Content-Based Retrieval of 3-D Objects Using Spin Image Signatures

Jürgen Assfalg, Marco Bertini, Alberto Del Bimbo, *Member, IEEE*, and Pietro Pala, *Member, IEEE*

***Abstract*—Retrieval by content of 3-D models is becoming more and more important due to the advancements in 3-D hardware and software technologies for acquisition, authoring and display of 3-D objects, their ever-increasing availability at affordable costs, and the establishment of open standards for 3-D data interchange. In this paper, we present a new method, referred to as Spin Image Signatures, that develops on the original spin images approach, with adaptations to support effective retrieval by content. According to the method proposed, a set of spin images is derived for each model, to obtain a view-independent description of its 3-D shape and a signature is evaluated for each spin image in the set. Clustering is hence performed on the set of Spin Image Signatures to obtain a compact representation. Experimental results are presented, showing the effectiveness of the Spin Image Signatures method for retrieval, also in comparison with other methods, and its sensitivity to model deformations.**

***Index Terms*—3-D content-based retrieval, 3-D shape description, spin images.**

I. Introduction

In RECENT years, 3-D digital models have become of widespread use in many application domains. Just to make a few examples: in manufacturing, 3-D models are used to model objects and components, and are reused for rapid prototyping of new products; in medicine, 3-D data are used to make more accurate diagnosis of human diseases; in entertainment, 3-D graphic scenes and models are used to develop videogames or to lower production costs of movies and TV fictions; in cultural heritage, archival of digital 3-D copies of original manufacts improves the quality of documentation and allows monitoring of deformations or changes that can occur through time. This is basically due to the availability of automatic techniques and tools, including tomography, magnetic resonance, 3-D laser scanners, structured light systems, and photogrammetry, that support the generation of 3-D digital models with untextured or textured surfaces at a wide range of resolutions. In this scenario, retrieval by content of 3-D models becomes an important subject of scientific investigation.

Solutions for retrieval of 3-D models differ in the representation used for modeling the object geometry, the information on the object appearance that is exploited (i.e., color and/or texture), and whether manual or automatic annotation is needed. An extended review of different solutions for 3-D shape representation and content-based retrieval of 3-D models is presented in [23]. In the following, we review and discuss some of the major approaches by grouping them into three different categories: retrieval based on *metadata*, retrieval based on projected *views*, and retrieval based on 3-D *structure*.

With retrieval based on metadata, annotations that are used for matching are typically input manually and refer to high level concepts describing object's type and attributes. Annotations are often subjective. Retrieval must cope with synonyms and semantic meaning similarity and requires the definition of useful ontologies. The cost of annotations limits the use of these methods to small sized 3-D model archives.

View-based approaches, assume that salient features of a 3-D object can be extracted from a set of 2-D views of the object. Matching is therefore reduced to assessing the similarity between the views of the query object and those of the models in the database. With this approach, results of content-based image retrieval [4], [18] can be exploited effectively. Retrieval of 3-D models based on projected views has been proposed in [2], where representations of the projected object contours are considered. Other similar solutions have been proposed by [13] and [15]. In [25], views, referred to as *light fields*, of the object are taken from observation points uniformly distributed on the surface of a sphere centered in the object's center of mass. For each of these views, Zernike moments and Fourier descriptors are computed so as to reduce the 2-D information to a 1-D feature vector. Computational complexity of retrieval is reduced by a multistep approach supporting early rejection of nonrelevant models. Retrieval based on views presents two main problems: on the one hand, self-occlusions may impair the completeness of the representation; on the other hand, a large number of views can degrade the efficiency of retrieval. This makes these approaches mostly suitable for simple 3-D objects.

Structure-based 3-D approaches rely on a topological or structural description of models that is computed directly in the 3-D space. Structural descriptions are obtained from either analytical or discrete models. Analytical models are difficult to obtain, due to limitations in providing effective parameterizations of arbitrary object shapes. An example of this approach is shown in [10], where retrieval of 3-D objects is based on the similarity of surface segments corresponding to potential docking sites of molecular structures that are modelled through parametric surface patches. The strong assumptions on the shape of the surface constrain the application of the approach to the original context; moreover, since the model does not

Manuscript received February 8, 2005; revised June 12, 2006. The associate editor coordinating the review of this paper and approving it for publication was Dr. Arnold W. M. Smeulders.

The authors are with the Dipartimento Sistemi e Informatica, Università di, Firenze, 50139 Firenze, Italy (e-mail: assfalg@dsi.unifi.it; bertini@dsi.unifi.it; delbimbo@dsi.unifi.it; pala@dsi.unifi.it).

Color versions of Figs. 2, 5–8, 10, and 12 are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2006.886271

represent the entire surface of molecules, retrieval based on global similarity is not supported.

Discrete models define 3-D object descriptors on the basis of the spatial properties of the 3-D polygonal mesh approximating the object surface. These are, by far, the most widely employed solutions, with notable examples in the recent literature [1], [5], [6], [11], [14], [16], [17], [21], [22], [24]. The system developed within the Nefertiti project [17] supports retrieval of 3-D models based on both mesh geometry and object color and texture. Mesh geometry is defined by the volume aspect ratio, a wavelet decomposition of the voxel-based representation of the volume enclosed by the mesh, and the distribution of *chords'* angles and lengths, being chords defined as the vectors from the center of mass to the center of the polygons of the mesh surface. Using chords rather than surface normals makes the system less sensitive to local variations of their orientations.

Shape functions, mapping geometric properties to a scalar value, computed on couples or triples of random points on the surface, have been used in [16] and [24]. A histogram of these values is then used as a signature for the model. Since signatures are evaluated for the entire object, characteristics that occur more frequently might have higher relevance than less frequent yet distinctive features. Kolonias *et al.* have used aspect ratios of the object bounding box, a binary voxel-based representation of geometry [11], and a set of paths (*model routes*) defined for the mesh shape. Model routes are computed according to a set of heuristics, and take into account edges that form an almost straight polyline. Evaluation of all model routes may yield a very large number of features, so making the method unsuitable for retrieval. Funkhouser *et al.* [21] evaluate spherical harmonics on binary voxel-based representations of polygonal surfaces and derive signatures. A major limitation of spherical harmonics for the purpose of retrieval relates to the impossibility of distinguishing between models that differ by a rotation of the interior part. In [14], feature points of the 3-D object are taken at the median curvature and the torsion maxima of the surface. To reduce the sensitivity to noise, a preliminary iterative smoothing is applied. Different orientations of the surface of the object are stored in a hash table mapping triples of surface points into their orientations [12]. Since this approach has been conceived for object recognition, an efficient matching scheme is required for large model repositories. In [5], moments (up to the fourth to seventh order) of surface points are used as features for retrieval of 3-D models. Evaluation of moments is not affected by (self-)occlusions. Use of moments and Fourier transform coefficients of the 3-D volume is also reported in [22]. It is recognized that 3-D moments, as well as Fourier transform coefficients, have poor capability to capture perceptually distinguishing traits of 3-D model geometry. In [1], a complex method is proposed according to which 3-D models are warped and projected onto a map. This provides view-independent descriptions but can be applied effectively only to objects modeled by a simply connected 3-D region.

In this paper, we present content based retrieval of 3-D models according to new descriptors based on spin images [8] that we will refer to as *Spin Image Signatures*. While the original formulation of the spin-image approach was conceived for object recognition, the proposed new signature exploits modifications that make the method suitable for access to large 3-D object repositories, still maintaining high discriminatory capability. The method can be considered hybrid view-based and structure-based. In fact, similar to view-based approaches, each spin image is the object view as seen from one object vertex. On the other hand, different from view-based approaches, the spin-image view taken from a generic object vertex includes information of both visible and cluttered vertices. In that, spin images provide an object-centered description that is insensitive to rigid transformations and robust to geometric deformations that alter the model surface. To make the spin image representation as compact as possible, spin image content is reduced to the 1-D histogram of the spin-image distribution in a spatial quantization grid. Clustering of histograms is used in order reduce the number of descriptors, still keeping evidence of the most characterizing elements.

The paper is organized as follows. Section II presents an overview of the original spin images approach [8] and discusses its limitations and drawbacks for content based retrieval. Section III presents in detail the new proposed approach and the definition of Spin Images Signatures. In Section IV, clustering of Spin Image Signatures and similarity matching of 3-D models with the new descriptors are expounded. Experimental results and a comparative analysis with other solutions in the literature are presented in Section V. Finally, conclusions are drawn in Section VI.

II. Spin Images

Spin images encode the density of mesh vertices projected onto a 2-D space. Given $O = \langle \boldsymbol{v}, \boldsymbol{n_v} \rangle$ an *oriented point* on the surface of the object, where $\boldsymbol{v}$ is a vertex on the meshed surface of the object and $\boldsymbol{n_v}$ the normal to the tangent plane in $\boldsymbol{v}$, the spin image for vertex $\boldsymbol{v}$ is derived by projecting the mesh vertices onto the tangent plane with normal $\boldsymbol{n_v}$ and building the 2-D histogram of the projected vertex coordinates.

Construction of spin images is controlled by a few parameters that influence the quality of representation, the storage requirements, as well as the coverage of the object to be described. These are the following.

- *Bin size*: the quantization step used in the 2-D projected space. It is usually defined with respect to mesh resolution. Large bin size values reduce noise sensitivity, but yield coarse descriptions.
- *Support angle*: the upper limit of the angle between $\boldsymbol{n_v}$ and the surface normal at mesh vertex $\boldsymbol{x}$. If the angle exceeds the limit, the vertex $\boldsymbol{x}$ is not accounted in the spin image. Low values of the support angle cause the spin image to account only for vertices whose normal is close to $\boldsymbol{n_v}$; differently, for large values, the spin image accounts for more vertices.
- *Attenuation function*: a function that weights the contribution of each point $\boldsymbol{x}$ according to its distance from vertex $\boldsymbol{v}$. Near and far points from $\boldsymbol{v}$ can be treated in the same way, or weighted differently.

TABLE I
INFLUENCE OF THE NUMBER OF SECTORS AND CROWNS OVER PERFORMANCE OF RETRIEVAL

	Sectors						
Crowns	3	4	5	6	7	8	9
3	0.93103	0.76552	0.52857	0.95862	0.57037	0.59310	0.96552
4	0.51724	0.95172	0.64828	0.55862	0.97241	0.59310	0.67586
5	0.53333	0.51724	0.95862	0.72414	0.74286	0.97131	0.91429
6	0.96552	0.54483	0.71034	**0.97241**	0.84828	0.84138	0.96552
7	0.71034	0.96552	0.88276	0.88276	0.96552	0.94483	0.86207
8	0.78621	0.74483	0.95172	0.92414	0.83448	0.96552	0.90345
9	0.95862	0.91724	0.91034	0.96552	0.87586	0.91034	0.95172

— *Image size*: the size of the spin image expressed in pixels. The product of *bin size* and *image size* defines the *support distance*, which determines the range of points that are included in the spin image.

A complete description of a 3-D object requires as many spin images as the number of the object vertices. In the attempt to reduce this huge amount of data, Johnson and Hebert used principal component analysis (PCA) to extract a compact representation of spin images: images are reduced to a vectorial representation and projected in a new base according to PCA. In the new base, elements are ordered in decreasing order of relevance, so that the projection of a generic vector (spin image) on the first elements of the base keeps information on the prominent characteristics of the original vector. A comparison of spin images with a 3-D extension of shape context and with harmonic shape context [28], for the purpose of 3-D object recognition in range images, was reported in [26].

However, using PCA to extract a compact representation of spin images has some drawbacks that make the approach unsuitable for content-based retrieval of 3-D models. Firstly, if the database is frequently updated, the addition/deletion of models requires that a new base is computed from the new set of spin images thus requiring high computational efforts to have coherent descriptions. Secondly, the most significant elements in the new base can be selected through the application of a threshold on the cumulative covariance of each element. In this case, the dimension of the representation space is dependent on the content of the database. Finally, although PCA enables the reduction of the dimension of the representation, the number of descriptors is not reduced. Since high-resolution 3-D models may require thousands of spin images, in this case, memory management becomes a critical issue for both the computation of the PCA base and indexing and matching in the retrieval phase.

In the following section, we show how Spin Image Signatures overcome these critical points.

III. SPIN IMAGE SIGNATURES

Spin image signatures are defined as n-dimensional vectors, following a region-based description of spin image content, similarly to the shape matrix description model [7].

For each spin image $I(i,j)$, three independent sets of regions are defined: sectors of circular crowns (for the $\beta > 0$ half-plane and for the $\beta < 0$ half plane), and circular sectors (see Fig. 1). Instead of sampling the shape at the intersection between radial and circular lines, like in the shape matrix approach, for each region, the number of spin image points that are included is considered, according to

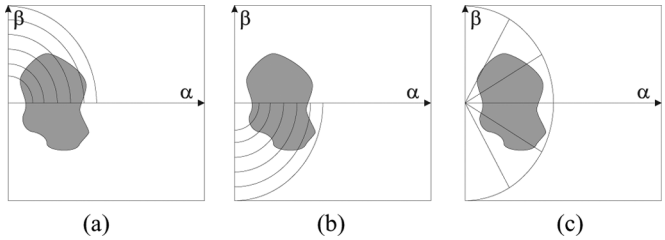

Fig. 1. Spin image signature: (a) crowns in the half-plane $\beta > 0$, (b) crowns in the half-plane $\beta < 0$, and (c) sectors.

$$cp_k = \frac{\sum_{[i,j]\in CP_k} I(i,j)}{\sum_{[i,j]} I(i,j)} \quad k = 1,\ldots,np$$

$$cn_k = \frac{\sum_{[i,j]\in CN_k} I(i,j)}{\sum_{[i,j]} I(i,j)} \quad k = 1,\ldots,nn$$

$$s_k = \frac{\sum_{[i,j]\in S_k} I(i,j)}{\sum_{[i,j]} I(i,j)} \quad k = 1,\ldots,ns$$

where CP_k is the kth "positive" crown, CN_k is the kth "negative" crown, S_k is the kth circular sector, np is the number of "positive" crowns, nn is the number of "negative" crowns, and ns the number of circular sectors.

The Spin Image Signature is obtained as the concatenation of three vectors, $C^p = \langle cp_1,\ldots,cp_{np}\rangle$, $C^n = \langle cn_1,\ldots,cn_{nn}\rangle$, and $S = \langle s_1,\ldots,s_{ns}\rangle$, representing the number of spin image points in the positive and negative crowns and sectors, respectively. The representation is invariant to translation, rotation, and scaling.

A satisfactory tradeoff between compactness and accuracy of the representation was found with $np = nn = ns = 6$. Table I reports performance values for $ns, np, nn \in \{3,\ldots,9\}$ in terms of precision and recall. Influence of the number of sectors (ns) and crowns $(np = nn)$ for Spin Image Signatures over the performance of retrieval was assessed by measuring precision P and recall R for different ns and $np = nn$ values, and taking the intersection of the $P(R)$ curve with the $P = R$ line. Experiments were carried out with the core-art database (described later in Section V).

From Table I, it can be noticed that the optimal combination is obtained for $ns = np = nn = 6$.

IV. Description and Matching of 3-D Models

Using Spin Image Signatures, one signature is computed for every vertex of the model mesh, so that the number of Spin Image Signatures per model is equal to the number of the mesh vertices. In the case of large archives of models taken at high resolution (in the order of thousands of mesh vertices), this results into an unmanageable number of descriptors and prevents efficient retrieval by content. Solutions are therefore required to reduce the number of Spin Image Signatures per model, still maintaining an adequate level of accuracy in retrieval.

In our approach, a solution is achieved by clustering Spin Image Signatures according to fuzzy clustering [3]. For a generic model, let $\{D_1, \ldots, D_m\}$ be the set of m Spin Image Signatures in $\mathbb{R}^{18}$ $\left(D_i \in \mathbb{R}^{18}\right)$. Signature vectors are clustered into c clusters, according to fuzzy c-means, by minimizing the following function:

$$J_{fcm} = \sum_{i=1}^{m}\sum_{j=1}^{c}(u_{ij})^{\mu} d^2(D_i, \mathbf{v}_j) \tag{1}$$

where $\mu \in (1, \infty)$ (*fuzzy exponent*) determines the fuzziness of the clusters, $\mathbf{v}_j$ $j = 1, \ldots, c$ are the cluster centers, u_{ij} is the membership degree of vector D_i in cluster $\mathbf{v}_j$, and $d(\cdot)$ is the Euclidean distance in $\mathbb{R}^{18}$. Values of u_{ij} are subjected to the following constraints:

$$\sum_{j=1}^{c} u_{ij} = 1, \quad \forall\, i, \text{ and } 0 \leq \sum_{i=1}^{m} u_{ij} < m, \quad \forall\, j$$

where the case of $\sum_{i=1}^{m} u_{ij} = 0$ leaves open the possibility of empty clusters.

Computation of the optimal number of clusters c is accomplished according to the approach proposed in [9]: given two functions that express respectively a measure of under-partitioning and over-partitioning, the optimum partition, i.e., the trade-off between under- and over-partitioning, is defined as the number of clusters that minimizes the sum of the two functions. To this end, mean intra-cluster distance (MICD) and inter-cluster minimum distance (ICMD) are defined, such that if the set of signatures is under-partitioned, at least one cluster has a large MICD value, and for over-partitioned sets, MICD and ICMD values are small (because at least one dense cluster collecting similar signatures is subdivided)

$$\mathrm{MICD}(i) = \sum_{\mathbf{x} \in \chi_i} \frac{|\mathbf{v}_i - \mathbf{x}|}{n_i} \tag{2}$$

where χ_i is the set of data elements comprising the ith cluster, n_i its cardinality and $\mathbf{v}_i$ the cluster center, and

$$\mathrm{ICMD} = \min_{i \neq j} |\mathbf{v}_i - \mathbf{v}_j|. \tag{3}$$

The optimal value c_{opt} is obtained for each model, by computing for different values of c the under- and over-partition functions $v_u(c)$ and $v_o(c)$, defined as

$$v_u(c) = \frac{1}{c}\sum_{i=1}^{c} \mathrm{MICD}^{(i)} \quad v_o(c) = \frac{c}{\mathrm{ICMD}}.$$

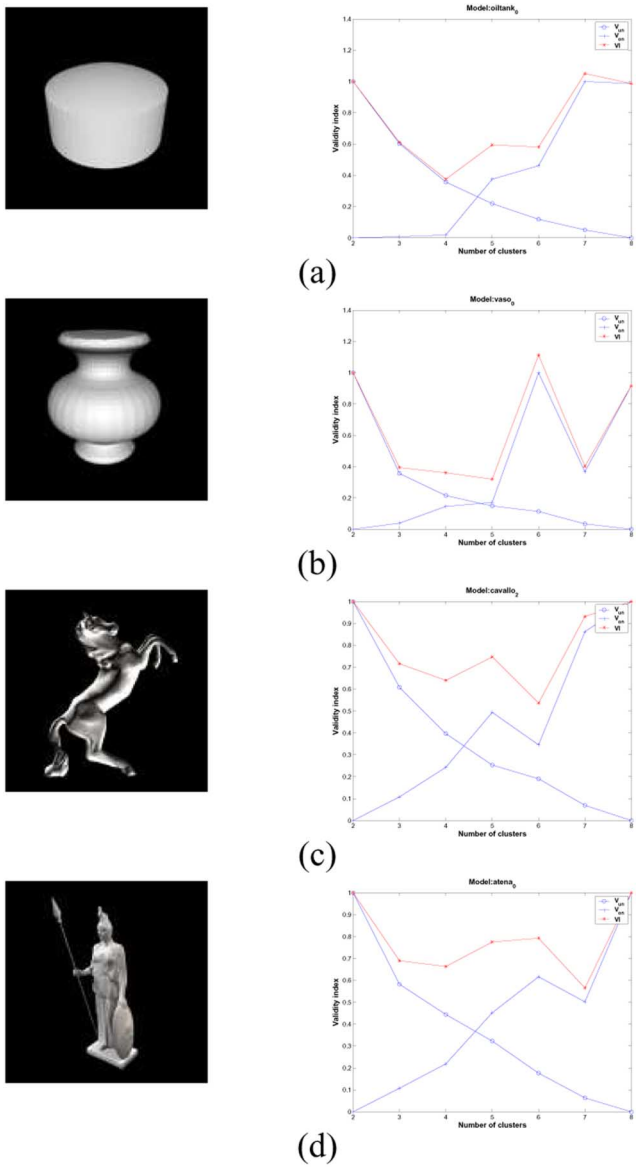

Fig. 2. Plots of over- and under-partitioning (V_{on} and V_{un}) and validity index (VI) functions for sample object models. The optimal number of clusters c_{opt} corresponds to the minimum of the validity index.

Values of the under- and over-partition functions are normalized according to

$$v_{u\min} = \min_c v_u(c) \quad v_{u\max} = \max_c v_u(c)$$
$$v_{o\min} = \min_c v_o(c) \quad v_{o\max} = \max_c v_o(c)$$
$$v_{uN}(c) = \frac{v_u(c) - v_{u\min}}{v_{u\max} - v_{u\min}}$$
$$v_{oN}(c) = \frac{v_o(c) - v_{o\min}}{v_{o\max} - v_{o\min}}$$

and the *validity index* (VI) is computed as

$$\mathrm{VI}(c) = v_{uN}(c) + v_{oN}(c). \tag{4}$$

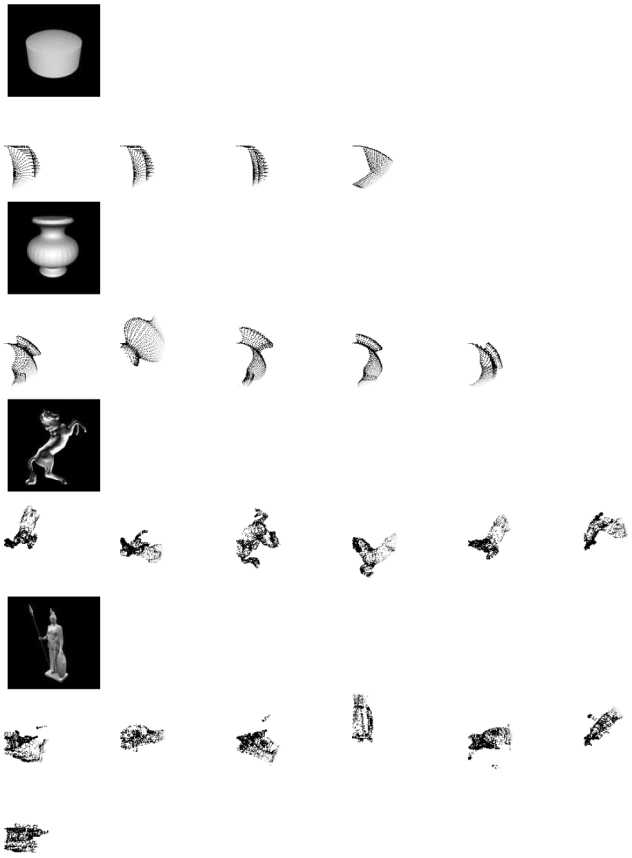

Fig. 3. Spin images of the cluster centers identified after fuzzy clustering for four sample object models.

The optimal number of clusters c_{opt} is finally obtained as the value of c that minimizes $\text{VI}(c)$:

$$c_{\text{opt}} = \arg\left\{\min_{c} \text{VI}(c)\right\}.$$

In Fig. 2, values of VI, v_{uN} and v_{oN} are plotted for spin images signatures of 3-D models with different structural complexity. The optimum number of clusters is also shown as the minimum of the validity index curve. Simple models (with a number of mesh vertices ranging from 10^2 to 10^3) have values of c_{opt} equal to 3 or 4; higher values (between 6 and 7) are observed for more complex models (with a number of vertices in the range 10^3-10^4). Typically, the clustering process reduces the number of model Spin Image Signatures from several thousands to less than ten signature clusters. The cluster centers are assumed as the model descriptor. In Fig. 3, the spin images of the cluster centers for four sample models are presented.

Due to the indeterministic nature of clustering, different initializations of the clustering process may determine different cluster centers. We have investigated the extent to which random initialization influences clustering of Spin Image Signatures of 3-D models. In more detail, we assume that the effect of clustering initialization is to move the final position of the cluster centers. According to this, let $\left\{C_i^{(1)}\right\}$ and $\left\{C_j^{(2)}\right\}$ be the sets of cluster centers resulting from two different initializations. If the two clustering processes were independent from the initialization, the two sets would be identical, i.e., $\forall\, C_i^{(1)} \exists C_j^{(2)}$ such that $d\left(C_i^{(1)}, C_j^{(2)}\right) = 0$. Instead, due to the dependency of clustering on initialization, the distance $d\left(C_i^{(1)}, C_j^{(2)}\right)$ is greater than zero. Therefore, the distance between $C_i^{(1)}$ and the nearest $C_j^{(2)}$ can be regarded as a measure of the *shift* of cluster centers induced by the initialization.

TABLE II
EFFECTS OF RANDOM INITIALIZATION ON CLUSTER CENTERS SHIFTS OF SPIN IMAGE SIGNATURES

N. clusters	Mean shift	σ^2 shift	Mean intra-cluster distance
3	0.2500	0.0402	0.6131
4	0.2712	0.0374	0.6321
5	0.2629	0.0382	0.6367
6	0.2513	0.0352	0.6258
7	0.2405	0.0336	0.6133

Table II shows mean values and variance of cluster center shifts, and the average intra cluster distances, for 3-D models of different complexity (with final number of clusters ranging in [3,7]), after 10 random initializations applied to all the models of the Core-art database. The average intra cluster distance can be used as a reference to appreciate the amount of cluster centers shifts. It can be noticed that the average mean shift is always smaller than the mean intra cluster distance.

To assess the similarity between two 3-D models, the sets of the cluster centers $\mathcal{D} = D_i,\ i = 1\ldots$ for the two models are compared to each other. The best cluster-center correspondence function is defined as the permutation $p : \{1,\ldots,l\} \rightarrow \{1,\ldots,k\}$ that minimizes the sum of distances between corresponding cluster centers, that is

$$\Delta\left(\mathcal{D}, \hat{\mathcal{D}}\right) = \min_{p}\left\{\sum_{i=1}^{l} \delta(D_i, \hat{D}_{p(i)})\right\} \tag{5}$$

where D_i is the ith cluster center in $\mathcal{D}$ and $\hat{D}_{p(i)}$ the $p(i)$th cluster center in $\hat{\mathcal{D}}$. The solution p to (5) is obtained according to a *greedy search*, by scanning all the cluster centers in $\mathcal{D}$ and associating to each cluster center the most similar cluster center not yet associated in $\hat{\mathcal{D}}$. Although this yields a suboptimal solution of (5), it represents a satisfactory tradeoff between accuracy of similarity and computational complexity.

V. EXPERIMENTAL RESULTS

In the following, we present and discuss some experimental results that demonstrate, under different perspectives, the potential of using Spin Image Signatures of 3-D models for the purpose of content based retrieval. In particular, the following aspects have been analyzed.

i) the capability of capturing the salient structural properties of 3-D models and dependency on model resolution. We provide precision/recall figures of Spin Image Signatures in comparison with other methods in the literature (Section V-A). Tests have been carried out over the following.
   — *Core-art* database, including about 300 medium-resolution models (from 3180 to 5150 mesh vertices) from miscellaneous sources in the web (many are high quality scans of 3-D models from the De Espona 3-D Models Encyclopedia http://www.deespona.com).

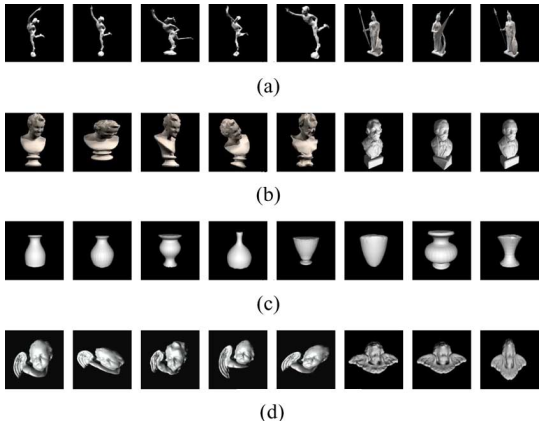

Fig. 4. Examples of content based retrieval by Spin Image Signatures. Query models (leftmost images for each row) have different structural properties. Retrieved models are ranked from left to right. The query model is always retrieved in the first position.

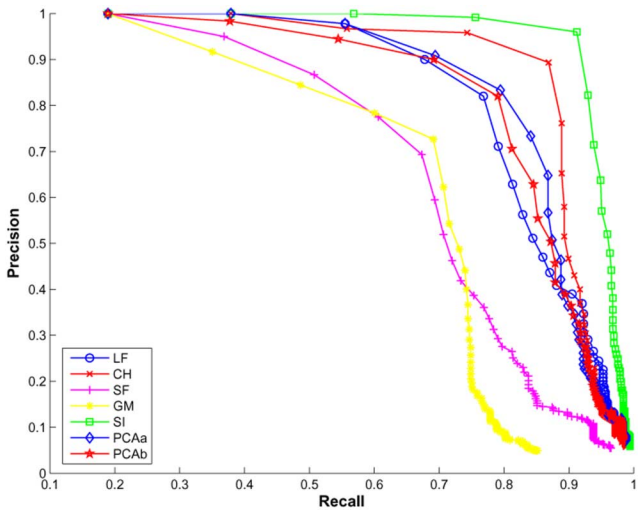

Fig. 5. Precision/recall figures for Light Field (LF), Curvature Histograms (CH), Shape Functions (SF), Geometric Moments (GM), Spin Image Signatures (SI) and PCA Spin Images (PCAa, PCAb) using the Core-art database.

— *Princeton-subset* database including the first 500 models of the Princeton Shape Benchmark [29] archive (models from 22 up to 10120 mesh vertices).

The two databases include models of humans, animals, buildings, body-parts…

ii) effectiveness for retrieval by part similarity (Section V-B). Tests have been made with the *Primitive-shape* database, including about 200 low resolution models that were obtained by aggregating simple 3-D models such as spheres, pyramids, rings, cones, and cubes, to form more complex 3-D forms.

iii) the sensitivity to geometric deformations (Section V-C). Tests have been carried out over the *Extended-art* database, including about 10000 medium-resolution models, that were obtained from the Core-art database by applying selected geometric deformations to each original model.

iv) Storage requirements and computational complexity (Section V-D).

Models in the test datasets are labelled with metadata indicating the type of object. They are in VRML format and composed of mesh vertices and polygonal faces according to the IndexedFaceSet data structure of VRML.

A. Retrieval Performance Comparison

Fig. 4(a)–(d) shows examples of retrieval by similarity, that provide a qualitative view of the capabilities of Spin Image Signatures to capture different structural properties of 3-D models. The examples are 3-D models taken from the Extended-art database. From Fig. 4(a)–(c), we can observe that the models retrieved have structural properties that follow those of the query model, in all the three different cases. Fig. 4(d) shows that using Spin Image Signatures, retrieval by model part similarity can also be dealt with to some extent. Using a single-winged putto as a query model, models of puttos with two wings are also retrieved.

Fig. 5(a) shows the precision versus recall curves obtained with the core-art database with Spin Image Signatures and several different solutions expounded in the literature. They were derived by averaging precision and recall over different sample queries, using models of different structural complexity (considering the number of extruding/intruding parts) and profiles (considering the model elongation), but all with almost the same resolution (average resolution approximately 4200 vertices).

The methods that have been put in comparison are the Light Field [25], the Curvature Histograms [19], the Shape Function [24], the Geometric Moments [5], and Spin Images in their original formulation [8]. In particular:

— the Light Field representation was obtained taking 100 views of the model from ten observation points uniformly distributed on the surface of the sphere centered in the model's center of mass. For each view, Zernike moments and Fourier descriptors were computed. Retrieval is performed according to a multistep approach, with early rejection of nonrelevant models.

— Curvature histograms were constructed by calculating the values of curvature at the vertices of the mesh. Retrieval was performed using histogram intersection to compute the similarity between curvature histograms.

— For Shape Function, object shapes were described using the probability distribution of the chord length between pairs of vertices of the 3-D model. The probability distribution was approximated through the 128-bin chord histogram. The dissimilarity between two histograms was evaluated according to the Kullback-Leibler divergence.

— To evaluate Geometric Moments of surface points, the coordinates of vertices of the mesh $\{P_i = (x_i, y_i, z_i)\}_{i=1}^{N}$ were considered. Moments were computed up to the sixth order and evaluated as

$$m_{pqr} = \frac{1}{N}\sum_{i=1}^{N} w_i x_i^p y_i^q z_i^r$$

where w_i is a weight that is proportional to the area of the portion of surface associated to vertex P_i. Independency from the position of the model was obtained by computing

the first-order moments m_{100}, m_{010} and m_{001}, and hence evaluating higher order moments with respect to them.

— According to the original formulation of the Spin Image method, each 3-D model was represented by a set of spin images, each of them corresponding to a specific vertex of the model mesh; principal component analysis (PCA) was then used to represent the spin image content through a reduced size feature vector. The dissimilarity between models M^a and M^b was measured as

$$\mathcal{D}(M^a, M^b) = \sum_{i=1}^{N_a} \min_{j \in \{1,\ldots,N_b\}} d_{Eucl}\left(e_i^{(a)}, e_j^{(b)}\right) \qquad (6)$$

where $M^a = \left\{e_i^{(a)}\right\}$ $i = 1, \ldots, N_a$ and $M^b = \left\{e_j^{(b)}\right\}$ $j = 1, \ldots, N_b$ are respectively the PCA-reduced sets of descriptors of the object models M^a and M^b, and d_{Eucl}. is the Euclidean distance.

Software programs for Curvature Histograms, Geometric Moments, and Shape Functions were implemented according to the information provided in [5], [19] and [24], respectively. For the Light Field method, we used the public software available at http://3d.csie.ntu.edu.tw/~dynamic/3DRetrieval/. Two different versions of the PCA-based Spin Image approach were implemented: in the first the description dimension was set to 108 (as derived from optimal eigenvector selection according to the Kaiser's criterion [30]); in the second, the description dimension was set to 18 to perform a fair comparison with the dimension of the Spin Image Signatures description. For each query model, precision and recall were evaluated counting how many models of the same category (as defined in the metadata) were retrieved.

From the results, we can affirm that for medium-resolution models, as those in the Core-art database, Spin Image Signatures exhibit very good precision/recall closely followed by the Light Field, Curvature Histogram, and the PCA-based Spin Image approaches.

Performance comparison was also carried out with 3-D models of the Princeton Shape Benchmark database. Due to the extreme variability of model configurations present in this database, very different performance figures can be obtained, depending on what models are used as sample queries. Fig. 6 shows some examples of retrieval with the Spin Image Signatures method with models of different structural complexity and resolution, using the Princeton-subset database. In particular, we selected models numbered 29, 122, 106, and 399. Models 106 and 399 (horse and church, respectively) have several protruding parts; model 29 (bird) has only few minor protrusions, and model 122 (human body) has a very elongated shape, with few protrusions all aligned in the same direction. For the case of horse models, the Princeton-subset database includes six models with very different resolutions (from 146 to 16049 vertices, resulting in about 110 max to min ratio). In all the other cases, resolution of models spans a reduced range of values. Fig. 7 shows the precision/recall performance of Spin Image Signatures, Light Field, Shape Functions, Curvature Histograms, Geometric Moments and PCA Spin Images for the queries of Fig. 6.

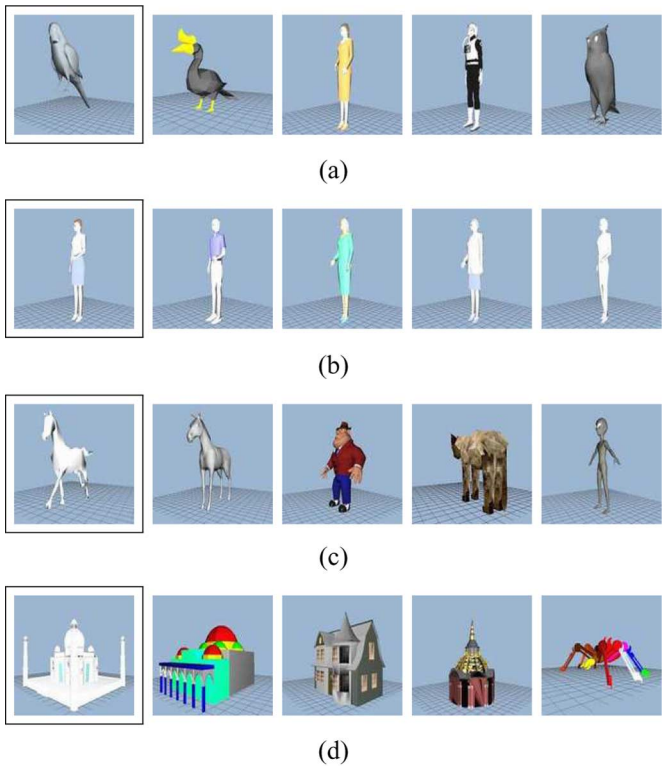

Fig. 6. Examples of retrieval with Spin Image Signatures from the Princeton-subset database. Queries correspond to models (a) 29, (b) 122, (c) 106, and (d) 399 of the Princeton Shape Benchmark database.

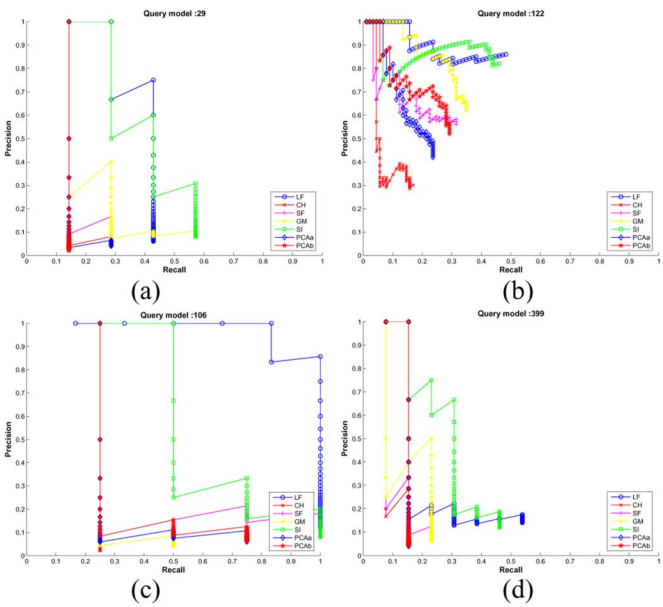

Fig. 7. Precision versus recall curves of Spin Image Signatures (SI), Light Field (LF), Shape Functions (SF), Curvature Histograms (CH), Geometric Moments (GM), and PCA Spin Images (PCAa and PCAb) for queries shown in Fig. 6.

It can be observed that Spin Image Signatures and Light Field always have the highest performance. From the case of the horse model, we can observe that model resolution affects sensibly the performance of the Spin Image Signatures method. In particular, similarly to all the methods where model descriptions are directly obtained from the analysis of mesh vertices, Spin

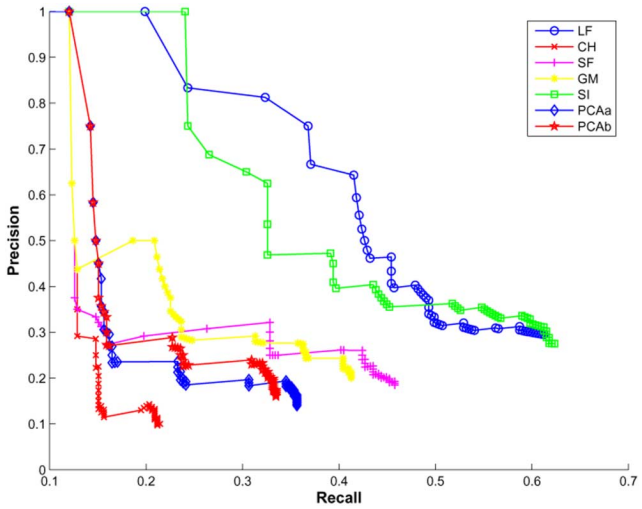

Fig. 8. Precision/recall figures for Light Field (LF), Curvature Histograms (CH), Shape Functions (SF), Geometric Moments (GM), Spin Image Signatures (SI), and PCA Spin Images (PCAa, PCAb) using the Princeton-subset database.

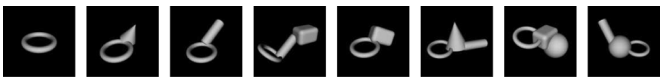

Fig. 9. Example of retrieval by part similarity: ring model used as a query.

Image Signatures has very low performance at low resolutions. In contrast, the Light Field approach, that instead takes into account model views from external viewpoints, is only marginally affected by changes in model resolution. We can also observe that the Spin Image method in its original formulation based on PCA, performs definitely worse than Spin Image Signatures at any model resolution. This is basically due to the fact that, while preserving the statistical characterization of the data distribution, PCA does not guarantee separation between model classes. Average performance figures for the Princeton-subset database for the sample queries used are shown in Fig. 8.

B. Retrieval by Part Similarity

Following the idea presented in [20], we have implemented a 3-D object segmentation module. 3-D models are processed to extract their constituent parts and Spin Image Signatures are obtained both for the model and for each of its parts. Fig. 9 gives a qualitative view of the capability of Spin Image Signatures to support retrieval by parts of 3-D models. In the example shown, a simple ring model is used as a query. All the top ranked models retrieved are a combination of the ring model with other models. Fig. 10 shows the precision-recall curve of Spin Image Signatures for retrieval by part similarity, averaged over queries using primitive shapes. Experiments have been made with the primitive-shape database.

C. Sensitivity to Shape Deformations

The Spin Image Signatures method provides good robustness to geometric deformations of 3-D object models. We have derived performance figures for several geometric deformations,

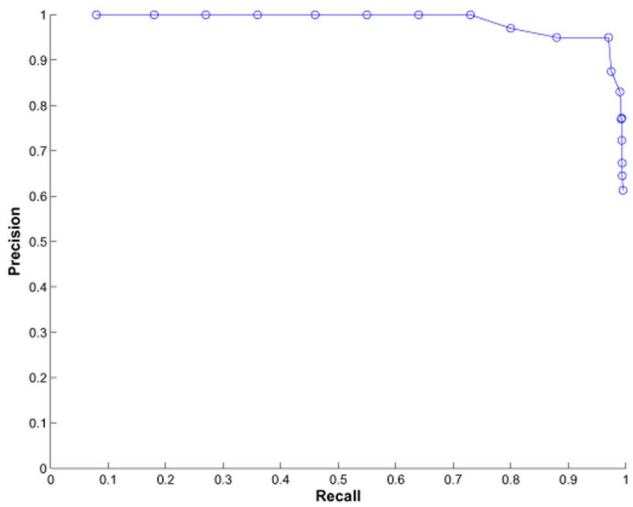

Fig. 10. Average precision-recall of Spin Image Signatures for retrieval by part similarity (Primitive-shape database).

using the Extended-art database. Mesh modifier functions analyzed were bending, relaxation, skewing, spherification (as provided in *3ds MAX*), and Gaussian noise deformation. The first four modifiers induce modifications that affect the structure of the model as a whole; the Gaussian noise modifier induces instead local perturbations, affecting each vertex independently. In particular

- — *Bending* produces a uniform bend in the object's geometry, up to 360 degrees with regard to a single axis.
- — *Relaxing* moves mesh vertices closer to, or away from, their neighbors, performing some model smoothing.
- — *Skewing* produces a uniform offset in the object's geometry, on one of three axes.
- — *Spherification* deforms a model into a spherical shape.
- — *Gaussian noise deformation* adds displacements to mesh vertices, according to a Gaussian distribution.

The effects of these modifiers are displayed in Fig. 11 for a sample 3-D model.

Precision and recall figures were computed for each modifier, in order to analyze the relationship between the amount of deformation applied to the model and the extent to which the Spin Image Signatures method supports retrieval effectively. Results are shown in Fig. 12. Drops of precision versus recall curves indicate the values of the deformation at which characteristic elements that are proper of the object type are no longer recognized in the model geometry. From Fig. 12, it can be observed that different deformations affect retrieval performance differently. For bending and relaxing, only a smooth degradation of performance is observed as the degree of the deformation applied increases. Differently, for skewing and spherification modifiers, some loss of performance can be observed also at moderate degrees of deformations applied. In fact, skewing and spherification are such that deformed models are soon transformed into models with very different shapes, respectively elongated and spherical. In that the database includes many models with significant elongation along a single axis, and many spherical models, in both cases, the description of the deformed model

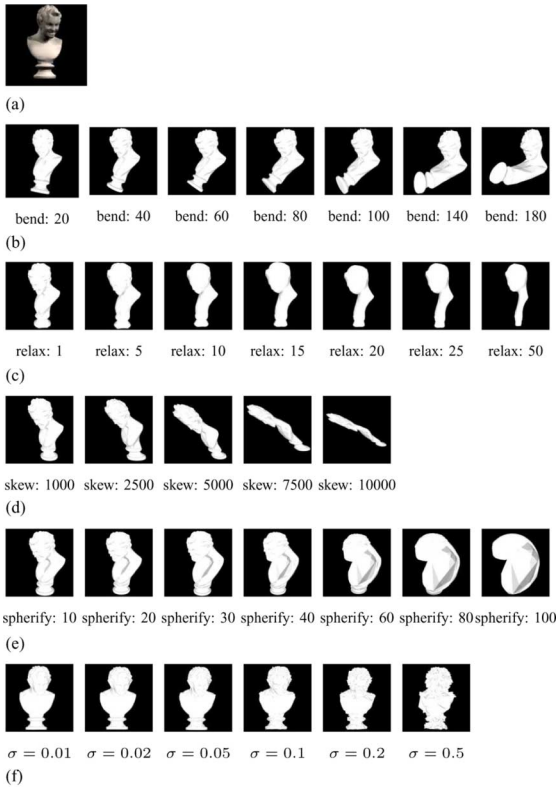

Fig. 11. Effects of different degrees of (b) bending, (c) skewing, (d) relaxation, (e) spherification, and (f) Gaussian noise deformation modifiers on a sample model (a).

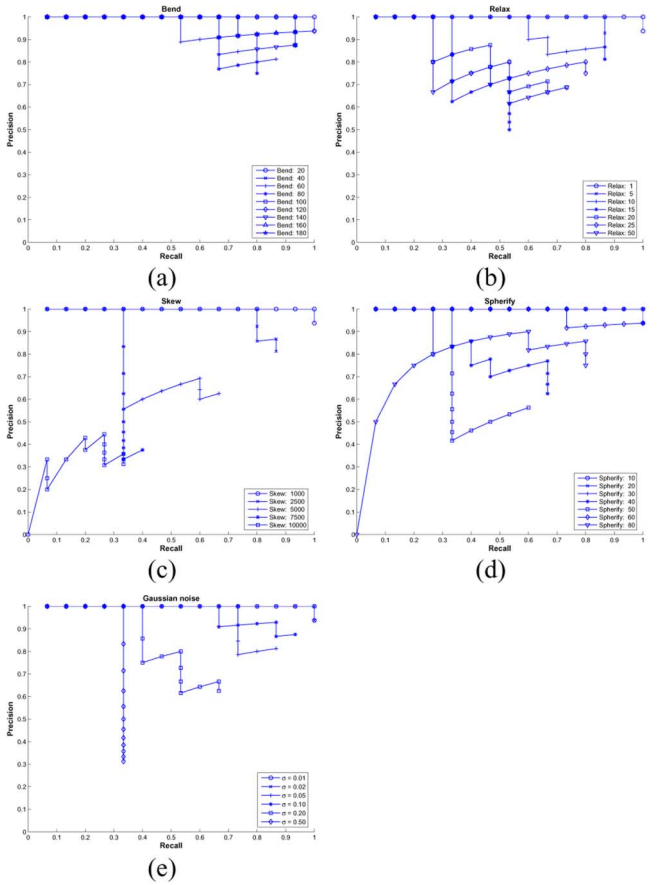

Fig. 12. Sensitivity to deformations introduced through bending, relaxing, skewing, spherification, and Gaussian noise deformation.

results closer to the descriptions of different models than to the original model.

For Gaussian noise deformation [see Fig. 12(e)], Spin Image Signatures shows good performance at small-moderate values of Gaussian noise intensity. The loss of performance at high noise levels can be explained with the fact that high noise produces substantial changes in the position of mesh vertices and, consequently, in the orientation of surface normals that are used to derive spin images. A similar behavior was also reported for the original Spin Image method by [26].

D. Computational Complexity and Storage Requirements

Computational requirements of the Spin Image Signatures method are analyzed separately for archiving and retrieval. In fact, the process of signature clustering permits efficient retrieval at the expense of increased cost at archival time. With the Spin Image Signatures method, adding a new model to the database requires: 1) computation of one spin image for each vertex of the 3-D model mesh; 2) computation of the Spin Image Signatures for the set of spin images obtained; and 3) clustering of Spin Image Signatures. If n_v is the number of mesh vertices, ns, np, and nn are the number of spin image sectors and crowns (positive and negative), T is the number of iterations of clustering, and c is the number of clusters obtained, the process of spin images extraction is $O\left(n_v^2\right)$; computation of the Spin Image Signature is $O((ns+np+nn)(n_v))$; and computational costs of clustering can be approximated to $O(\sum_c(ns + np + nn)Tcn_v)$ ([27]). Therefore, the overall computational complexity of the Spin Image Signatures method at archival time can be estimated as $O\left(n_v^4\right)$.

For retrieval, two different cases must be considered. If the model used for querying is already available in the database, assuming that retrieval is performed with sequential scanning with no index structure, if N is the number of database models and $\overline{c}$ the average number of clusters, computational complexity of retrieval can be estimated as $O(N\,\overline{c}!)$. If the query model is new, its Spin Image Signature must be obtained to perform matching. This implies an additional computational cost of $O\left(n_v^4\right)$.

The average computational costs per model, computed for a Pentium IV (2.6-GHz clock and 1.2-GB RAM) machine, are reported in Table III for both Spin Image Signatures and the original PCA-based Spin Image method. Costs of extraction and clustering are computed for the entire archive and normalized with respect to the number of models in the archive. Costs of retrieval are obtained by averaging over 100 model comparisons.

Storage space requirements for the Spin Image Signatures method, and the other methods analyzed are as follows. The Spin Image Signature method requires $18 * 4$ bytes to store values of each cluster center, each cluster center requiring a 18-D array of 4-bytes float values. The overall storage cost is therefore $c * 18 * 4$ bytes, c being the number of cluster centers. The PCA Spin Images method requires $n_v * s * 4$ bytes to store the PCA reduced spin images of the model, where n_v

TABLE III
COMPARISON OF COMPUTATIONAL COSTS FOR ARCHIVING AND RETRIEVAL OF SPIN IMAGE SIGNATURES AND THE ORIGINAL PCA-BASED SPIN IMAGE METHOD. AVERAGE VALUES ON PENTIUM IV (2.6-GHZ CLOCK AND 1.2-GB RAM) MACHINE

	Spin Image Signatures	*PCA-based spin images*
Descriptor extraction	22.4 s	27.6 s
Descriptor matching	0.6 ms	2.1 s

TABLE IV
MEMORY SPACE REQUIREMENTS OF THE DIFFERENT METHODS FOR A DATABASE OF 1000 MODELS WITH 2000 VERTICES PER MODEL, ON AVERAGE

Approach	*Memory*
Spin Image Signatures	0.43MBytes
PCAa Spin Images	864MBytes
PCAb Spin Images	144MBytes
Light Field	4.5MBytes
Shape Functions	0.51MBytes
Curvature Histograms	0.51MBytes
Geometric Moments	0.22MBytes

is the number of model vertices and s is the number of relevant PCA component that are not discarded (108 for PCAa and 18 for PCAb in our experiments), and each PCA reduced spin image is represented as an s-dimensional array of 4-bytes float values. The Light Field method requires $(35+10)$ bytes to store values of 35- and 10-D 1-byte integer vectors used to encode Zernike moments and Fourier descriptors of each silhouette. Since for each model 100 silhouettes are extracted from different viewpoints, the overall storage cost is $(35+10)*100$ bytes. The Curvature Histograms method requires $N_{\text{curv}}*4$ bytes to store the N_{curv}-dimensional array of 4-bytes float values representing the relative frequency of curvature values quantized into N_{curv} levels. The Shape Functions method needs $N_{\text{quant}}*4$ bytes to store the N_{quant}-dimensional array of 4-bytes float values representing the relative frequency of chord-length values quantized into N_{quant} levels. The Geometric Moments approach requires instead $54*$ 4 bytes to store values of all the 54 different 3-D moments up to the sixth order, each moment being represented through a 4-byte float value.

Table IV shows a comparative estimation of the memory space requirements (in bytes) assuming a database of 1000 models with 2000 vertices per model, in average. Parameters used are $c=6$, $N_{\text{curv}}=128$ and $N_{\text{quant}}=128$.

It can be observed that all the methods, except the PCA-based Spin Image method, have comparable memory space requirements. The PCA-based Spin Image method has memory space requirements that do not scale with the number of database models. This makes the method unsuitable for large databases.

VI. CONCLUSIONS

In this paper, we have presented a new method for 3-D content-based retrieval based on Spin Image Signatures. The method develops from the Spin Images approach, originally conceived for the purpose of 3-D object recognition, but defines new solutions to obtain a more concise, yet effective, description. Experimental results reported have shown that the Spin Image Signatures approach is well suited for the purpose of 3-D model retrieval and provides superior performance with respect to the other methods in the literature particularly at medium-high resolutions. It also showed good insensitivity to geometric deformations like bending and relaxing of the model mesh. Future work is to address relationships between spin images, Spin Image Signatures, and object parts, in order to support a more effective retrieval by part similarity. Indexing models will also be studied to improve the efficiency of retrieval.

REFERENCES

[1] J. Assfalg, A. Del Bimbo, and P. Pala, "Curvature maps for 3D CBR," in *Proc. Int. Conf. Multimedia and Expo (ICME'03)*, Baltimore, MD, Jul. 2003.

[2] S. Berchtold and H. P. Kriegel, "S3: Similarity search in CAD database systems," in *Proc. the ACM SIGMOD Int. Conf. Management of Data*, Tucson, AZ, May 1997, pp. 564–567.

[3] J. C. Bezdek, J. Keller, R. Krishnapuram, and N. R. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Boston, MA: Kluwer, 1999.

[4] A. Del Bimbo, *Visual Information Retrieval*. New York: Academic, 1999.

[5] M. Elad, A. Tal, and S. Ar, "Content based retrieval of VRML objects—An iterative and interactive approach," *EG Multimedia*, pp. 97–108, Sep. 2001.

[6] T. T. Elvins and R. Jain, "Web-based volumetric data retrieval," in *Proc. VRML 95, 1st Annu. Conf. Virtual Reality Modeling Language*, Dec. 1995.

[7] A. Goshtasby, "Description and discrimination of planar shapes using shape matrices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 7, pp. 738–743, 1985.

[8] A. E. Johnson and M. Hebert, "Using spin-images for efficient multiple model recognition in cluttered 3-D scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433–449, 1999.

[9] D.-J. Kim, Y.-W. Park, and D.-J. Park, "A novel validity index for determination of the optimal number of clusters," *IEICE Trans. Inform. Syst.*, vol. E84-D, no. 2, pp. 281–285, Feb. 2001.

[10] H. P. Kriegel and T. Seidl, "Approximation-based similarity search for 3D surface segments," *GeoInformatica J.*, vol. 2, no. 2, pp. 113–147, 1998.

[11] I. Kolonias, D. Tzovaras, S. Malassiotis, and M. G. Strintzis, "Content-based similarity search of VRML models using shape descriptors," in *Proc. Int. Workshop on Content-Based Multimedia Indexing*, Brescia, Sep. 19–21, 2001, vol. I.

[12] Y. Lamdan and H. J. Wolfson, "Geometric hashing: A general and efficient model-based recognition scheme," in *Proc. IEEE Int. Conf. Computer Vision*, Tampa, FL, 1988, pp. 238–249.

[13] S. Mahmoudi and M. Daoudi, "3D models retrieval by using characteristic views," in *Proc. 16th Int. Conf. Pattern Recognition*, Aug. 11–15, 2002, vol. 2, pp. 457–460.

[14] F. Mokhtarian, N. Khalili, and P. Yeun, "Multi-scale free-form 3D object recognition using 3D models," *Image and Vis. Comput.*, vol. 19, no. 5, pp. 271–281, 2001.

[15] R. Ohbuchi, M. Nakazawa, and T. Takei, "Retrieving 3D shapes based on their appearance," in *Proc. MIR'03*, Berkeley, CA, Nov. 2003, pp. 39–46.

[16] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Matching 3D models with shape distribution," in *Proc. Shape Modeling International*, Genova, Italy, 2001, vol. I.

[17] E. Paquet and M. Rioux, "Nefertiti: A query by content system for three-dimensional model and image database management," *Image Vis. Comput.*, vol. 17, no. 2, pp. 157–166, 1999.

[18] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[19] J.-P. Vandeborre, V. Couillet, and M. Daoudi, "A practical approach for 3D model indexing by combining local and global invariants," in *Proc. 1st Int. Symp. 3D Data Processing, Visualization, and Transmission (3DPVT'02)*, 2002.

[20] K. Wu and M. D. Levine, "3D part segmentation using simulated electrical charge distributors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 11, pp. 1223–1235, Nov. 1997.

[21] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, and D. Jacobs, "A search engine for 3D models," *ACM Trans. Graph.*, vol. 22, no. 1, pp. 83–105, Jan. 2003.

[22] C. Zhang and T. Chen, "Efficient feature extraction for 2D/3D objects in mesh representation," in *Proc. ICIP 2001*, Thessaloniki, Greece, 2001.
[23] J. W. H. Tangelder and R. C. Veltkamp, "A survey of content based 3D shape retrieval methods," in *Proc. Int. Conf. Shape Modeling and Applications 2004*, Genova, Italy, Jun. 7–9, 2004, pp. 145–156.
[24] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," *ACM Trans. Graph.*, vol. 21, no. 4, pp. 807–832, Oct. 2002.
[25] D. Y. Chen, X. P. Tian, Y. T. Shen, and M. Ouhyoung, "On visual similarity based 3D model retrieval," *Eurographics, Comput. Graph. Forum (EG 2003 Proceedings)*, vol. 22, no. 3, 2003.
[26] A. Frome, D. Huber, R. Kolluri, T. Bulow, and J. Malik, "Recognizing objects in range data using regional point descriptors," in *Proc. European Conf. Computer Vision (ECCV)*, May 2004.
[27] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley.
[28] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 24, pp. 509–522, 2002.
[29] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton shape benchmark," in *Proc. Shape Modeling International*, Genova, Italy, 2004.
[30] H. F. Kaiser, "The application of electronic computer to factor analysis," *Educ. Psychol. Meas.*, vol. 20, pp. 141151–141151, 1960.

Jürgen Assfalg received the Laurea degree in electronics engineering in 1998 and the Ph.D. degree in information and telecommunications engineering in 2002, both from the Università di Firenze, Firenze, Italy.

He was then a Research Associate at the Visual Information Processing Lab until the end of 2003. He is now with the local government of the Provincia di Firenze. His current research interests include multimedia retrieval systems, advanced user interfaces, and software architectures.

Dr. Assfalg is a member of the ACM.

Marco Bertini received the M.S. degree in electronic engineering in 1999 and the Ph.D. degree in 2004, both from the Università di Firenze, Firenze, Italy.

He holds a research grant and carries out his research activities at the Department of Systems and Informatics, University of Florence. His main research interest is content-based indexing and retrieval of videos. He is the author of more than 50 papers in international conference proceedings and journals, and is a reviewer for international journals on multimedia and pattern recognition.

Alberto Del Bimbo (M'90) is a Full Professor of computer engineering at the Università di Firenze, Firenze, Italy. Since 1998, he has been the Director of the Master's degree in Multimedia. His scientific interests include pattern recognition, image databases, multimedia and human computer interaction. He is the author of over 170 publications in the most distinguished international journals and conference proceedings. He is author of the Visual Information Retrieval monography (New York: Morgan Kaufman) on content-based retrieval from image and video databases.

Dr. Del Bimbo is a Fellow of the International Association for Pattern Recognition (IAPR). He is presently an Associate Editor of *Pattern Recognition*, the *Journal of Visual Languages and Computing*, the *Multimedia Tools and Applications Journal*, and *Pattern Analysis and Applications*. He was an Associate Editor of IEEE TRANSACTIONS ON MULTIMEDIA and IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. He was the Guest Editor of several special issues on image databases in highly respected journals.

Pietro Pala (M'92) graduated in electronics engineering in 1994 and received the Ph.D. degree in information and telecommunications engineering in 1997, both from the Università di Firenze, Firenze, Italy.

He is currently an Associate Professor of Computer Engineering at the Università di Firenze, where he teaches the courses "Database Management Systems", and "Image and Video Processing and Analysis". His main scientific and research interests include pattern recognition, image and video databases, and multimedia information retrieval. He is Vice Chair of the IAPR Technical Committee 12 on Multimedia and Visual Information Systems.