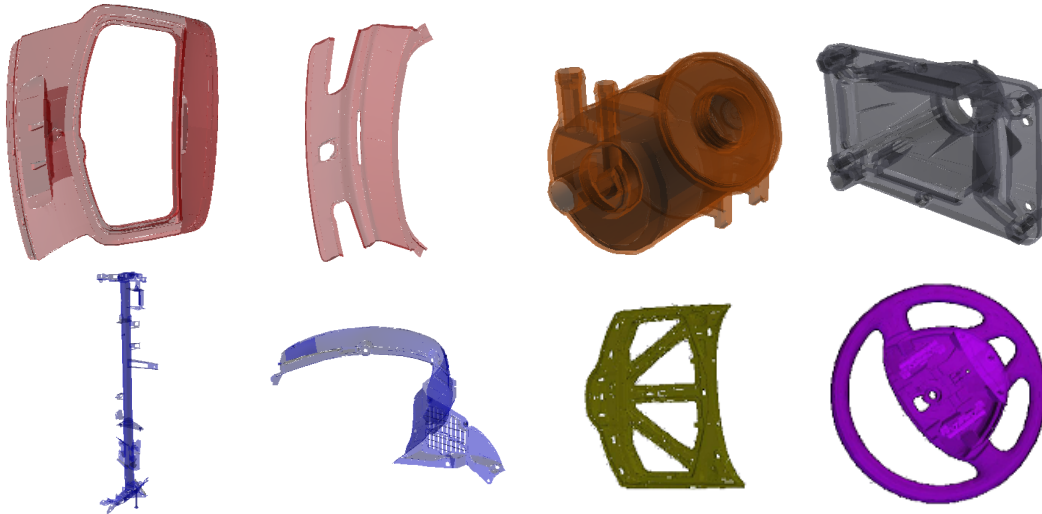


Fig. 10. Examples of Renault database 3D models.

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. 13.a shows that the best results are obtained for $n = 3$ and $m_\theta = m_\phi = 5$. Fig. 13.b presents the same experiment with different sizes of the discretization 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 [43] for the SH computation.

3) *Robustness*: To evaluate the robustness of the process, we distort a 3D object according to scaling, rotation, translation and noise. Table II 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 with the original one. The translation has no effect on the distance, because the pre-processing step centers the objects. For the same

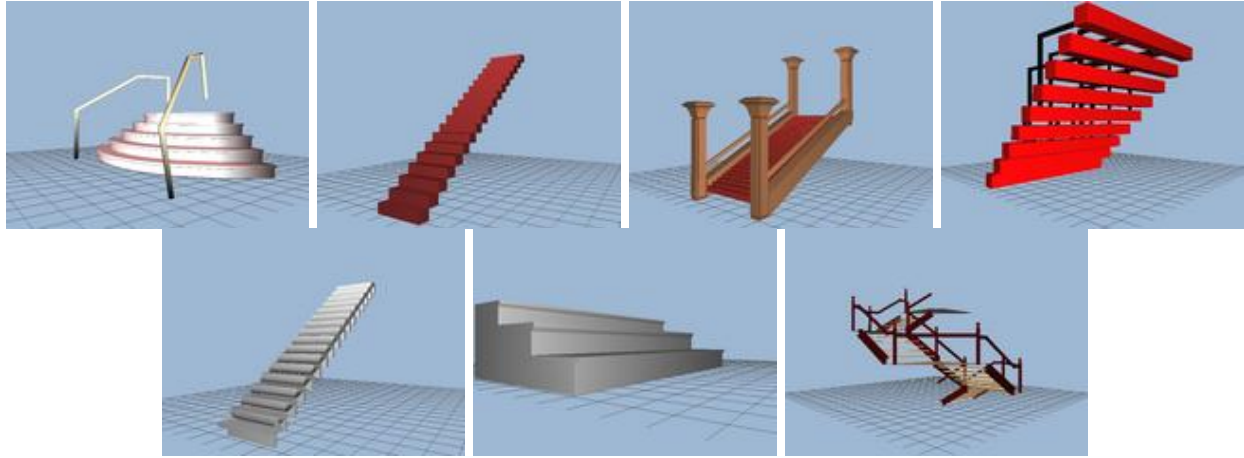


Fig. 11. Example of Princeton Shape Benchmark class: staircase.



Fig. 12. Example of Renault database class: seat belt part.

reasons, 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. 14 shows distorted objects by the noise distortions.

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. The figures 15.a and 15.b show the recall values for SH and 3D ART descriptors for the two databases. On the Princeton database (Fig. 15.a), the SH method gives a better description than the ART. The results on the Renault database are similar with the two methods (Fig. 15.b). ART description gives better results when the objects of a same class are similar.

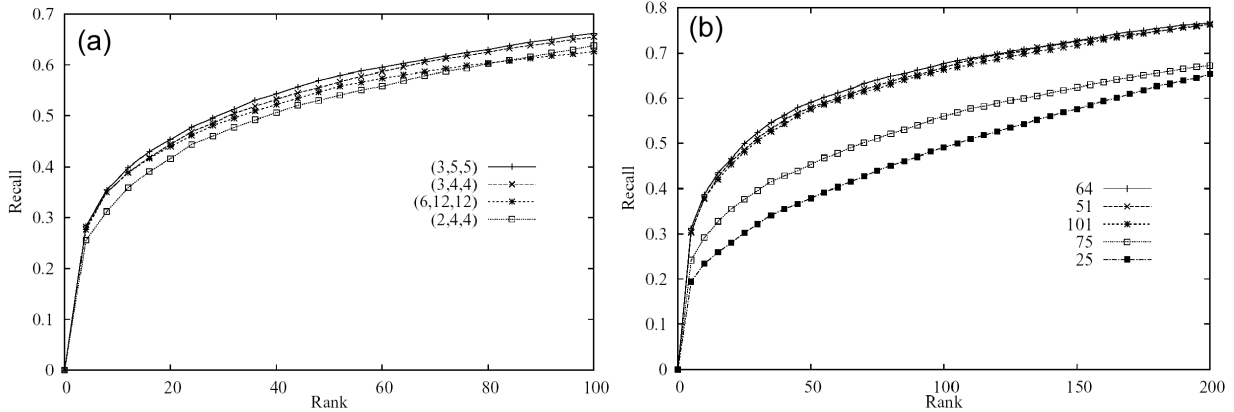


Fig. 13. Recall values to set up parameters.

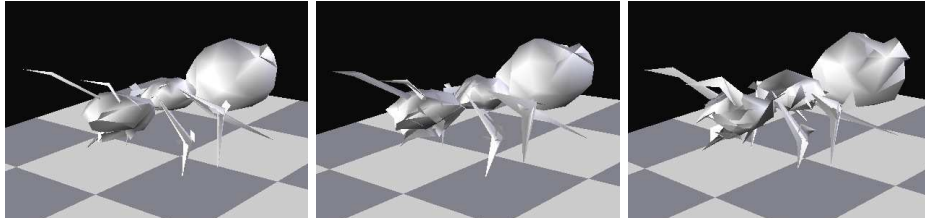


Fig. 14. Example of noise distortions for three distance values: 0%, 5% and 10%

5) *Complexity*: The computational cost and the size of the descriptors are significant comparison criteria (Table III). 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 the smallest possible.

VI. CONCLUSION

In the first part of this paper, we have proposed an extension of the 2D region based shape descriptor ART to color and deformed images, and to 3D models. The generalizations of the ART, to perspective projection and to

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

TABLE II

DISTANCE OBTAINED FOR SEVERAL DISTORTIONS.

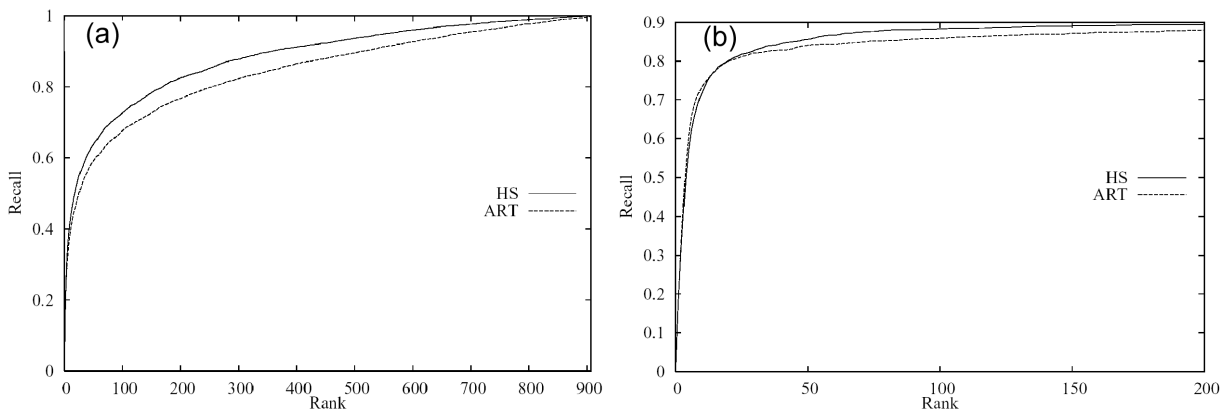


Fig. 15. Recall values on Princeton and Renault databases.

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

TABLE III

SIZE (IN FLOATING NUMBERS) AND INDEXING TIME (IN SECONDS) COMPARISON BETWEEN 3D ART AND SPHERICAL HARMONIC REPRESENTATION.

color, increase the numbers of ART uses and the definition domains while keeping the discriminating capacities. The optimized process makes possible to have a light online process and a quick answer for color image content-based retrieval. In the second part of this work, we have presented the generalization of the ART to 3D shape description. The proposed descriptor is robust to translations, scaling, multi representation (remeshing, weak distortions), noises and 3D rotations. It fulfills the requirements induced by the technical model database analysis: robustness and accuracy of the indexing, and fast retrieval processes and similarity computation index.

As a future work, we plan to investigate the possibilities to build a 2D/3D retrieval from 3D ART.

ACKNOWLEDGMENT

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

- [1] S. Jeannin, "Mpeg-7 Visual part of eXperimentation Model Version 9.0," in *ISO/IEC JTC1/SC29/WG11/N3914, 55th Mpeg Meeting*, Pisa, Italia, Jan. 2001.
- [2] W.-Y. Kim and Y.-S. Kim, "A new region-based shape descriptor," in *TR 15-01*, Pisa, Dec. 1999.

- [3] D. McWherter, M. Peabody, and W. C. Regli, "Clustering techniques for databases of CAD models," Drexel University, Tech. Rep., Sept. 2001.
- [4] M. Bober, "Mpeg-7 visual shape descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 1(6), June 2001.
- [5] D.-Y. Chen and M. Ouhyoung, "A 3d model alignment and retrieval system," in *International Computer Symposium (ICS 2002)*, Hualien, R.O.C, Dec. 2002.
- [6] J. Fang and G. Qiu, "Human face detection using angular radial transform and support vector machines," in *International Conference on Image Processing (ICIP 2003)*, 2003, pp. I: 669–672.
- [7] K. Nassau, *Color for Science, Art, and Technology*, E. Science, Ed., Amsterdam, 1998.
- [8] J. Wang, J. Li, and G. Wiederhold, "SIMPLiCity: Semantics-sensitive integrated matching for picture libraries," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, 2001.
- [9] J. Laaksonen, J. Koskela, S. P. Laakso, and E. Oja, "PicSOM: Content-based image retrieval with self-organizing maps," *Pattern Recognition Letters*, vol. 21(13-14), Dec. 2000.
- [10] M. Akcay, A. Baskurt, and B. Sankur, "Measuring similarity between color image regions," in *EUSIPCO*, vol. 1, European Signal Processing Conference (EUSIPCO'02), Sept. 2002, pp. 115–118.
- [11] K. Idrissi, J. Ricard, A. Anwander, and A. Baskurt, "An image retrieval system based on local and global color descriptors," *the 2nd IEEE Pacific-Rim Conference on Multimedia*, pp. 55–62, Oct. October 2001.
- [12] K. Idrissi, G. Lavou, J. Ricard, and A. Baskurt, "Object of interest-based visual navigation, retrieval, and semantic content identification system," *Computer Vision and Image Understanding*, vol. 94(1-3), pp. 271–294, 2004.
- [13] K. Idrissi, J. Ricard, and A. Baskurt, "An objective performance evaluation tool for color based image retrieval systems," *International Conference on Image Processing (ICIP 2002)*, Sept. 2002.
- [14] S. Kullback, *Information theory and statistics*, J. Wiley and Sons, Eds., New York, 1959.
- [15] J. R. Ullmann, "An algorithm for subgraph isomorphism," *Journal of the ACM*, vol. 1(23), pp. 31–42, Jan. 1976.
- [16] C. Dorai and A. K. Jain, "Shape spectrum based view grouping and matching of 3d free-form objects," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19(10), pp. 1139–1146, 1997.
- [17] D. McWherter, M. Peabody, W. C. Regli, and 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)*, 2001.
- [18] M. Peabody and W. C. Regli, "Clustering techniques for databases of cad models," Hewlett-Packard Research, Tech. Rep., Sept. 2001.
- [19] M. Hilaga, Y. Shinagawa, T. Kohmura, and T. L. Kunii, "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, 2001, pp. 203–212.
- [20] G. Reeb, "Sur les points singuliers d'une forme de pfaff completement integrable ou d'une fonction numerique [on the singular points of a completely integrable pfaff form or of a numerical function]," *Comptes Rendus de Acadmie des Sciences de Paris*, vol. 222), pp. 847–849, 1946.
- [21] T. Tung and F. Schmitt, "Augmented reeb graphs for content-based retrieval of 3d mesh models," in *(Accepted) International Conference on Shape Modeling and Applications (SMI'04)*. Genova, Italy: IEEE Computer Society Press, 2004.
- [22] I. Biederman, "Recognition by components: A theory of human image understanding," *Psychological Review*, vol. 94, pp. 115–147, 1987.
- [23] P. Irani and C. Ware, "Diagrams based on structural object perception," in *Proceedings of the working conference on Advanced visual interfaces*. ACM Press, 2000, pp. 61–67.
- [24] L. Zhou and C. Kambhampettu, "Representing and recognizing complete set of geons using extended superquadrics," in *ICPR02*, 2002, pp. III: 713–718.
- [25] L. Chevalier, F. Jaillet, and A. Baskurt, "Segmentation and superquadric modeling of 3d objects," in *The 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2003 (WSCG)*, vol. 11(1), Plzen, czech republic, Feb. 2003.
- [26] I. Park, J. Kim, D. Kim, I. Yun, and S. Lee, "3d perceptual shape descriptor: Result of exploration experiments and proposal for core experiments," in *ISO/IEC JTC1/SC29/WG11, MPEG02/M9210*, Japon, Dec. 2002.
- [27] S. Abbasi and F. Mokhtarian, "Affine-similar shape retrieval: Application to multi-view 3-d object recognition," in *IEEE Transactions on Image Processing*, vol. 10(1), Japon, Dec. 2001, pp. 131–139.
- [28] J.-L. Chen and G. Stockman, "3d free-form object recognition using indexing by contour feature," *Computer Vision and Image Understanding*, vol. 71(3), pp. 334–355, 1998.

- [29] D.-Y. Chen, X.-P. Tian, Y.-T. Shen, and M. Ouhyoung, "On visual similarity based 3d model retrieval," in *Computer Graphics Forum (EUROGRAPHICS'03)*, vol. 22(3), Barcelona, Spain, Sept. 2003, pp. 223–232.
- [30] S. Nayar, S. A. Nene, and H. Murase, "Real-time object recognition system," in *International Conference on Robotics and Automation*, 1996.
- [31] T. F. Ansary, M. Daoudi, and J. Vandeborre, "A bayesian approach for 3d models retrieval based on characteristic views," in *International Conference on Pattern Recognition (ICPR 2004)*, vol. 11(1), Cambridge, England, Aug. 2004.
- [32] T. Murao, "Descriptors of polyhedral data for 3d-shape similarity search," in *Proposal P177, MPEG-7 Proposal Evaluation Meeting*, UK, Feb. 1999.
- [33] M. Elad, A. Tal, and S. Ar, "Directed search in a 3d objects database using svm," Hewlett-Packard Research, Tech. Rep., 2000.
- [34] C. Zhang and T. Chen, "Efficient feature extraction for 2d/3d objects in mesh representation," in *International Conference on Image Processing (ICIP 2001)*, Thessaloniki, Greece, 2001, p. 935.
- [35] E. Paquet, M. Rioux, A. Murching, T. Naveen, and A. Tabatabai, "Description of shape information for 2-d and 3-d objects," in *Signal Processing: Image Communications*, Sept. 2000, p. 103.
- [36] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Matching 3d models with shape distributions," in *International Conference on Shape Modeling and Applications*, May 2001.
- [37] C.-H. Teh and R. T. Chin., "On image analysis by the methods of moments," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10(4), pp. 496–513, 1988.
- [38] M. Elad, A. Tal, and S. Ar, "Content based retrieval of vrml objects - an iterative and interactive approach," in *The Sixth Eurographics Workshop in Multimedia*, 2001, pp. 97–108.
- [39] E. Paquet and M. Rioux, "Influence of pose on 3-d shape classification," in *SAE International Conference on Digital Human Modeling for Design and Engineering*, Dearborn, MI, USA, June 2000.
- [40] N. Canterakis, "3d zemike moments and zernike affine invariants for 3d image analysis," in *11th Scandinavian Conf. on Image Analysis*, 1999.
- [41] D. V. Vranic and D. Saupé, "Description of 3d-shape using a complex function on the sphere," in *IEEE International Conference on Multimedia and Expo (ICME 2002)*, Lausanne, Switzerland, August 2002.
- [42] 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*, vol. 22(1), pp. 83–105, Jan. 2003.
- [43] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3d shape descriptors," *Symposium on Geometry Processing*, June 2003.
- [44] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The princeton shape benchmark," in *Shape Modeling International*, June 2004.