3D CONTENT-BASED RETRIEVAL WITH SPIN IMAGES

Jürgen Assfalg, Gianpaolo D’Amico, Alberto Del Bimbo, Pietro Pala

Dipartimento Sistemi e Informatica
Università di Firenze
{assfalg, damico, delbimbo, pala}@dsi.unifi.it

ABSTRACT

Along with images and videos, 3D models have recently gained increasing attention for a number of reasons: advancements in 3D hardware and software technologies, their ever decreasing prices and increasing availability, affordable 3D authoring tools, and the establishment of open standards for 3D data interchange.

The ever increasing availability of 3D models demands for tools supporting their effective and efficient management. Among these tools, those enabling content-based retrieval play a key role.

In this paper we present a novel approach to 3D content-based retrieval that is based on spin images. Spin images are used to derive a view-independent description of both database and query objects: a set of spin images is first created for each object; then, a descriptor is evaluated for each spin image in the set; clustering is performed on the set of image-based descriptors of each object to achieve a compact representation of the object, thus allowing for efficient indexing and matching.

Experimental results are presented for a test database of about 300 models. These results indicate that spin images can be successfully exploited for content-based retrieval of 3D objects.

1. INTRODUCTION

Content-based retrieval of 3D information is a new challenge for researchers and practitioners. Development of techniques supporting archival, indexing and retrieval of 3D models is of paramount importance in a variety of application domains. For instance, this is particularly the case in the fields of cultural heritage and historical relics, where there is an increasing demand for solutions enabling preservation of relevant artworks (e.g. vases, sculptures, handicrafts) as well as cataloguing and retrieval by content.

Tools supporting retrieval of 3D models are also expected to play a key role in educational programs, either traditional or computer-based.

In this paper we address the problems of description and matching of 3D objects. Inherent difficulties in this field concern the representation of objects, the spatial transformations that may affect them, self-occlusions, and varying levels of detail. We present a solution based on the use of spin images ([4]) to capture the global shape of 3D objects in a view independent manner. Since object description based on spin images entails a huge amount of information, feature extraction and clustering techniques are used to meet the specific storage and efficiency requirements of content-based retrieval. By relying on spin images, we provide for an object-centred description, which is insensitive to rigid transformations, and which can leverage achievements in the field of content-based image retrieval.

The paper is organised as follows: Sec.2 surveys related work in the field of 3D CBR; Sec. 3 introduces object description based on spin images; Sec. 4 expounds on feature extraction and clustering; then, in Sec. 5 experimental results are presented; finally, in Sec. 6 conclusions are drawn.

2. RELATED WORK

Description and retrieval of 3D objects based on description and retrieval of 2D views has been addressed in [9] and [11]. However, the effectiveness of these solutions is limited to description and retrieval of simple objects. In fact, as complex objects are considered, occlusions prevent to capture distinguishing 3D features using 2D views.

Description of 3D surface data for the purpose of recognition or retrieval has been addressed for some time. A few authors have investigated analytical 3D models, but this is not always a viable solution, as there are many limitations in providing parameterizations of arbitrary models. In [7] retrieval of 3D objects based on similarity of surface segments is addressed. Surface segments model potential docking sites of molecular structures. The proposed approach develops on the approximation error of the surface. However, assumptions on the form of the function to be approximated limit applicability of the approach to special contexts.

Much attention has been recently devoted to free-form (i.e. polygonal) meshes. While this representation of 3D models poses major hurdles to development and implementation of algorithms, it is indeed the most appealing field of application. The system developed within the Nefertiti project supports retrieval of 3D models based on both geometry and appearance (i.e. colour and texture) [12]. Also Kolonias et al. have used dimensions of the bounding box (i.e. its aspect ratios) and a binary voxel-based representation of geometry [8]. They further relied on a third feature, namely a set of paths, outlining the shape (model routes). In [10] a method is proposed to select feature points which relies on the evaluation of Gaussian and median curvature maxima, as well as of torsion maxima on the surface. In [5], Elad et al. use moments (up to the 4-7th order) of surface points as basic features to support retrieval of 3D models. Differently from the case of 2D images, evaluation of moments is not affected by (self-)occlusions.

In [1] description and retrieval of 3D objects is accomplished through a combination of warping and projection. This method captures prominent geometric features of 3D objects in a view independent manner. However, it can be applied only to objects whose surface defines the boundary of a simply connected 3D region. Moreover, warping may introduce irregular deformation of the object surface before its projection on a 2D map.
3. SPIN IMAGES

Spin images were introduced by Johnson and Hebert to support recognition of single objects in complex scenes [4]. Basically, spin images encode the density of mesh vertices projected onto an object-centred space: the three-dimensional mesh vertices are first mapped onto a two-dimensional space defined w.r.t. to the object itself; the resulting coordinates are then used to build a two-dimensional histogram.

More precisely, let $O = (p, n)$ an oriented point on the surface of the object, where $p$ is a point on the surface of the object and $n$ the normal of the tangent plane in $p$. For a generic oriented point $O$, a spin map can be defined, which maps any point $x$ in the three-dimensional space onto a two-dimensional space according to the following formula (see also Fig. 1 for notation):

$$S_O(x) \rightarrow [\alpha, \beta] = \left[\sqrt{||x - p||^2 - (n \cdot (x - p))^2}, n \cdot (x - p)\right]$$

In other words, the oriented point defines a family of cylindrical coordinate systems, with the origin in $p$, and with the axis along $n$. The spin map projection of $x$ retains the radial distance ($\alpha$) and the elevation ($\beta$), while it discards the polar angle. In so doing, the ambiguity is resolved which results from the fact that the oriented point does not define a unique cylindrical coordinate system. This projection ensures that, for any given oriented point, a unique spin map exists.

To produce a spin image of an object, a spin map is applied to points comprising the surface of the object. Hence, given a mesh representation of the object, the spin image can be obtained by applying the map to the vertices comprising the mesh. A simple binary image can be obtained by discretizing the projected coordinates and by setting the corresponding point on the image. However, more refined grey-level spin images encoding a measure of the density of vertices that insist upon the same image point are usually employed. To construct such an image, the projected coordinates $\alpha$ and $\beta$ of each mesh vertex are used to update the two-dimensional histogram $I(i, j)$ (i.e. the spin image) according to a bi-linear interpolation scheme that spreads the contribution of each vertex over the nearest points on the grid induced by the quantization of the image space. A sample model and one of its spin images are shown in Fig. 2.

Most outstanding characteristics of spin images are invariance to rigid transformations (as a consequence of the adoption of an object-centred coordinate system), limited sensitivity to variations of position of mesh vertices (which might result from the adoption of different sampling schemes), flexibility (since no hypotheses are made on the surface representation), and ease of computation.

4. DESCRIPTION, CLUSTERING AND MATCHING

4.1. Description

Spin images provide a powerful means to describe three-dimensional objects. However, the fact that many spin images are typically produced for a single object, and the fact that each image implies considerable storage requirements, prevent us to use them directly as descriptors for retrieval purposes. Therefore, we decided to rely on more compact descriptions extracted from spin images, synthesizing the content of each spin image.

Although a spin image is actually a 2D histogram, it can be regarded as a grey-level image. We therefore decided to investigate the usage of descriptors that have been used to describe images for the purpose of content-based image retrieval. In particular, our descriptor for spin images was inspired by region descriptors such as grid-based techniques or the shape matrix [3]. The latter approach appears particularly appealing because it is invariant to translation, rotation, and scaling. Instead of sampling the shape at the intersection between radial and circular lines, we decided to measure the relative density encompassed by each of the regions defined by those lines, so as to provide a more precise description of the spin image. However, exploratory experiments with such a descriptor displayed a poor performance, since the adopted scheme was too rigid, and did not effectively support retrieval of similar objects. We have therefore defined three independent sets of regions for the spin image: sectors of circular

![Figure 2: A 3D object, along with a spin map and a spin image of the object.](image)

![Figure 3: The attenuation function $a(d)$; $d$ is the distance of a given point from the reference point, normalized w.r.t. to the size of the bounding box of the object. In our experiments, $a_s = 0.3$ and $a_c = 0.5$.](image)
crowns for both the half-plane with $\beta > 0$ and the half plane with $\beta < 0$, and circular sectors. Each of these sets defines a descriptor $(C^p = \{c_{p1}, \ldots, c_{pn_p}\}, C^n = \{c_{n1}, \ldots, c_{nn}\}$, and $S = \{s_{n1}, \ldots, s_{ns}\}$, respectively, whose components represent the amount of surface points (or vertices) whose projections fall within the corresponding crown/sector:

$$
c_{pk} = \frac{\sum_{i,j} I(i,j) s_{pk}}{\sum_{i,j} I(i,j)} \quad k = 1, \ldots, n_p
$$

$$
c_{nk} = \frac{\sum_{i,j} I(i,j) s_{nk}}{\sum_{i,j} I(i,j)} \quad k = 1, \ldots, n_n
$$

$$
s_{sk} = \frac{\sum_{i,j} I(i,j) s_{sk}}{\sum_{i,j} I(i,j)} \quad k = 1, \ldots, n_s
$$

where $SP(k)$ is the $k$-th “positive” crown sector, $SN(k)$ is the $k$-th “negative” crown sector, and $S(k)$ is the $k$-th circular sector (see Fig. 4). The following constraints hold:

$$
\sum_{i=1}^{n_p} c_{pi} \leq 1 \quad \sum_{i=1}^{n_n} c_{ni} \leq 1
$$

$$
\sum_{i=1}^{n_p} c_{pi} + \sum_{i=1}^{n_n} c_{ni} = 1 \quad \sum_{i=1}^{n_s} s_{si} = 1
$$

Based on results of some preliminary experiments we chose $n_p = n_n = n_s = 6$ as these represent a satisfactory trade-off between compactness and selectivity of the representation. Hence, a 18-dimensional descriptor $D = \langle C^p, C^n, S \rangle$ is evaluated for each spin image.

![Compound object descriptors comprise descriptors for a) $n_p$ crowns in the half-plane $\beta > 0$, b) $n_n$ crowns in the half-plane $\beta < 0$, c) $n_s$ sectors. In our experiments $n_p = n_n = n_s = 6$.](image)

### 4.2. Clustering

Clustering algorithms partition the data set into groups such that the clusters describe an underlying structure within the data. Fuzzy clustering [2] is an extension to c-means procedure to avoid partitioning feature vectors into hard or crisp clusters. Furthermore, the use of fuzzy clustering guarantees lower sensitivity of cluster centres to outliers, with respect to the use of traditional hard c-means.

Let $\{D_1, \ldots, D_m\}$ be a set of $m$ descriptors in $\mathbb{R}^{18} \ (D_i \in \mathbb{R}^{18})$. According to the fuzzy c-means algorithms, feature vectors can be clustered into $c$ clusters by minimizing the function:

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

being $\mu \in (1, \infty)$ a weighting exponent that determines the fuzziness of the clusters (known as the fuzzy exponent), $v_j, j = 1, \ldots, c$ the cluster centres, and $u_{ij}$ the membership degree of vector $D_i$ in cluster $v_j$. In order to avoid the trivial solution, values of $u_{ij}$ are subject to the following constraints:

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

By allowing the value of $\sum_{i=1}^{m} u_{ij}$ to equal 0 it is left open the possibility for some clusters to be empty.

Minimization of function (1) is carried out through the following procedure:

1. Randomly set the initial values of $v_j, j = 1, \ldots, c$
2. Update values of the membership functions: $u_{ij} = (d(D_i, v_j))^{-2/\mu-1}$
3. Normalize values of the membership functions: $u_{ij} = \frac{u_{ij}}{\sum_{j=1}^{c} u_{ij}}$
4. Update values of cluster centres: $v_j = \sum_{i=1}^{m} u_{ij} D_i$
5. If the average motion vector of cluster centres is below a predefined threshold then stop. Otherwise, goto Step 2.

Computation of the optimal number of clusters $c_{opt}$ is accomplished according to the approach proposed in [6]. According to it, a Validity Index ($VI$) is defined based on values of mean intra cluster distance and inter cluster minimum distance. To determine the optimal value $c_{opt}$, data clustering is accomplished for several values of $c$. For each value, the validity index is computed; $c_{opt}$ corresponds to the value that minimizes $VI$.

### 4.3. Matching

Computation of the similarity between two 3D objects is accomplished by comparing their descriptors, each descriptor being in the form of a set of cluster centres $D = \{D_i, i = 1, \ldots, k\}$. Computing the distance $\Delta$ between two descriptors $D$ and $\hat{D}$ requires to find the best cluster-centre correspondence function. This is defined as the permutation $p : \{1, \ldots, k\} \rightarrow \{1, \ldots, k\}$ that minimizes the sum of distances between corresponding cluster centres, that is:

$$
\Delta(D, \hat{D}) = \min_p \left\{ \sum_{i=1}^{k} \delta(D_i, \hat{D}_{p(i)}) \right\}
$$

being $D_i$ the $i$-th cluster centre tile in the first descriptor and $\hat{D}_{p(i)}$ the $p(i)$-th cluster centre in the second descriptor.

The solution $p$ to Eq. (2) is approximated through a heuristic search approach that requires to scan all cluster centres for the first object in a predefined order and associate to each centre the most similar cluster centre not yet associated in the second object. This pairwise NN association yields a suboptimal solution.

### 5. EXPERIMENTAL RESULTS

Approximately 300 models were collected to build the test database. These comprise four classes of models: taken from the web, manually authored (with a 3D CAD software), high quality 3D scans.
from the De Espona 3D Models Encyclopedia\footnote{http://www.deespona.com}, and variations of the previous three classes (obtained through deformation or application of noise, which caused points surface to be moved from their original locations). Feature descriptors were then evaluated and added to the index.

Fig. 5 shows a retrieval example where the model of a statue portraying Mercur is used as a query. The result set displays all models of the Mercur statue in the first five positions. Other models retrieved feature similar shapes, characterized by a main body and protrusions that resemble Mercur’s elongated arm and leg.

In Fig. 6, the precision-recall curve of the proposed approach is compared with the curve of the approach based on curvature maps presented in [1] (which outperforms previous approaches based on 3D moments and curvature histograms).

6. CONCLUSIONS AND FUTURE WORK

In this paper we have presented an approach to 3D content-based retrieval developing on spin-images. Experimental results have shown that, besides object recognition, spin images are also suited for the purpose of retrieval. Furthermore, results showed that the approach performs better than previous approaches. Future work will address extension of the approach to retrieval of parts of 3D objects.

7. REFERENCES