FCT Fundação para a Ciência e a Tecnologia

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

PROJECTOS DE INVESTIGAÇÃO CIENTÍFICA E DESENVOLVIMENTO TECNOLÓGICO

DecorAR

Survey on 3D Shape Descriptors

Autores:

Lisha Zhang
Manuel João da Fonseca
Alfredo Ferreira

Acrónimo: **DecorAR**

Título: **Decoração Aumentada: Combinando Realidade Aumentada e Recuperação de Objectos 3D para a Decoração de Interiores**

Referência: **POSC/EIA/59938/2004**

União Europeia - Fundos Estruturais

Governo da República Portuguesa

Table of Contents

1. Introduction

The advancement of modelling, digitizing and visualizing techniques for 3D shapes has led to an increasing amount of 3D models, both on the internet and in domain specific databases. Determining the similarity between 3D shapes is a fundamental task in shape-based recognition, retrieval, clustering, and classification. There have been quite a few experimental search engines, such as the 3D model search engine at Princeton University[2], the 3D model similarity search engine at the University of Konstanz[1], the 3D model retrieval system at the National Taiwan University[3], and the 3D retrieval engine at Utrecht University[4].
Recently, a lot of researchers have investigated the specific problem of content based 3D shape retrieval. Iyer et al.[6] provide an extensive overview of 3D shape searching techniques. Atmosukarto and Naval[7] describe a number of 3D model retrieval systems and methods, but do not provide a categorization and evaluation.
The task of content-based retrieval research is to develop search engines that would allow users to perform a query by similarity of content. The fundamental ingredient of a retrieval system is shape matching, which is the process of determining how similar two shapes are[9]. Unfortunately, there are some difficulties for 3D shape matching ubiquitously in most of correlative shape retrieval applications. 3D models are not easily retrieved like text documents, but content based 3D shape retrieval methods that use shape properties of the 3D models to search for similar models usually perform better than text based methods[5]. On the other hand, although 2D shape matching methods have been improved quite well[8], most 2D methods, unfortunately, do not generalize directly to 3D model matching.
There are two main categories of approaches for shape matching: (1) matching by feature correspondences; (2) matching by global descriptors. The general idea of the former approach is to compute multiple local shape features for every object and to assess the similarity of any pair of objects by optimizing a cost function determined by the optimal set of feature correspondences at the optimal relative transformation. On the other hand, the global descriptor-based paradigm[10], which is considered as a mapping from the space of 3D objects to some finite-dimensional vector space, is the more common approach to shape matching. The vector encodes the information about the object’s shape by storing a vector of numerical attributes for each 3D object in the database as a representation, hereby, allow fast and reliable similarity searches. Such representations are named shape descriptors, which usually are high-dimensional vectors. When a query is presented, the retrieval engine calculates its descriptor(s) and compares it to all of the stored descriptors by a distance function. After the distance function measures the dissimilarity between the query and all of the objects in the database, the engine sorts database objects in terms of increasing distance values. Therefore, in order to improve the efficiency and effectiveness of 3D shape recognition and retrieval engines, appropriate 3D shape descriptors should be constructed, and the corresponding matching strategies are a crucial point in 3D object retrieval, especially when dealing with a large object database.
This survey reviews some of the most popular 3D shape descriptors for 3D object classification and retrieval. We analyze the important and fundamental problems on feature representation and spatial partition, and engage in the classification and comparison of these shape descriptors. This survey also chooses several appropriate 3D shape descriptors for shape recognition.
In the following section we introduce some important theoretical context of 3D shape descriptors. In section 3 we classify the most commonly used 3D shape descriptors, describing and comparing them with respect to the 3D shape categories they belong to. Finally in section 4,

we present some conclusions, describe the 3D objects involved in our work, and discuss and compare the descriptors we choose for our 3D shape classification.

2. Theoretical context

3D descriptor research has concentrated exclusively on shape, as given by the object's surface or its interior, rather than the attributes like color and texture that are commonly used in 2D image recognition and retrieval. The main reason is probably the fact that the most information about the similarity between 3D objects is beared and congregated by their shapes, and furthermore, color and texture information is not always guaranteed to be provided during recognition or retrieval processes. Thus, many researches focus on designing 3D shape descriptors, especially depending on surface information of 3D objects, which results in a variety of representation formats. For instant, a 3D object is usually represented by a collection of parameterized surface patches, such as implicit surfaces, superquadrics, NURBS(non-uniform rational B-splines) surfaces, or using constructive solid geometry techniques, such as voxel data, ray-based data, and point clouds. However, the most popular format of representation is the polygonal mesh(usually triangular mesh), which arises in virtual reality, entertainment, and web applications. A certain application demands a specific representation more suitable for application dependent tasks, while, there is always a possible mapping to switch one format of representation to another.

In our opinion, shape representation is not simply a direct process of feature extraction from 3D objects. Actually, it often consists of some preprocesses, for example, original feature representation, spatial partition, and pose normalization.

2.1 Original feature representation

Much work has been done on 3D pose determination and shape recognition on polyhedral objects yielding good results. The varieties of relative methods of both 3D pose extraction and shape recognition are based on different fundamental features of 3D objects, which we call original feature representations. Most of these features are implicated in the surfaces of 3D shapes. However, the information on object surface itself is usually not used, especially in the pose determination process. In practice, a large amount of methods adopt local features, such as the object boundaries, creases, limbs, 3D curves, surface patches, and so on.

Furthermore, besides polyhedral objects, the smooth objects, which are usually not piecewise planar but rather smooth and continuous in nature, as curved objects, are also universal in real world and significantly more difficult to express. Accordingly, there have been a great amount of original feature representation methods for 3D objects.

In the past two decades, there have been many feature representations based on local features of object surface. Faugeras[43][44] proposed a 3D object recognition algorithm based on geometrical matching between primitive surfaces. It was actually implemented as planar faces although quadric surface algorithms are presented as well. Bhanu[45] uses planar polygonal faces to represent the object and later match the object shape with a relaxation-based scheme(so-called stochastic face labelling). Stein and Medioni[46] use small surface patches(splashes) and 3D curves corresponding to depth or orientation discontinuities to represent the object for recognition and pose normalization. Bolle and Cooper[47] present the approach of collected surfaces to 3D pose normalization, and the surface of an object is modelled as a collection of planar, cylindrical and spherical patches. Bolles et al.[48] developed 3DPO system based on matching of several features or feature clusters involving object-specific features such as circular arc of a specific radius and edges. Nevatia and Binford[49] use generalized cylinders to match curved objects. Fan et al.[50], use jump boundaries, creases, limbs to match and locate 3D objects. Recently, aiming at curved object recognition, Ponce and Kreigman[51] employ a

monocular intensity image, considering the image contours as the basis for recognition and location of the object, under the assumption that the contour equations are parametrically known.
Horn[26] uses EGI to represent object with surface information, which avoid the more difficult problem of local feature matching by directly extracting the object surface area distribution with the surface normal.
Kang and Ikeuchi[27] introduced Complex EGI(CEGI) which also employed global surface features to represent object, including smooth and continuous objects, with surface area as the magnitude of the feature weight, additionally with the normal distance of the surface as the phase of the weight.

2.2 Object spatial partition

Feature representation is strongly related to the partition methods of the spheres on the 3D objects. The feature vector achieved by a particular definition of partitions is possibly quite better than that of another partition definition in vector sizes, information redundancy and robustness. Additionally, shape sphere partitioning is usually integrated into the feature extraction processes and can not be divided clearly. The methods for object spatial partition fall mainly into two types: basic models and variant models.

2.2.1 Basic models

Among all the object spatial partition methods, there are several basic partition models that are comparably simple in structure, which can be usually modified and improved to construct some complex variations. These basic models chiefly include shell model, sector model, ray-based model and voxel-based model.
(1) Shell model
In shell model[22], the 3D space is decomposed into concentric shells around the center point. This representation is particularly independent from the rotation of the objects, i.e. any rotation of an object around the center point of the model results in the same histogram.
(2) Sector model
In sector model[22], the 3D space is decomposed into sectors that emerge from the center point of the model. The sectors are defined as follows: distribute the desired number of points uniformly on the surface of a sphere, which usually use the vertices of regular polyhedrons and their recursive refinements. This model is invariant against scaling.
(3) Ray-based model
Ray-based model is always a popular object spatial partition method. For a normalized object I in the canonical coordinate frame, define a unit sphere S^2 with the center in the origin (i.e., center of the sphere coincides with the center of mass of the model). Further, one defines the function $r(u)$

$$r : S^2 \to R$$
$$u \mapsto \max\{r \geq 0 \mid ru \in I \cup \{0\}\},$$

where 0 is the origin. This function $r(u)$ measures the extent of the object in directions given by $u \in S^2$. Similarly, one may consider a rendered perspective projection of the object on an enclosing sphere. Thus, the function $r(u)$ is as follows:

$$r : S^2 \to R$$
$$r(u) = x(u) + iy(u)$$
$$x : S^2 \to [0,+\infty) \in R, \quad y : S^2 \to [0,1] \in R,$$

where i is the imaginary unit. The function $x(u)$ measures the extent of the object from the origin 0 in directions given by $u \in S^2$

$$x(u) = \max\{x \geq 0 \mid xu \in I \cup \{0\}\}.$$

The imaginary part of $r(u)$ is defined as follows

$$y(u) = \begin{cases} 0, & if\ x(u) = 0 \\ u \cdot n(u), & otherwise \end{cases}$$

where $n(u)$ is the normal vector of the mesh at the point $ux(u)(x(u) \neq 0)$. The function $y(u)$ can also be described as a rendered perspective projection of the model on an enclosing sphere.
In this ray-based partition, a number of samples $x(u)$ can be used directly as a feature vector in the spatial domain. However, this feature vector is sensitive to small perturbations of the object. Vranic et al.[13] improved the robustness of the feature vector by taking samples of the spherical function a number of samples $x(u)$ at many points, but characterizing the map by just a few coefficients in the spectral domain, such as spherical harmonics, the combination of spherical harmonics function(SHF) and fast Fourier transform(FFT).
(4) Voxel-based model
A voxel, which is a portmanteau of the words *volumetric* and *pixel*, is a volume element, representing a value on a regular grid in three dimensional space. This is analogous to a pixel, which represents 2D image data. Voxels are frequently used in the visualisation and analysis of medical and scientific data. As with pixels, voxels themselves typically do not contain their position in space(their coordinates), but rather, it is inferred based on their position relative to other voxels, i.e. their position in the data structure that makes up a single volume image.
3D voxel data can be represented as a collection of spherical functions $f_r(\theta,\varphi)$, where r corresponds to the distance from the origin of the voxel grid and (θ,φ) to spherical coordinates. The binary function is sampled for a sufficient number of radii $r = 1,\cdots R$ and angles (θ,φ).
The motive for the voxelization is to achieve a better robustness w.r.t variances of the polygonal surface, but it is considered that many fine details are lost in the voxel grid[33].

2.2.2 Variant models

The variant models of object spatial partition are relatively complicated in dividing the spheres of a certain 3D objects, which are usually improved according to some particular basic partition models. In this section, we introduce some variant forms based on the basic shape models mentioned in 2.2.1, which are spiderweb model on the basis of shell model and sector model, ray-based spherical harmonics and voxel-based variants.
(1) Spiderweb model
Spiderweb model[22] is a combined model representing more detailed information than pure shell models and pure sector models. A simple combination of two fine-grained 3D decompositions results in high dimensionality. However, since the resolution of the space decomposition is a parameter in any case, the number of dimensions may easily be adapted to the particular application.
(2) Ray-based spherical harmonics
Ray-based spherical harmonic 1 (RH1)[33] is a ray-based partition method. It defines a function on each sphere using the values of the intersection points between the polygonal mesh and the casting rays from the origin uniformly.
In [33], the RH1 is described as follows. Firstly, it casts rays from the origin in many directions $u(\theta,\varphi)$, find all points of intersection with the polygonal mesh, and define several functions on the sphere using the intersection points. Definition of function values is depicted in Figure 1. Secondly, let a and b be rays (cast from the origin O) intersecting the mesh at three and one points, respectively. The distances from the intersection points to the origin are a_0, a_1, a_2, and b_3. Let f_1, f_2 and f_3 be the functions on the spheres S_1, S_2 and S_3, respectively. Finally, for each

intersection point we determine the closest sphere and set the corresponding value of the function on that sphere. In the given example, we set $f_1(a) = a_1$, $f_2(a) = a_2$, $f_3(a) = 0$, $f_1(b) = 0$, $f_2(b) = 0$, and $f_3(b) = b_3$. If two intersection points lying on the same directional vector are closest to the same sphere, then the longer distance determines the function value ($a_1 > a_0$). In practice, Vranic took R concentric spheres to define the functions and 16384 directional vectors u. Centers of all spheres lie at the origin. Radii of the spheres take values t/R, $2t/R$,…, t, where t is an empirically determined constant (usually set t =8). Using the constant value of parameter t rather than the radius of hounding sphere increases robustness w.r.t. outliers. Later, Vranic proposed RH2 [33] which is similar to RH1 except for defining several functions on concentric spheres instead of a single one.

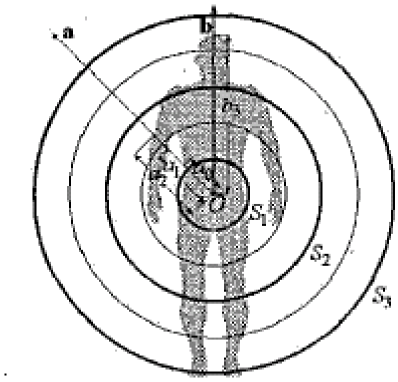

Figure 1 RH1spatial partition

(3) Voxel-based variants

(3.1) In the voxel-based variant model mentioned by Kang and Ikeuchi [27], the sphere of a 3D object is discretized into 240 sampling view directions located at the center of each face of the two-frequency dodecahedron (tessellated pentakis dodecahedron). The normal direction space is discretized into 240 cells as well.

(3.2) Zaharia and Prêteux [32] divide the enclosing sphere of a 3D object into a set of planes by uniformly sampling the spherical angle coordinates, which leads to partitions of the unit sphere into "meridians" and "parallels", as shown in Figure 2.

A plane $\Pi \in \Re^3$ is uniquely defined by a triplet (s,θ,φ), where $s \geq 0$ denotes the distance from the origin of the coordinate system to plane Π, and $\theta \in [0,2\pi)$ and $\varphi \in [-\pi/2,\pi/2)$ respectively denote the two angles(azimuth and elevation) associated with the spherical representation of the plane's unit length normal vector n.

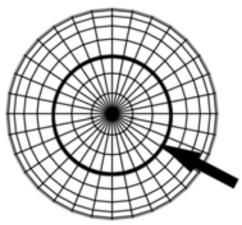

Figure 2 Voxel-based variants partition

(3.3) Regular polyhedron partition

A regular polyhedron partition[32] is obtained by projecting the vertices of any regular polyhedron(such as octahedron) onto the unit sphere, as seen in Figure 3. The following figure illustrates the nice behavior of the partition cells in grey(as figure 3(a)) being one-to-one mapped into corresponding cells(as figure 3(b)), associated with a different principal component analysis (PCA) coordinate system.

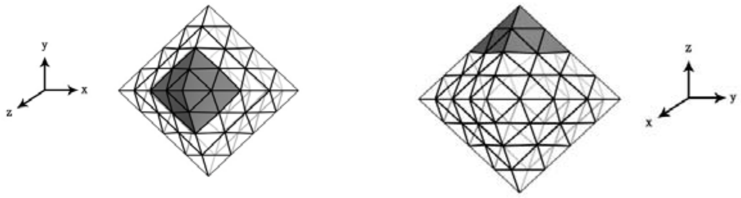

Figure 3 Regular polyhedron partition
(a) Before mapping (b) After mapping

The advantage of this partition is to make some basic PCA coordinate systems(such as generation configurations(GCs)) equivalent, in the sense that there exists a one-to-one mapping between them. In addition, multiple granularity levels can be obtained for such partitions, by recursively subdividing each of the polyhedral faces. It is possible for construction of other invariant partitions, but researchers always retain uniquely the octahedron-based partition for reasons of simplicity.

(3.4) Voxelization with bounding cube

The bounding cube(BC) of a 3D model is defined to be the tightest cube in the canonical coordinate frame that encloses the model, with the center in the origin and the edges parallel to the coordinate axes. As described by Vranic and Saupe[36], after determining the BC, one can perform voxelization in the following manner: subdivide the BC into N^3 (N is a power of 2) equal sized cubes and calculate the proportion of the total surface area of the mesh inside each of the new cubes(cells). The cell with the attributed value is regarded as the voxel at the given position. Obviously, of all voxels inside BC the fraction having values greater than zero decreases with increasing N. Thus, a suitable way of storing a voxel-based feature vector is an octree structure and then an efficient hierarchical feature representation is constructed.

(3.5) BF and IDF

Dutagaci et al. [37] suggest rendering the mesh representation of the object in a 3D voxel grid of size $N \times N \times N$, such that the object's center of mass coincides with the center of the 3D grid. In this approach, the object center, x_{center}, is calculated from the triangular mesh as follows:

$$x_{center} = \frac{1}{N_t}\sum_t A_t x_t \ ,$$

where A_t is the area and x_t is the center of mass of triangle t, and N_t is the number of triangles in the object. Then the object is scaled so that the maximum distance from the center of mass to the surface is $N/2$. Finally, a 3D binary function(BF) $v(x)$ on the voxel grid is obtained such that $v(x)$ is 1 if $x=[x_{1,}x_2, x_3]$ is on the object surface, and is 0 otherwise.

Another voxel representation, $v_d(x)$, called inverse distance function(IDF), is a function of 3D distance transform $d(x)$:

$$v_d(x) = \frac{1}{d(x)+1}$$

where $d(x)$ is the minimum L_1 distance of x to the object surface, and the 3D function $v_d(x)$ is equal to 1 on the object surface and decreases as the one moves away from the object surface. An example of BF and IDF can be seen in figure 4.

Since $v_d(x)$ decays rapidly to zero towards the corners of the bounding box, one can assume that the range of $v_d(x)$ does not get affected significantly by rotation. Dutagaci et al. [37] demonstrates in the experiment that in fact the features derived from IDF representation are almost invariant to rotation.

IDF is advantageous in that it fills the 3D space so that at any cross section, either planar or spherical, one has more information about the shape content. IDF also provides spatial smoothing so that high-frequency components due to sharp shape details are reduced. Moreover, IDF concentrates the spectral energy at the center while binary function supports the larger frequency. However, the IDF representation is not totally rotation invariant due to the distance transform values at the corners of the bounding box. Figure 4 shows cross sections of BF and IDF of a voxelized object.

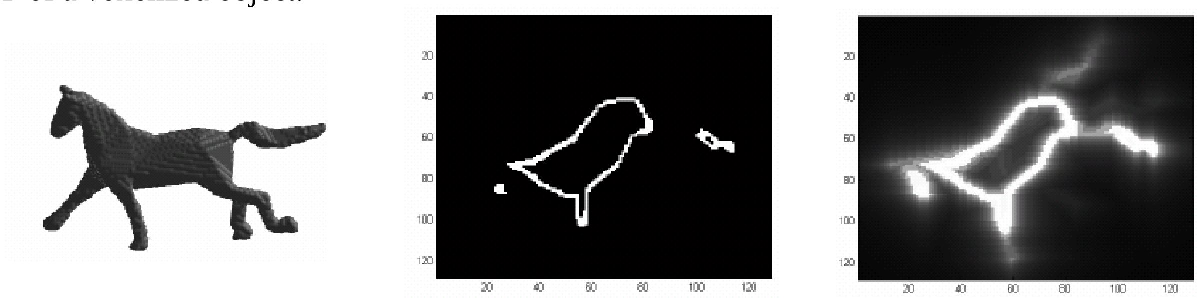

Figure 4 An example of BF and IDF

(a) Voxelized object (b) Cross section of the BF (c) Cross section of theIDF

2.3 Object pose normalization

The shape of an object is the geometrical information that remains after the effects of translation, rotation, and isotropic rescaling, i.e., the effects of affine transformations are removed [11]. A shape descriptor should accordingly be invariant to such transformations by itself, or combined with a certain pose normalization procedure in advance. Here, the pose of an object specifies completely its orientation and position with respect to a predefined frame or coordinate system. It is difficult to recommend each of them since current research groups advocate one or another and support their favorites with experiments [12][13][14][15].In fact, the descriptors designed for invariance come usually with a certain loss of shape information that might be valuable for a specific application, while, defining a canonical reference frame is still an open issue. In the following sub-sections, we'll briefly introduce some commonly used pose normalization methods: principal component analysis(PCA), per-object alignment method, partial matching of SGF, extended Gaussian image(EGI) and complex EGI.

2.3.1 Principal component analysis

Principal component analysis(PCA) is the most commonly used method for pose estimation before 3D shape description [52] [38]. It employs the local features, such as the object boundary and edges to align the 3D object to have a centered canonical reference frame and scale.

Typically, PCA first determines a rigid-body transformation, such as rotation and a uniform scale, which align two models together as closely as possible, before measuring the distance between them. Such global alignment method does not discriminate between the details, and can easily cause similar local features to be misaligned and, consequently, result in an improper global similarity measure. Such a global analysis is prone to errors when the shapes disagree even in an apparently minor region. PCA performs better when models are different in global geometric features.

Conventionally, the PCA is applied only to a set of points, e.g., the centioids of triangles, thus, the different sizes of triangles are not taken into account. Vranic and Saupe[52] introduced weighting factors associated to vertices in order to account the different sizes of triangles of a mesh, which provides improvements to the classical PCA. In [38], continuous PCA(CPCA) is presented, with which the accuracy is limited only by the applied arithmetic and one does not have any systematical errors. The CPCA is very efficient in many cases and can be applied even if the mesh model is not orientable or a closed polygonal surface, while is not time-consuming.

2.3.2 Per-object alignment method

Podolak et al. proposed a per-object alignment method based on finding symmetry axes[16]. Whenever such symmetries exist within the object, this approach may be promising and useful for obtaining semantically more meaningful reference frames. However, the computational complexity is not comparable to that of PCA, which makes PCA more attractive and be widely used as a pose normalization tool.

2.3.3 Partial matching of SGF

Partial matching of the salient global features(SGFs) [53][54] can be used to align two models that have little global similarity. At first, this method searches for matches between pairs of salient features one from each model. For each such match, the associated transformation gets a grade that reflects the number of salient features it successfully aligns. Then, the most successful transformation is voted, and applied to bring the two given models close wherever possible. Afterwards, the voted transformation defines a correspondence between the two models. Once this correspondence is defined, the corresponding features can be brought closer in the least-squares sense.

2.3.4 Complex extended Gaussian images

Extended Gaussian images(EGI) is designed mainly for pose determination and the application of object recognition in an industrial environment. As a pose determination method, EGI representation uses surface information. Extended Gaussian images(EGI)[29] is defined as a histogram that records the variation of surface area with surface orientation[26]. The weights in EGI represent the associated visible face area of the object, which are scalars and do not contain any direct distance information. Thus, it is translation invariant, and it can be seen that the EGI representation rotates in the exact manner as the object in space.
Complex extended Gaussian images(CEGI) also employs the object surface information as features[27]. Its difference from EGI is that the complex weight is composed of the magnitude denoting the corresponding visible face area and the phase denoting the normal distance of the face from the designated origin in the direction of the normal. This method combines the pose determination into shape descriptors and even into the shape matching process.

3. Descriptors for 3D shapes

A shape descriptor, in general, can be viewed as a mapping from the space of 3D objects to some high-dimensional, yet finite, vector space. The main purpose of the shape descriptor researches is to design such mappings that can preserve as much information as possible and to build the resulting representing vector in a possibly low-dimension, which can be considered as effectiveness and efficiency, respectively. These two criteria are contradictory, but also somewhat complementary, and we will discuss it in section 4.

There are lots of 3D shape descriptors being reported since a decade before. The most up-to-date and complete reviews in this rapidly developing field are given in[9][10]. Among all these shape descriptors, some are suitable to describe the simple-structural shapes, and also easy and fast in matching process, while yield poor results when representing and discriminating relatively complicated shapes, which have more detailed information on surfaces of the objects than simple shapes.

3.1 Classification of 3D shape descriptors

Shape descriptors can be roughly divided into three categories as shown in figure 6: (1) feature based shape descriptors, (2) graph based descriptors and (3) other descriptors. The categories of shape descriptors are not completely disjoined. For example, a distribution-based descriptor, in some way, describes the global features of the shape, such as shape distribution descriptor; shape histograms descriptor can be considered a spatial map-based method as well as a global feature distribution-based one according to the different partitions of object surfaces and the histogram structures. Similarly, a graph-based shape descriptor can also be viewed as a global feature-based descriptor.

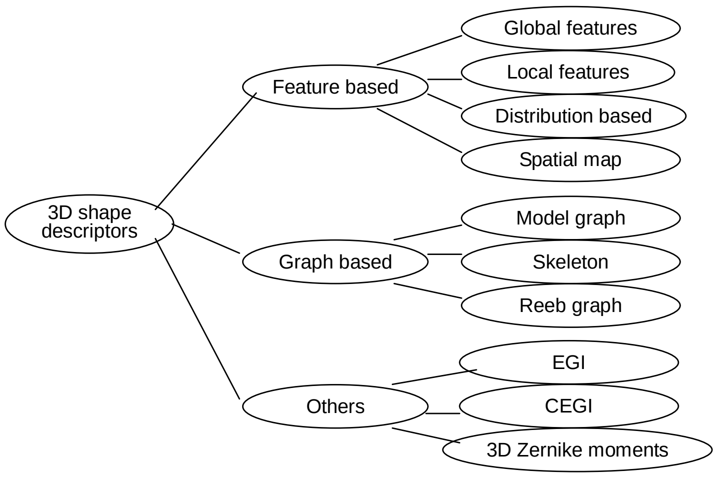

Figure 5 Classification of 3D shape descriptors

In the context of 3D shape matching, features denote geometric and topological properties of 3D shapes. Thus, 3D shapes can be discriminated by measuring and comparing their features.

3.1.1 Feature-based descriptors

Feature based methods can be divided into four categories according to the type of shape features: (1) global features, (2) local features, (3) distribution based, and (4) spatial features. Feature based methods from (1) (3) (4) represent features of a shape using a single descriptor consisting of a d-dimensional vector of values, where the dimension d is fixed for all shapes. In contrast with them, local feature based methods describe for a number of surface points the 3D shape around the point. For this purpose, for each surface point a descriptor is used instead of a single descriptor.

The most common approach to shape based retrieval of 3D objects is to represent every object by a single global shape descriptor representing its overall shape. Global features characterize the global shape of a 3D model. Examples of these features are the moments, invariants, Fourier transform descriptors, and geometry ratios. A lot of efficient ways to calculate these features from the mesh representation of an object were demonstrated in[59]. Generally, global features analyze the shape in an overview perspective and fail to capture the specific details of a 3D

shape. Such feature representation is constrained in describing shapes accurately and discriminating among locally dissimilar shapes.
For example, geometric parameters and ratios are usually used as such shape descriptors. The mostly used features are surface area to volume ratio, compactness (non-dimensional ratio of the volume squared over the cube of the surface area), crinkliness (surface area of the model divided by the surface area of a sphere having the same volume as the model), convex hull features, bounding box aspect ratio, and Euler numbers. However, the discriminating characteristics of these descriptors are very limited.
However, the concept of global feature based similarity has been refined recently by comparing distributions of global features instead of the global features directly.
Spatial maps are the representations that capture the spatial location of an object. The map entries correspond to physical locations or sections of the object, and are arranged in a manner that preserves the relative positions of the features in an object. Thus, the spatial maps based methods are more related to the partition approaches of the object space and, consequently, the transformation and matching methods of the descriptors are more or less selected or designed in terms of the spatial partitions.Spatial maps are generally not invariant to rotations, except for specially designed maps. Therefore, typically a pose normalization is needed first.
Local feature based methods provide various approaches to take into account the surface shape in the neighborhood of points on the boundary of the shape.

3.1.2 Graph-based descriptors

Graph-based approaches are essentially different from other vector-based descriptors either in structures or in design purposes. The graph-based descriptors are more complex and, accordingly, difficult to be constructed and derived. On the contrary, they have the potential of encoding geometrical and topological shape properties in a more faithful and intuitive manner than vector-based descriptors. Therefore, one of the most prominent characteristics is that the graph can represent the relationships, even including semantic relationships, between any two of the sub-parts of the 3D object. However, they do not generalize easily to all representation formats and require dedicated dissimilarity measures and matching schemes. Thus, they are consequently not efficient for general-purpose retrieval applications.

3.1.3 Other descriptors

As for the other category of descriptors, we consider it including Extended Gaussian Image(EGI)[26], complex EGI(CEGI)[27], 3D Zernike moments[42], and so on.

In the following sub-sections, we describe briefly some of the examples of descriptors in the three categories of our taxonomy, i.e., feature-based, graph-based and other descriptors, depending on whether it is suitable for recognizing shapes of broad categories or for locally dissimilar shapes.

3.2 Descriptors for shapes of broad categories

Objects from different broad categories are possible to possess quite different features that are intuitive for people and can be easily represented by some global shape descriptors. Apparently, the global feature based descriptors, such as cord and angle histograms[22][25][26], a typical one of this type, and shape distributions[21] which is also a distribution based descriptor, can be used to represent objects from different broad categories. However, they are not restricted to the type of global feature based descriptors. For example, Shape histograms[22], Radial-cosine transform[34], 3D shape contexts[25], and Extended Gaussian Image[29], which are respectively distribution-based descriptor, local feature based descriptor, semi-local feature descriptor and histogram-based method, can also be employed for shapes of broad categories.

3.2.1 Cord and angle histograms

Cord and angle histograms is a global based descriptor[22][25][26]. A cord is defined as a ray segment which joins the barycenter of the mesh with a triangle center. The histograms of the length and the angles of these rays are used as 3D shape descriptors. This descriptor is easy and efficient to be calculated. However, it simplifies triangles to their centers and does not consider the size and shape of the mesh triangles[17][18]. Therefore, the triangles of all sizes have equal weight in the final distribution. Moreover, centers may not represent adequately the impact of the triangle on the shape distribution because of arbitrary triangle orientations. Paquet and Rioux [19] improve the similar descriptor to consider the angles between surface normal directions and the reference axes. They claim that the bivariate histogram of the angles between the surface normal direction and the first two axes of the reference frame is sensitive to the level of detail by which the object is represented, although it may contain more information that the univariate histograms. They carry out the experiments to prove their claim, which is rather contrast for retrieval in a large database experienced by Akgul[20].

Since only global features are used to characterize the overall shape of objects this method is not very discriminative about object details, but their implementation is straightforward. These methods can be used as an active filter, after which more detailed comparisons can be made, or they can be used in combination with other methods to improve results.

3.2.2 Shape distributions

Shape distributions descriptor is a global feature and distribution based method. Osada et al.[21] use a collection of shape functions i.e. geometrical quantities computed with random sampling of the surface of the 3D object. The main idea of shape distributions is to represent the signature of an object as a shape distribution sampled from a shape function measuring global geometric properties of an object. The purpose of it is to reduce the shape matching problem to the comparison of probability distributions, which is simpler than traditional shape matching methods that require pose registration, feature correspondence, or model fitting. The challenge is to select discriminating shape functions, to develop efficient methods for sampling them, and to robustly compute the dissimilarity of probability distributions.

Shape distributions measure properties based on distance, angle, area and volume measurements between random surface points. The above shape functions are used as shape features: (1) distance of a surface point to the center of mass of the model; (2) distance between two surface points; (3) area(square root) of the triangle defined by three surface points; (4) volume(cube root) of the tetrahedron defined by four surface points; (5) angle formed by three random surface points(mutually visible).

Firstly, histograms of a set of the following shape functions are constructed. Secondly, reconstruct the representation for shape distribution by a piecewise linear function with V equally spaced vertices. Finally, store each shape model as a sequence of V integers. The randomization of the surface sampling process improves the estimation (over approach in[17]), and a more representative and dense set of the surface points is obtained. Additionally, the histogram accuracy can be controlled with sampling density. The descriptor of shape distributions is fast, simple to implement, and useful for 3D shapes discrimination. It works directly on the original polygons of a 3D model, which needn’t model reconstruction from degenerate 3D data. Despite all the advantages above, the shape functions used are not adequate to describe the 3D shape effectively, so it is better to be considered as pre-classification prior to more exact similarity comparison methods.

It is noticeable that shape distributions distinguish models in broad categories very well, such as aircraft, boats, people, animals, etc. However, they perform often poorly when involving the discrimination between shapes that have similar gross shape properties but vastly different detailed shape properties, as we mentioned above.

3.2.3 Shape histograms

Shape histograms is also a distribution-based method. A shape histogram is defined by partitioning the 3D space into concentric shells and sectors around the center of mass of a 3D model, and consequently constructed by accumulating the surface points in the bins based on a nearest-neighbor rule. Ankerst et al.[22] employ three kinds of models, i.e. shell model, sector model and spiderweb(Combined) model, and the experimental comparison shows the histograms based on spiderweb model perform best.

The shape histograms method is an intuitive and discrete representation of complex spatial objects. A Mahalanobis-like quadratic form distance functions are employed for the shape histograms to take into account the distances between histogram bins, which also take small shifts and rotations into account. On the other hand, 3D objects represented by polygonal meshes need to be voxelized prior to descriptor extraction. This approach needs pose normalization before feature representation, because, for instance, the sector model is only scaling invariant, while the shell model is only rotation invariant. Moreover, further reduction of dimensionality of feature vectors are generally needed here.

3.2.4 Radial-cosine transform

The radial cosine transform(RCT)[34] of the 3D function $v(x)$ [23]:

$$V_{RCT}(m) = \sum_x v(x)\phi_m(|x|)$$

where $\phi_m(r)$ are radial cosine transform basis functions and are defined as follows:

$$\phi_m(r) = \begin{cases} 1, & if\ r = 0 \\ 2\cos(\pi m r), & otherwise \end{cases}, \text{ for } m = 0,1,2,\cdots M$$

The RCT coefficients constitute a set of rotation invariant shape descriptors. It represents models with a small number of features, accordingly is easy and fast to be calculated. However, the retrieval results of RCT are generally worse than DFT, Zernike and spherical harmonic descriptors (Evaluation: precision-recall curves, first tier, second tier, E-measure, discounted cumulative gain, average precision). It is always considered to be suitable for pre-classification.

3.2.5 3D shape contexts

3D shape contexts are a natural extension of 2D shape contexts introduced by Belongie et al. for 2D image recognition[24]. Kortgen et al.[25] try to apply 3D shape contexts for 3D shape retrieval and matching.

3D shape contexts are semi-local descriptions of object shape centered at points on the surface of the object. The shape context of a point is defined as a coarse histogram of the relative coordinates of the remaining surface points. The bins of the histogram are constructed by the overlay of concentric shells around the centroid of the model and sectors emerging from the centroid. In the matching process, it consists of local matching and global matching stages, in the former stage, for a point p the best matching point q is found on the other shape, while in the latter stage, correspondences between similar sample points on the two shapes are found. The descriptor of 3D shape contexts is less efficient than the other currently used methods, the indexing is not straightforward, and the obtained dissimilarity measure does not obey the triangle inequality.

3.2.6 Extended Gaussian images

As described in section 2.3.4, EGI is designed mainly for pose determination and the application of object recognition in an industrial environment. As a pose determination method, EGI representation uses surface information. Extended Gaussian images(EGI)[29] is defined as a histogram that records the variation of surface area with surface orientation[26]. The weights in EGI represent the associated visible face area of the object, which are scalars and do not contain any direct distance information. Thus, it is translation invariant, and it can be seen that the EGI representation rotates in the exact manner as the object in space.

EGI has been proposed to avoid the more difficult problem of local feature matching by directly extracting the object surface area distribution with the surface normal. Additionally, it is translation invariant, for the weights in the EGI representation do not contain any direct distance information. Translation invariance is a primary advantage of the EGI, and also the primary drawback of it, so that the translation of a recognized 3D object cannot be recovered[27]. As for the convex polyhedra recognition, EGI can uniquely determine them and it appears to be ideal for convex object recognition without occlusions. Non-convex objects are handled in general by creating a separate orientation histogram for every view in a discrete set of views and matching against this enlarged data structure. Although occlusion is not a common problem in CAD/engineering applications, many engineering parts are non-convex.

3.3 Descriptors for locally dissimilar shapes

The so-called complex objects are those 3D shapes that belong to different but close categories, or those belong to the same category but with locally dissimilar features in shapes or with different poses. Therefore, information of local features inherent in the 3D shapes plays an important role in discrimination of complex objects. Local feature based descriptors, including probability density-based shape descriptors[20], 3D Hough transform descriptor(C3DHTD) [34][32], voxel-based 3D Fourier transform[36], 3D ray-based spherical harmonics[38], 3D voxel-based spherical harmonic[14][15], PCA-spherical harmonics transform(SHT) [12][13][33], and so on, are distinctly the most appropriate methods in order to discriminate and recognize these locally dissimilar 3D shapes. However, graph-based descriptors are also a fairly good choice to represent the structural information of the shapes. Besides, 3D Zernike moment[42], as a kind of moment with the advantages of capturing global information about the 3D shape and not requiring closed boundaries as boundary-based methods, is a wonderful descriptor for 3D shapes dissimilar in local parts.

3.3.1 Shape spectrum

3D Shape Spectrum Descriptor (3D SSD)[34] is defined as the distribution of the shape index over the entire mesh. The shape index [28] is a local geometric attribute of a 3D surface, expressed as the angular coordinate of a polar representation of the principal curvature vector. The principal curvatures is, then, defined as the eigenvalues of the Weingarten map(W) given by the following expression: $W = \ddagger T^{-1} \ddagger U$, where $\ddagger T$ and $\ddagger U$ denote respectively the first and the second fundamental differential forms.

Concerning the original feature itself, the shape index represents salient elementary shapes(convex, concave, rut, ridge, saddle, and so on) and is invariant with respect to scale and Euclidean transforms. As for the 3D SSD, the descriptor locally characterizes free-form surfaces represented as discrete polygonal 3D meshes. 3D SSD possesses the characteristics of: (1)

generality, since 3D meshes may include open surfaces that have not an associated volume; (2) invariance to scale and Euclidean transforms; (3) robustness that different triangulations of the same object are permitted and it successfully retrieves articulated objects with different postures. Koendering[28] experiments a compact descriptor of minimum size of 100 bits/3D mesh model which allows fast browsing and search of 3D model databases.
On the other hand, the 3D SSD requires a non-trivial preprocessing phase for meshes that are not topologically correct or not orientable. For example, a transformation between data formats before extraction of 3D SSD is possibly needed: VRML data are initially intended for graphics purposes, hence, transformation of such rough 3D data into some useful geometrical surfaces is demanded before applying differential geometry-based analysis. Moreover, this descriptor, as a simple local feature representation, is better to be combined with some global representation schemes.

3.3.2 Probability density-based shape descriptors

Akgul[20] defines a geometric feature over the surface of the 3D object, which is calculated on each triangle of the mesh and a set of observations are obtained, each providing a local characterization. Employing the set of observations and kernel density estimation (KDE) [29][30][31], Akgul estimates the probability density of the local geometric feature at target points chosen on the domain of the feature. Consequently, the vector of estimated density values is used as 3D shape descriptor.
The probability density-based shape descriptor considers three different sorts of multidimensional local geometric features of a point p:

(1) Radial feature S_r: It consists of a magnitude component measuring the distance of the point p to the origin and a direction component pointing to the location of the point p.
(2) Tangent plane-based feature S_t: It consists of a magnitude component which stands for the distance of the tangent plane at p to the origin and a direction component of the unit surface normal vector of the tangent plane.
(3) Cross-product feature S_c: It encodes the interaction between the first two features above. It is decoupled into a magnitude component which is the same as that of S_r, and a direction component of the cross product between the vector representation of point p and the unit surface normal vector at p.

Akgul uses the most general form of the kernel approach to estimate the probability density function (pdf) of the local geometric feature sampled as a set of given observations, and takes into account the multitude of points uniformly distributed over the triangle geometry, for only the barycentric sampling of per triangle is not the best option because of possible shape and size non-uniformities of triangles. He also applies Simpson’s 1/3 numerical integration formula to approximate the expected local feature value over the mesh triangle, computing the integral with respect to each vertices of triangle and averaging the three integration results in order to remove the arbitrariness of the vertices. Finally, the performance of descriptors using precision-recall curves and discounted cumulative gain(DCG) values is presented.
As Akgul mentioned in his paper, when performing on the Princeton Shape Benchmark (PSB) Test Set, the KDE-based descriptor has the highest DCG score among all other well-known 3D shape descriptors, including Cord and Angle Histogram(CAH)[17][18], D2 Distribution[21], 3D Hough Transform Descriptor[32], EGI[26], CEGI[27], Radicalized EXtent Descriptor(REXT)[33], and so on. However, this descriptor still suffers a problem that the three features are neither scale- nor rotation-invariant. Since the method depends on them, we must perform prior pose normalization of the mesh.

3.3.3 3D Hough transform

3D Hough transform descriptor(C3DHTD)[34][32] is based upon the principle of accumulating points within a set of planes in 3D space. A plane in $\Re^3$ is uniquely defined by a triplet (s,θ,φ), where $s \geq 0$ denotes the distance from the origin of the coordinate system to plan, and $\theta \in [0,2\pi)$ and $\varphi \in [-\pi/2,\pi/2)$ respectively denote the two angles, i.e. azimuth and elevation, associated with the spherical representation of the plane's unit length normal vector n. Each axis of the parameter space (s,θ,φ) is uniformly sampled and every center p_i of the mesh triangle gives an additive contribution to the bin corresponding to $(s_i^{'},\theta_j\varphi_k)$, where $s_i^{'}$ is the closest value to s_i, which is the distance to the coordinate system' origin of the plane passing through p_i and with orientation (θ_j,φ_k). In order to avoid the drawbacks of PCA in pose normalization, Zaharia and Preteux[32] propose the octahedron-based partitions of the 3D spherical object and construct the canonical 3DHTD completely specified by an unique 3D HT by deriving one PCA coordinate system(PCA CS) from each other with appropriate permutations of the HT coefficients.

3DHT can be considered as a generalized version of EGI. They are similar except for the way the contributions of the triangles are assessed. Akgul et al.[35] have experimentally proven the conjecture that 3DHT descriptor captures the shape information better than the EGI descriptor.

Firstly, 3DHT uses PCA before shape feature extraction. Secondly, it defines the 3D Hough Transform: $h:(s,\theta,\varphi) \to R$ mapping parameter space (s,θ,φ) to R, where, each axis of the parameter space is uniformly sampled, and each centric point p of the mesh faces gives an additive contribution to each element $h(\tilde{s}_{jk}^{p},\theta_j,\varphi_k) \in (s,\theta,\varphi)$ according to a pre-defined distance-orientation rule. Finally, it constructs the canonical 3D HTD completely specified by a unique 3D HT, associating with an arbitrary PCA CS from the 48 CSs, and defines the similarity measure between two C3DHT.

During the construction process of 3DHT descriptor, a critical point is the resolution to the drawbacks of PCA[32]. If consider all the 48 possible PCA-based coordinate systems(PCA CS) and generate all the corresponding 3DHTs, it is time consuming. Although 3 generating configurations(GCs) are sufficient to ensure a complete Hough representation, a specific domain on the uniformly sampled unit sphere does not fit anymore the resulting partition after changing the PCA CS. Moreover, it requires a high complexity in terms of descriptor size and matching computation time. A practical resolution is to define partitions on unit sphere to ensure there exists a 1-1 mapping between the 3 GCs(3 GCs are equivalent). For instance, one can abandon the principle of uniform sampling of the unit sphere, construct the partitions by projecting the vertices of any regular polyhedron(here octahedron) onto the unit sphere, which is invariant to changes between the 48 PCA CSs, and accordingly called canonical 3D HTD(C3DHTD).

The C3DHTD deserves a lot of advantages. It associates with a spatial alignment procedure to be geometric invariant and is completely independent of the mesh topology without considering whether the size of the mesh polygons is well-adapted to the HT specific granularity. Moreover, it satisfies the storage and computational complexity requirements to similarity-based retrieval applications. Zaharia and Preteus[32] carry out the experimental proves that C3DHTD outperforms the MPEG-7 3DSSD and the EGI, using a subset of 362 models of the MPEG-7 3D test set[34] and under some pre-set parameters(thresholds) and the evaluation of Bull-Eye Percentage(BEP). Additionally, it is also mentioned that the approving retrieval performances do not dramatically degrade when dealing with large amount of data because of the scalability of the C3DHTD. Nevertheless, this descriptor is, intrinsically, not geometrical transformations invariant and the PCA before feature representation is still needed. In [32], only classic EGI, but not the CEGI is involved in the experimental comparison.

3.3.4 Voxel-based 3D Fourier transform

Vranic et al.[36] introduce voxel-based 3D Fourier transform descriptor after they proposed a ray-based spherical harmonics method[38]. The voxel-based 3DFT defines a binary voxel-based feature after voxelization using the so-called bounding cube(BC), which is the tightest cube in the canonical coordinate frame that encloses the model, with the center in the origin and the edges parallel to the coordinate axes.

In order to extract the features of a 3D model, a pose normalization is needed before the voxelization using the BC in the spatial domain. After that, the descriptor represents the feature in the frequency domain by 3D Discrete Fourier Transform(3DFT).

Comparing to previous voxel-based methods, 3DFT significantly reduces the size of feature vector when choosing a greater value of N(number of subdivision of BC). It makes calculation of feature distances with L1 of L2 norms much efficiently and leads to good experimental results. In another perspective, when comparing with other methods such as cords-based descriptor, rotation invariance(90 degrees) descriptor, ray-based descriptor with spherical harmonics, 3DFT presents best overall performance on categories of “cars” and “airplanes”. On the other hand, since the 3DFT coefficients are not rotation invariant, DFT is applied after alignment to principal axes. Moreover, 3DFT leads to problems with outliers because of the use of BC. Theoretically, using bctree to represent features in spatial domain can solve it, but the problem of large size of feature vector still exists.

Later, Dutagaci et al.[37] propose a 3D discrete Fourier transform descriptor to the two different voxel representations of 3D objects, namely, binary denoting object and background space, and continuous after distance transformation, and finally compare the experiments on them and showed the latter is better. The voxel-based feature is defined as a representation of inverse distance function(IDF), as already described in section 2.2.2.

The 3D DFT employs a measure of the spectral energy(SE) in a sphere of radius r, because the spectral energy in a sphere centered at the origin of the frequency domain remains constant under rotation. Afterwards, it defines the incremental spectral energy(ISE) as the difference of the spectral energies contained within concentric spheres and normalize the ISE by r^2 and take its square root to balance out large values accumulated in the low-pass shells. The normalized spectral energy(NSE), which has the property of rotation invariance, is used as the 3DFT-based descriptors of the object.

This descriptor of 3D DFT achieves a multi-resolution representation that NSE descriptors values at small radii (low-pass region) carry information about the gross shape of the object, while shape details are encoded in the spectral shells at high-frequency radii. Therefore, 3DFT-based scheme, gives good results in both binary and IDF case when comparing with RCT-based descriptors, and gives better results in E-measure, discounted cumulative gain and average precision when comparing with 3D SH-based descriptors. Additionally, the size of the DFT-based descriptor is far smaller than that of SH. On the other hand, slight deviation from rotation invariance can occur due to voxelization distortion. Unfortunately, IDF representation is not totally rotation invariant due to the distance transform values at the corners of the bounding box.

3.3.5 Spherical harmonics transform

The Princeton group and the Konstanz group had considerable impact in 3D shape descriptors research to date, and spherical harmonics transform(SHT), which have become a very popular tool in the field of 3D shape descriptors, to a larger extent has been one of the main tools used in a great deal of descriptors developed by these two groups. However, there is an on-going debate between Princeton and Konstanz groups on the use of SHT for 3D shape description, emphasizing on whether PCA normalization should be applied prior to SHT and, whether the

magnitude of transform coefficients or the energies in different bands of the transform domain should be used as the descriptor. Both of them carry out retrieval experiments to support their opinions.
As a transform function, SHT is suitable to reduce the descriptor size considerably without loosing too much shape information. At present, there are two types of typical methods involving spherical harmonics transform: rotation invariant spherical harmonics(RISH) and PCA-spherical harmonics transform(PCA-SHT).

3.3.5.1 Rotation invariant spherical harmonics(RISH)

Funkhouser and Kazhdan et al. present a general approach based on spherical harmonics to transform rotation dependent shape descriptors into rotation independent ones[14][15].
The shape descriptor is defined as 3D voxel-based spherical harmonic(VH) feature vector. The basic idea in forming any considered descriptor of a model is to divide the model into R spheres(with radii 1 through R) and apply the Spherical Fourier Transform(SFT) to a function on each of a specific sphere(with radius r_i). Firstly, one can define a function $r = r(\theta,\varphi)$ on a sphere(with radius r_i), approximate the function using spherical harmonic basis function($Y_{l,m}$), figure out the complex Fourier coefficients $\hat{r}(l,m)$ and form the corresponding feature vector with absolute values $\left|\hat{r}(l,m)\right|$ of dimension $\dim = (l_{\max}+1)l_{\max}/2$, which represents the sphere. Secondly, use the L2-norm to represent the shape model, securing the rotation invariance(RI). There are l_m L2-norms corresponding to each sphere with radius r_i, and totally $l_m \times R$ L2-norms to be the components of feature vector, which represents the model. The most prominent advantage of this SHT descriptor is the rotation invariance(RI). However, it should be noticed that voxel grid-based representation loses many fine details and rotation invariance properties of SHT descriptors should be understood with caution as it comes with a certain loss of shape information.
Overall, Kazhdan et al. provide mathematical support for rotation invariance of the descriptor[15]. Basically, this mathematical justification relies on the fact that the energy in a certain frequency band of the ST does not change when the object is rotated around its center of mass. However, the rotation invariance SHT descriptors are considered as losing a certain amount of shape information, which partly leads to the controversy between the Princeton and Konstanz group.

3.3.5.2 PCA-spherical harmonics transform(PCA-SHT)

The Konstanz group uses SHT to transform spherical functions densely sampled over the surface to obtain spherical harmonic(RH) feature vector of a PCA-normalized object[12][13][33]. These SHT descriptors need PCA or continuous PCA(CPCA) to achieve pose normalization before feature extraction.
In [12], the so called ray-based or extent descriptor gives the SHT-transformed version of the maximal distance from the center of mass as a function of the spherical angle. In [13], Vranic et al. obtain the feature vector by forming a complex function on the sphere, apply the Fast Fourier Transform(FFT) on the sphere and generate Fourier coefficients for spherical harmonics. The feature vector is then composed of the absolute values of the coefficients. In [33], Vranic improves the previous descriptor and proposes two 3D shape descriptors based on functions on concentric spheres. It combines ray-based notion into voxel-based sphere partition methods, finds all points of intersection with the polygonal mesh, and then defines several functions on the concentric spheres with different radii, and afterwards again with SFT. This improved descriptors consider the internal structure of a model by using functions on concentric sphere and

increase retrieval effectiveness, for instance, outperform the previous ray-based ones[12][13] and the voxel-based SHT descriptors[14].

The descriptor of PCA-SHT leads good experimental results outperforming the RI-SHT by Funkhouser[14]. Furthermore, the pose determination method used here, i.e. continuous PCA(CPCA), although it shows certain weaknesses, is still very efficient and not time-consuming in many cases, which can be applied even if the mesh model is not orientable or a closed polygonal surface.

3.3.6 Multi-resolution Reeb graphs

Reeb graphs at multiple levels of resolution of a function μ is defined over the object's surface[39][40]. The function μ can be the height of a point on the surface, the curvature value of the point, or the integrated geodesic distance at that point. According to the definition of Reeb graph, each node in each Reeb graph corresponds to a connected component of the object in the sense that μ-values in that component fall within the same interval determined by the resolution at which the graph is constructed. The parent-child relationships between nodes represent adjacent intervals of these μ-values for the contained object parts.

According to the function chosen, the resulting descriptor provides certain invariance properties. A graph at a coarser level is encoded as the ancestor of a graph at a finer level which contains more detailed information than its ancestor. The singular points that correspond to the finest resolution locations are valuable in studying the topology of the underlying object. In spite of all the good properties mentioned above, Reeb graphs are not applicable to all classes of shapes, and the choice of Reeb function affects results significantly[6].

3.3.7 Skeletal graphs

The skeletal graph is obtained from object voxel data as a directed acyclic graph(DAG). According to the definition of the skeletal graph, each node of the DAG is associated with a set of geometric features and a signature vector that encodes topological information. The topological signature vector(TSV) is derived from the eigen-decomposition of the graph's adjacency matrix.

Sundar et al.[41] describe a method for searching and comparing 3D objects via skeletal graph matching in order to build an interactive system that allows part-matching. The visualization of the results helps the user to define and interactively change his/her query. There are two stages of matching procedure, where the second one can be used to refine the possible set of retrieved database objects.

3.3.8 Complex extended Gaussian images

Although complex EGI(CEGI)[27] representation is a global feature based method, it is considered suitable for objects that belong to close categories and are locally dissimilar in shapes. The complex weight of a CEGI is comprised of a magnitude of the corresponding visible face area and a phase of the normal distance of the face from the designated origin in the direction of the normal.

The difference of CEGI from conventional EGI is that it has the distances encoded in its weights in a different manner and it is not translation invariant. CEGI essentially allows both the orientation and translation of a given 3D object to be determined separately, in which determining the translation parameters is also simple based on the least squares formulation.. It can differentiate larger classes of objects than the conventional EGI, for CEGI disambiguates objects having similar EGIs. Furthermore, CEGI is identical for both polyhedral and smooth objects.

On the hand, CEGI shows some disadvantages either. It derives better results for a smooth object with a more even Gaussian distribution and a lower degree of concavity. Moreover, the accuracy of the derived translation parameters is sensitive to the angular error of the surface normals, the magnitude of the actual translation parameters, and the distribution of the surface normals.

3.3.9 3D Zernike moments

3D Zernike moment is a kind of moment with the advantages of capturing global information about the 3D shape and not requiring closed boundaries as boundary-based methods[42]. Zernike moments are a projection of the function defining the object onto a set of orthonormal functions within the unit ball. They can be considered as the magnitudes of a set of orthogonal complex moments of the 3D shape and the natural extensions of spherical harmonics based descriptors.
Comparing to the regular(un-orthogonal) moments, it is easier to recover 3D shape from Zernike moments and there will be less computationally expensive. Additionally, the information content of the recovered 3D shape has no redundancy because of the orthonormality. Comparing to other orthogonal moments, 3D Zernike moments possess a useful rotation invariance property. It is also easy and instructive for 3D shape reconstruction from Zernike moments. Firstly, it is intuitively that 3D Zernike moments are able to separate out the individual contribution of each order moment to the reconstruction process. Secondly, the maximum order(required number of features) of the moments can be determined by the close degree of reconstructed 3D shape to the original one. Moreover, the contribution of moments to the reconstruction process can instruct the weights of the corresponding features. Thereby, the reconstruct approximations of the original object from 3D Zernike moments possess completeness. However, 3D Zernike moments suffer by the severe instability of geometrical moments and hence always in case of high orders. Anyway, the descriptor of 3D Zernike moments are superior over others in terms of noise sensitivity, information redundancy and discrimination power and still reported to be among the most successful representations at present.

3.4 Discussion on shape descriptors

The common objective in 3D descriptor research is to design mappings from the space of 3D objects to some high-dimensional, yet finite, vector space, in a way to preserve as much information as possible and to keep the resulting vector as low-dimensional as possible. With regard to the 3D shape descriptors mentioned above, some of them are theoretically similar to each other, while some are quite different in essence. We compare all these categories of shape descriptors in some most important and typical aspects, such as original shape features, spatial partition methods, pose normalization, transformation invariance, advantages and disadvantages, and so on, as shown in table 1.
No matter in what ways a descriptor is obtained, there always remains the ambiguity about the similarity notion to associate with it. As we know, the approach to describe a 3D shape results in descriptors embedded in high-dimensional vector spaces which are generally obscure from a mathematical perspective. To clarify, it is usually not known theoretically which distance or norm would be the most suitable for retrieval and classification. For the time being, since there is no universal mathematical theory for 3D shapes and for designing such mappings in particular, the usual practice is to experimentally try a set of distance functions, and consequently evaluate the effectiveness and efficiency - two criteria of shape descriptors - of a descriptor largely on experimental terms, and eventually report their classification and retrieval performances.

Table 1 Comparison on 3D shape descriptors

Category of methods	Descriptors	Shape feature	Spatial partition	Transformation invariance	Computation cost	Advantages	Limitations	Refs.
Global feature-based	Cord and angle histograms	Histograms of the length and the angles of the cord rays	Ray-based	R		a. Consider the angles between surface normal directions and the reference axes	a. Simplify triangles to their centers; b. intolerant to impact of shape distribution	[17][18][19]
Distribution-based	Shape distributions	A collection of shape functions	Randomly selected points of surfaces	T, R	O(S log N) where S is number of samples and N is number of triangles	a. Fast, simple, useful to discriminate 3D shapes; b. Randomization of the surface sampling process improves the estimation; c. Histogram accuracy can be controlled with sampling density.	a. Shape functions are not adequate to describe the 3D shape; b. Better to be pre-classification.	[21]
Spatial map-based	Shape histograms	Accumulation of the surface points in the bins based on a nearest-neighbor rule	Shell model, sector model, and spiderweb	R(shell model) S(sector model)	O(N3B) where N is the number of voxels along each axis and B is the number of histogram bins	a. Intuitive; b. make use of Quadratic form distance functions to take into account the distances between histogram bins.	a. Voxelization is needed prior to descriptor extraction; b. Need further reduction of dimensionality of feature vectors.	[22]
	Radial cosine transform	Coefficients of radial cosine transform of a 3D function	Shell-like model	R		a. Easy to calculate, fast; b. Represent models with small number of features	a. Retrieval results are generally worse than commonly used methods; b. Suitable to be pre-classification.	[23]
	3D Hough transform	3D Hough Transform: $h:(s,\theta,\varphi)\rightarrow R$	a. Uniformly sampled each axis of parameter space; b. Octahedron-based partition of the unit sphere.	R(canonical 3D HTD)		a. Associate with a spatial alignment procedure to be geometric invariant; b. Satisfy the storage and computational complexity requirements; c. Completely independent of the mesh topology; c. Stable retrieval performances	a. Only the canonical 3DHTD of octahedron-based partition is rotation invariant.	[34][32]
	Voxel-3D Fourier transform(3DFT)	Binary Voxel-based feature	Voxelization using the bounding cube	-		a. Significantly reduce the size of feature vector; b. Efficient calculation of feature distances; c. Good experimental results	a. Alignment to principal axes is needed; b. 3DFT leads to problems with outliers because of the use of BC.	[36] [37]
	Rotation invariant spherical harmonics	Spherical functions of each specific sphere; Voxel-based spherical harmonic feature vector	Voxel-based division in terms of radii	R	O(N3K) where N is the number of voxels along each axis and K is the number of spherical functions	a. Satisfy rotation Invariance(RI).	a. Voxel grid-based representation loses many fine details.	[14][15]
	PCA-Spherical harmonics Transform(SHT):	A complex function on the sphere.	Ray-based division in terms of given directions within a unit sphere.	-		a. SHT reduces the descriptor size considerably without loosing too much shape information; b. outperform the RI-SHT in[38 in thesis]	a. CPCA shows certain weaknesses.	[12][13]
		Spherical functions on concentric spheres; Ray-based spherical harmonic feature vector	Combined ray- and voxel-based partitions: RH1 and RH2	R	O(N3K) where N is the number of voxels along each axis and K is the number of spherical functions	a. Consider the internal structure of a model by using functions on concentric spheres; b. Rotation invariant by using CPCA.		[33]
Local feature-based	3D shape contexts	A coarse histogram of the relative coordinates of the remaining surface points	Concentric shell and sector models	-			a. Less efficient; b. Indexing is not straightforward; c. The obtained dissimilarity measure does not obey the triangle inequality.	[25]
	Shape spectrum	The distribution of the shape index over the entire mesh		T, S		a. General to 3D meshes including open surfaces; b. Robust to different triangulations of the same object and retrieving articulated objects with different postures; c. Allow fast browsing and search of 3D model databases.	a. Transformation of rough 3D data into useful geometrical surfaces is demanded; b. Better to combine with some global representation schemes.	[28]

	Probability density-based shape descriptors	The probability density of the three local geometric features over the surface.	Multi-points of every triangle on the mesh	-		a. Outperforms all other well-known 3D shape descriptors with highest DCG score.	a. Not invariant with any geometric transformation.	[20]
Graph based	Multi-resolution Reeb Graphs	Reeb graphs at multiple levels of resolution of a function over the surface	Topology of the object	R	O(V log V) where V is number of mesh vertices	a. Have the potential of encoding geometrical and topological shape properties; b. The singular point locations are valuable in studying the topology of the object.	a. Not applicable to all classes of shapes; b. Choice of Reeb function affects results significantly.[6]	[39][40]
	Skeletal graphs	A directed acyclic graph(DAG) associated with a set of geometric features and a signature vector	Voxel data	R	O(N3) where N is the number of voxels along each axis	a. Allows part-matching.	a. Mre elaborate and complex in calculation; b. Not efficient for general-purpose retrieval applications.	[41]
Others	Extended Gaussian images	A histogram recording the variation of surface area with surface orientation	Surface-based	T		a. Avoids the more difficult problem of local feature matching; b. Translation invariant; c. Unique representation for convex objects without occlusions.	a. Does not contain any direct distance information; b. Confusion in non-convex objects.	[26]
	Complex Extended Gaussian images	A histogram recording the variation of surface area and distance with surface orientation	a. Two-frequency dodecahedron-based 240 sampling view directions of the sphere; b. A simple ray-tracing technique	-		a. Encodes distance in its weights; b. Identical for both polyhedral and smooth objects; c. Disambiguates objects having similar EGIs.	a. Accuracy of the derived translation parameters is sensitive to some parameters.	[27]
	3D Zernike moments	Magnitudes of a set of orthogonal complex moments of the object	Voxel-based partition	R	O(N3) where N is the number of voxels along each axis	a. Information content has no redundancy; b. Easy for 3D shape reconstruction from Zernike moments. c. Superior over others in terms of noise sensitivity, information redundancy and discrimination power.	a. Severe instability of geometrical moments; b. High orders.	[42]

4. Conclusions

A 3D shape descriptor can be designed in a variety of ways according to the shape characteristics of 3D objects and the requests of applications. In the scope of DecorAR project, we aim at recognizing a small set of shapes in a fast and accurate manner in order to be used in real time interaction. In this context, we are concerned with shapes possessing obvious visual distinction on the whole shapes from each other, which can always be detected by human eyes. The majority of these shapes can be discriminated with its outlines. Therefore, it is not necessary to distinguish these objects with detailed information implicated in every local part. As a result, we are supposed to focus mainly on the global features of shapes, which are easier to capture than local features, while integrate most of the general features of the shapes as a whole. In addition, some local features, although not dominating features for shape classification, can also be utilized to recognize simple shapes so as to better describe the 3D objects from a local and more detailed perspective and improve the 3D shape descriptors.

Based on the analysis of descriptors introduced in section 3, we select respectively two 3D shape descriptors for broad categories classification and two for locally similar objects classification, which are cord and angle histograms, shape distributions descriptor, 3D complex function FFT descriptor, and complex EGI.

Cord and angle histograms[22][25][26] and shape distributions[21] are both global feature based descriptors. The former uses the histograms of the length and the angles of cords as 3D shape descriptors, which are easy and efficient to be calculated and implemented. The latter uses a collection of shape functions measuring global geometric properties of an object, and represents the signature of an object as a shape distribution sampled from these shape functions. Shape distributions distinguish models in broad categories very well, such as aircraft, boats, people, animals, etc. However, both of these methods are not very discriminative about objects details, because only global features are used to characterize the overall shape of the objects.

3D complex function FFT descriptor [13] engages spherical harmonics to merge two features represented by real functions (x(u) and y(u)) by embracing them into a single complex function, apply the Fast Fourier transform(FFT) on the sphere and generate Fourier coefficients for spherical harmonics. Some experiments show the complex feature vector FFT performs better than ray-based and shading-based feature vector.

Complex EGI descriptor is a global feature based method which is considered suitable for objects locally dissimilar in shapes that belong to close categories. The CEGI encodes the normal distance and the visible area of each outward surface normal as a complex weight of the feature representation, which captures the rich information of shape surface, whereas, with appropriate computation complexity. It can also distinguish a convex object from a non-convex object or disambiguate non-convex objects having similar EGIs, and accordingly achieve satisfactory experimental results. Therefore, the shape descriptor of CEGI is relatively more suitable for our 3D shape recognition design.

As we know, the approach to describe a 3D shape results in descriptors embedded in high-dimensional vector spaces which are generally obscure from a mathematical perspective. Few descriptors can be theoretically proved which distance or norm would be the best for recognition and classification. Researchers usually experimentally try a set of distance functions and compare their performances, and the function that produces the best scores is consequently considered as the most suitable for the particular descriptor and the object database. In the same way, we plan to design several experiments for these 3D shape descriptors on a set of 3D shapes and compare their effectiveness and efficiency for shape classification.

5. References

[1] 3D model search engine, http://shape.cs.princeton.edu/.

[2] 3D model similarity search engine, http://merkur01.inf.uni-konstanz.de/CCCC/.

[3] 3D model retrieval system, http://3d.csie.ntu.edu.tw/~dynamic/cgi-bin/DatabaseII_v1.8/index.html.

[4] 3D shape retrieval engine, http://www.cs.uu.nl/centers/give/imaging/3Drecog/3Dmatching.html.

[5] P. Min, M. Kazhdan, and T. Funkhouser. A comparison of text and shape matching for retrieval of online 3D models. In Proc. European Conference on Digital Libraries, Bath, UK, September 2004.

[6] Iyer, N., Jayanti, S., Lou, K., Kalyanaraman, Y., and Ramani, K. "Three Dimensional Shape Searching: State-of-the-Art Review and Future Trends", Computer-Aided Design, Volume 37, Issue 5, 15 April 2005, pp.509-530.

[7] I. Atmosukarto and P. Naval. A survey of 3D model retrieval systems, www.comp.nus.edu/˜indri/research.htm, 2003.

[8] R. C. Veltkamp and M. Hagedoorn. State-of-the-art in shape matching. In M. Lew, editor, Principles of Visual Information Retrieval, pages 87–119. Springer Verlag, 2001.

[9] J. W. H. Tangelder, R. C. Veltkamp. “A survey of content based 3D shape retrieval methods”. In Proc. of the Shape Modeling International 2004 (SMI’04), pages 145-156, Genoa, Italy, 2004.

[10] B. Bustos, D. Keim, D. Saupe, T. Schreck, D. Vranic: Feature-based Similarity Search in 3D Object Databases, ACM Computing Surveys (CSUR) 37(4):345-387, Association for Computing Machinery, 2005.

[11] D. G. Kendall. The diffusion of shape. Adv. Appl. Prpbab., 9:428-430, 1977.

[12] D. Saupe and D. V. Vrani´c. 3D model retrieval with spherical harmonics and moments. In Proceedings of the DAGM 2001, Munich, Germany, September 2001.

[13] D. V. Vrani´c and D. Saupe. Description of 3D-shape using a complex function on the sphere. In Proc. of the IEEE International Conference on Multimedia and Expo (ICME 2002), pages 177–180, Lausanne, Switzerland, August 2002.

[14] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, and D. Jacobs. A search engine for 3D models. ACM Trans. Graph., 22(1):83–105, 2003.

[15] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation invariant spherical harmonic representation of 3D shape descriptors. In Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing (SGP ’03), pages 156–164, Aire-la-Ville, Switzerland, 2003. Eurographics Association.

[16] J. Podolak, P. Shilane, A. Golovinskiy, S. Rusinkiewicz, and T. Funkhouser. A planar-reflective symmetry transform for 3D shapes. ACM Transactions on Graphics (Proc.

SIGGRAPH), 25(3), July 2006.
[17] E. Paquet, A. Murching, T. Naveen, A. Tabatabai, and M. Rioux. Description of shape information for 2D and 3D objects. Signal Processing: Image Communication, 16:103–122, 2000.
[18] E. Paquet and M. Rioux. Nefertiti: a query by content software for three-dimensional models databases management. In Proc. of the International Conference on Recent Advances in 3-D Digital Imaging and Modeling (NRC ’97), page 345, Washington, DC, USA, 1997. IEEE Computer Society.
[19] E. Paquet and M. Rioux. Nefertiti: A tool for 3D shape databases management. SAE Transactions: Journal of Aerospace, 108:387–393, 2000.
[20] Ceyhun Burak Akgul. “3D Shape Descriptors and Similarity Learning”. Thesis.
[21] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape distributions. ACM Trans. Graph., 21(4):807–832, 2002.
[22] M. Ankerst, G. Kastenm¨uller, H.-P. Kriegel, and T. Seidl. 3D shape histograms for similarity search and classification in spatial databases. In Proc. of the 6th International Symposium on Advances in Spatial Databases (SSD ’99), pages 207–226. Springer-Verlag, 1999.
[23] H. Duta˘gacı, B. Sankur, and Y. Yemez. Transform-based methods for indexing and retrieval of 3D objects. 3DIM, 00:188–195, 2005.
[24] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. PAMI, 24(4):509–522, 2002.
[25] M. K¨ortgen, G.-J. Park,M. Novotni, and R. Klein. 3D shape matching with 3D shape contexts. In The 7th Central European Seminar on Computer Graphics, April 2003.
[26] B. K. P. Horn. Extended Gaussian images. Proc. of the IEEE, 72:1671–1686, 1984.
[27] S. B. Kang and K. Ikeuchi. The complex EGI: A new representation for 3D pose determination. IEEE Trans. Pattern Anal. and Mach. Intell., 15(7):707–721, 1993.
[28] J. Koendering. Solid shape. The MIT Press, 1990.
[29] R. O. Duda, P. E. Hart, and D. G. Stork. Pattern Classification. Wiley-Interscience, 2000.
[30] W. H¨ardle, M. M¨uller, S. Sperlich, and A. Werwatz. Nonparametric and Semiparametric Models. Springer Series in Statistics. Springer, 2004.
[31] D. W. Scott. Multivariate Density Estimation. Theory, Practice and Visualization. Wiley, New York, 1992.
[32] T. Zaharia and F. Prˆeteux. Shape-based retrieval of 3D mesh models. In Proc. of the IEEE International Conference on Multimedia and Expo (ICME’2002), Lausanne, Switzerland, August 2002.
[33] D. V. Vrani´c. An improvement of rotation invariant 3D shape descriptor based on functions on concentric spheres. In Proceedings of the IEEE International Conference on

Image Processing (ICIP 2003), pages 757-760, Barcelona, Spain, September 2003.
[34] T. Zaharia and F. Prˆeteux. Three-dimensional shape-based retrieval within the MPEG-7 framework. In Proceedings SPIE Conference on Nonlinear Image Processing and Pattern Analysis XII, Vol. 4304, pp. 133-145, San Jose, CA, January 2001.
[35] C. B. Akg¨ul, B. Sankur, Y. Yemez, and F. Schmitt. A framework for histogram-induced 3D descriptors. In European Signal Processing Conference (EUSIPCO ’06), Florence, Italy, 2006.
[36] D. V. Vrani´c and D. Saupe. 3D shape descriptor based on 3D Fourier transform. In Proc. of the EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services (ECMCS ’01), Budapest, Hungary, September 2001.
[37] H. Dutaˇgacı, B. Sankur, and Y. Yemez. Transform-based methods for indexing and retrieval of 3D objects. 3DIM, 00:188–195, 2005.
[38] D. V. Vrani´c, D. Saupe, and J. Richter. Tools for 3D-object retrieval: Karhunen-Loeve-transform and spherical harmonics. In Proc. IEEE 2001 Workshop on Multimedia Signal Processing, Cannes, France, (in press), 2001.
[39] M. Hilaga, Y. Shinagawa, T. Kohmura, and T. L. Kunii. Topology matching for fully automatic similarity estimation of 3D shapes. In ACM SIGGRAPH, pages 203–212, Los Angeles, CA, USA, August 2001.
[40] T. Tung and F. Schmitt. The augmented multiresolution Reeb graph approach for content-based retrieval of 3D shapes. International Journal of Shape Modeling (IJSM), 11(1), June 2005.
[41] H. Sundar, D. Silver, N. Gagvani, and S. Dickinson. Skeleton based shape matching and retrieval. In Proc. of the Shape Modeling International 2003 (SMI ’03), page 130. IEEE Computer Society, 2003.
[42] M. Novotni and R. Klein. 3D zernike descriptors for content based shape retrieval. In Proc. of the 8th ACM symposium on Solid modeling and applications (SM ’03), pages 216-225, New York, NY, USA, 2003. ACM Press.
[43] O. D. Faugeras, “New steps toward a flexible 3-D vision system for robotics,” in Proc. Int. Conf. Patt. Recognition, July 1984, pp. 796-805.
[44] O. D. Faugeras and M. Hebert, “A 3-D recognition and positioning algorithm using geometrical matching between primitive surfaces,” in Proc. Int. Joint Conf. Artificial Intelligence, Aug. 1983, pp. 996-1002.
[45] B. Bhanu, “Representation and shape matching of 3-D objects,” IEEE Trans. Patt. Anal. Machine Intelligence, vol. PAMI-6, no. 3, pp. 340-350, 1984.
[46] F. Stein and G. Medioni, “Structural indexing: Efficient 3-D recognition,” IEEE Trans. Patt. Anal. Machine Intelligence, vol. 4, no. 2, pp. 125-145, 1992.
[47] R. M. Bolle and B. D. Copper, “On optimally combining pieces of information, with

application to estimating 3-D complex-object position from range data," IEEE Trans. Patt. Anal. Machine Intelligence, vol. PAMI-8, no. 5, pp. 619-638, 1986.

[48] R. C. Bolles, P. Horaud, and M. J. Hannah, "3DPO: A three-dimensional part orientation system," in Proc. First Int. Symp. Robotic Res., 1984, pp. 413-424.

[49] R. Nevatia and T. 0. Binford, "Description and recognition of curved objects," Artificial Intelligence, no. 8, pp. 77-98, 1977.

[50] T. J. Fan, G. Medioni, and R. Nevatia, "Matching 3-D objects using surface descriptions," in Proc. Int. Conf. Robotics Automat., Apr. 1988, pp. 1400-1406.

[51] J. Ponce and D. J. Kriegman, "On recognizing and positioning curved 3-D objects from image contours," in Proc. Image Understanding Workshop, May 1989, pp. 461-470.

[52] D. V. Vranic and D. Saupe, "3D Model Retrieval". Proc. SCCG 2000, May 3-6, Budmerice, Slovakia, p.89, 2000.

[53] Hoffman, D. D., Singh, M. Salience of visual parts. In Cognition 63, 1997, 1, 29-78.

[54] Ran Gal, Daniel Cohen-Or. Salient Geometric Features for Partial Shape Matching and Similarity, ACM Transactions on Graphics, January 2006, 25(1).

[55] http://immi.inesc-id.pt/cali.

[56] Manuel J. Fonseca and Joaquim A. Jorge: "Using Fuzzy Logic to Recognize Geometric Shapes Interactively", Proceedings of the 9th Int. Conference on Fuzzy Systems, (FUZZ-IEEE'00), pages 191-196, San Antonio, USA, 05/2000.

[57] Manuel J. Fonseca and Joaquim A. Jorge: "Experimental Evaluation of an On-line Scribble Recognizer", Pattern Recognition Letters, 22(12): 1311-1319, 01/2001.

[58] Manuel J. Fonseca, Alfredo Ferreira, and Joaquim A. Jorge: "Generic Shape Classification for Retrieval", GREC 2005, HK, China, 08/2005.

[59] Zhang C, Chen T. Efficient feature extraction for 2D/3D objects in mesh representation. IEEE International conference on Image Processing (ICIP 2001), Thessaloniki, Greece 2001.