

# 3D model retrieval with spherical harmonics and moments

Dietmar Saupe and Dejan V. Vranić  
Institut für Informatik, Universität Leipzig

**Abstract.** We consider 3D object retrieval in which a polygonal mesh serves as a query and similar objects are retrieved from a collection of 3D objects. Algorithms proceed first by a normalization step in which models are transformed into canonical coordinates. Second, feature vectors are extracted and compared with those derived from normalized models in the search space. In the feature vector space nearest neighbors are computed and ranked. Retrieved objects are displayed for inspection, selection, and processing. Our feature vectors are based on rays cast from the center of mass of the object. For each ray the object extent in the ray direction yields a sample of a function on the sphere. We compared two kinds of representations of this function, namely spherical harmonics and moments. Our empirical comparison using precision-recall diagrams for retrieval results in a data base of 3D models showed that the method using spherical harmonics performed better.

## 1 Introduction

Currently methods for retrieving multimedia documents using audio-visual content as a key in place of traditional textual annotation are developed in MPEG-7 [6]. Many similarity-based retrieval systems were designed for still image, audio and video, while only a few techniques for content-based 3D model retrieval have been reported [2, 5–11]. We consider 3D object retrieval in which a 3D model given as a triangle mesh serves as a query key and similar objects are retrieved from a collection of 3D objects. Content-based 3D model retrieval algorithms typically proceed in three steps:

1. *Normalization (pose estimation).* 3D models are given in arbitrary units of measurement and undefined positions and orientations. The normalization step transforms a model into a canonical coordinate frame. The goal of this procedure is that if one chose a different scale, position, rotation, or orientation of a model, then the representation in canonical coordinates would still be the same. Moreover, since objects may have different levels-of-detail (e.g., after a mesh simplification to reduce the number of polygons), their normalized representations should be similar as much as possible.
2. *Feature extraction.* The features capture the 3D shape of the objects. Proposed features range from simple bounding box parameters [8] to complex image-based representations [5]. The features are stored as vectors of fixed dimension. There is a tradeoff between the required storage, computational complexity, and the resulting retrieval performance.
3. *Similarity search.* The features are designed so that similar 3D-objects are close in feature vector space. Using a suitable metric nearest neighbors are

computed and ranked. A variable number of objects are thus retrieved by listing the top ranking items.

We present an empirical study extending our contribution [11] in which we introduced a modification of the Karhunen-Loeve transform and the application of spherical harmonics to the problem of 3D object retrieval. We first review the 3D model retrieval problem and previous work. Then we recall our approach based on spherical harmonics and present an alternative using moments. We describe our experiments that we designed to evaluate and contrast the two competing methods. Finally, the results and conclusions are presented.

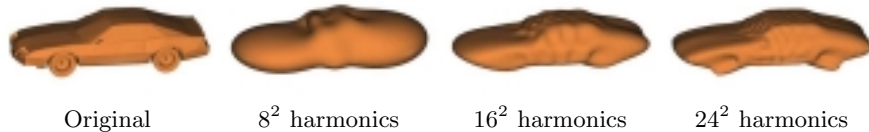
## 2 Previous work

The normalization step is much simpler than the pose estimation deeply studied in computer vision where a 3D pose must be inferred from one or more images, i.e., projections of a 3D object. Here, the 3D models for the retrieval problem are already given in 3D space, and, thus, the most prominent method for normalization is the *principle component analysis* (PCA) also known as the Karhunen-Loeve transform. It is an affine transformation based on a set of vectors, e.g., the set of vertices of a 3D model. After a translation of the set moving its center of mass to the origin a rotation is applied so that the largest variance of the transformed points is along the  $x$ -axis. Then a rotation around the  $x$ -axis is carried out so that the maximal spread in the  $yz$ -plane occurs along the  $y$ -axis. Finally, the object is scaled to a certain unit size. A problem is that differing sizes of triangles are not taken into account which may cause widely varying normalized coordinate frames for models that are identical except for finer triangle resolution in some parts of the model. To address this issue we introduced appropriately chosen vertex weights for the PCA [10], while Paquet et al. [8] used centers of mass of triangles as vectors for the PCA with weights proportional to triangle areas. Later we generalized the PCA so that *all* of the (infinitely many) points in the polygons of an object are equally relevant for the transformation [11].

Feature vectors for 3D model retrieval can be based on Fourier descriptors of silhouettes [1, 5], on 3D moments [8], rendered images or depth maps [5], or on volumetric representation of the model surface [2] or the corresponding volume (if the surface bounds a solid) [7, 6, 8].

Using special moments of 3D objects the normalization step may be joined with the feature extraction. In [3] a complete set of orthogonal 3D Zernike polynomials provides spherical moments with advantages regarding noise effects and with less information suppression at low radii. The normalization is done using 3D moments of degree not greater than 3. There were no examples demonstrating the performance of these feature vectors in 3D model retrieval.

In this paper we consider a particular method to generate feature vectors for 3D object retrieval. In a first step the 3D shape is characterized by a function on the sphere. For this function we empirically compare two kinds of representation, one using spherical harmonics and the other by computing moments.



**Fig. 1.** Multi-resolution representation of the function  $r(\mathbf{u}) = \max\{r \geq 0 \mid r\mathbf{u} \in I \cup \{\mathbf{0}\}\}$  used to derive feature vectors from Fourier coefficients for spherical harmonics.

### 3 Functions on the sphere for 3D shape feature vectors

In this section we describe the feature vectors used in our comparative study. As 3D models we take triangle meshes consisting of triangles  $\{T_1, \dots, T_m\}$ ,  $T_i \subset \mathbb{R}^3$ , given by vertices (geometry)  $\{\mathbf{p}_1, \dots, \mathbf{p}_n\}$ ,  $\mathbf{p}_i = (x_i, y_i, z_i) \in \mathbb{R}^3$  and an index table with three vertices per triangle (topology). Then our object is  $I = \bigcup_{i=1}^m T_i$ , the point set of all triangles. We may assume that our models are normalized by a modified PCA as outlined in Section 1. For details we refer to [11].

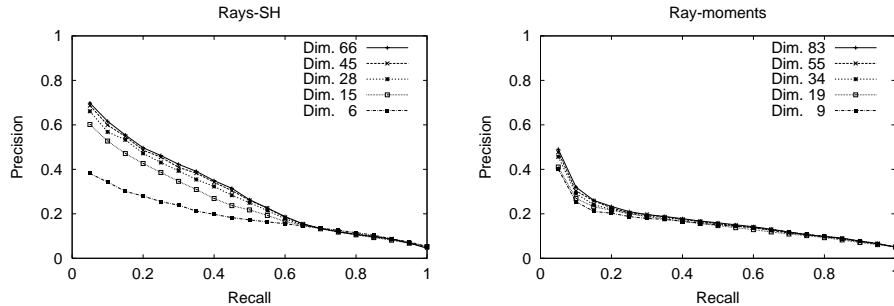
Some feature vectors can be considered as samples of a function on the sphere  $S^2$ . For example, for a (normalized) model  $I$  define

$$r : S^2 \rightarrow \mathbb{R} \\ \mathbf{u} \mapsto \max\{r \geq 0 \mid r\mathbf{u} \in I \cup \{\mathbf{0}\}\}$$

where  $\mathbf{0}$  is the origin. This function  $r(\mathbf{u})$  measures the extent of the object in directions given by  $\mathbf{u} \in S^2$ . In [10] we took a number of samples  $r(\mathbf{u})$  as a feature vector, which, however, is sensitive to small perturbations of the model. In this paper we improve the robustness of the feature vector by sampling the spherical function  $r(\mathbf{u})$  at many points but characterizing the map by just a few parameters, using either spherical harmonics or moments. Other definitions of features as functions on the sphere are possible. For example, one may consider a rendered perspective projection of the object on an enclosing sphere, see [5].

The Fourier transform on the sphere uses the spherical harmonic functions  $Y_l^m$  to represent any spherical function  $r \in L^2(S^2)$  as  $r = \sum_{l \geq 0} \sum_{|m| \leq l} \hat{r}(l, m) Y_l^m$ . Here  $\hat{r}(l, m)$  denotes a Fourier coefficient and the spherical harmonic basis functions are certain products of Legendre functions and complex exponentials. The (complex) Fourier coefficients can be efficiently computed by a spherical FFT algorithm applied to samples taken at points  $\mathbf{u}_{ij} = (x_{ij}, y_{ij}, z_{ij}) = (\cos \varphi_i \sin \theta_j, \sin \varphi_i \sin \theta_j, \cos \theta_j)$ , where  $\varphi_i = 2i\pi/n$ ,  $\theta_j = (2j+1)\pi/2n$ ,  $i, j = 0, \dots, n-1$ , and  $n$  is chosen sufficiently large. We cannot give more details here and refer to the survey and software in [4]. One may use the spherical harmonic coefficients to reconstruct an approximation of the underlying object at different levels, see Figure 1. An example output of the absolute values of the spherical Fourier coefficients (up to  $l = 3$ ) is given here:

			1.161329			
		0.063596	0.162562	0.063596		
	0.213232	0.037139	0.373217	0.037139	0.213232	
0.016578	0.008051	0.009936	0.008301	0.009936	0.008051	0.016578



**Fig. 2.** Precision vs. recall results for varying dimensions in spherical harmonics (left) and moments (right). Results were averaged over all retrieval results in the class of airplane models.

Feature vectors can be extracted from the first  $l + 1$  rows of coefficients. This implies that such a feature vector contains all feature vectors of the same type of smaller dimension, thereby providing an embedded multi-resolution approach for 3D shape feature vectors. We have chosen to use only the absolute values as components of our feature vectors. Because of the symmetry in the rows of the coefficients (for real functions on the sphere coefficients in rows are pairwise complex conjugate) we therefore obtain feature vectors of dimension  $\sum_1^{l+1} k = (l + 1)(l + 2)/2$  for  $l = 0, 1, 2, \dots$ , i.e., 1, 3, 6, 10, 15, and so forth.

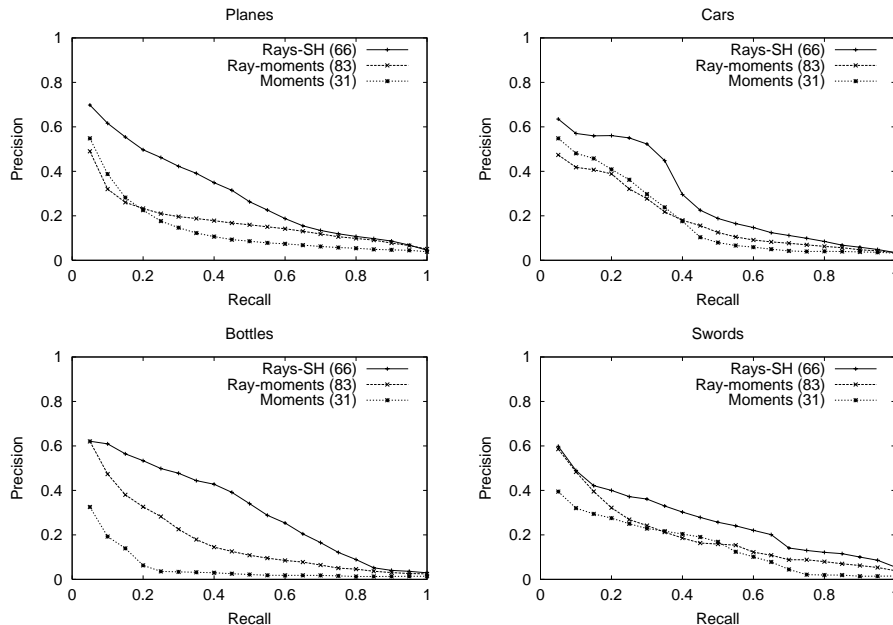
An alternative to the representation of a spherical function by spherical harmonics is given by moments. To be consistent we sample the spherical function  $r(\mathbf{u})$  at the same  $n^2$  points  $\mathbf{u}_{ij}$ ,  $i, j = 0, \dots, n - 1$ , as for the representation by spherical harmonics. As moments we define

$$M^{q,r,s} = \sum_{i,j=0}^{n-1} r(\mathbf{u}_{ij}) \Delta s_{ij} x_{ij}^q y_{ij}^r z_{ij}^s$$

for  $q, r, s = 0, 1, 2, \dots$ . The factor  $\Delta s_{ij}$  represents the surface area on the sphere corresponding to the sample point  $\mathbf{u}_{ij} = (\cos \varphi_i \sin \theta_j, \sin \varphi_i \sin \theta_j, \cos \theta_j)$  and compensates for the nonuniform sampling. For example, when  $n = 128$  we have  $\Delta s_{ij} = \frac{\pi}{64} (\cos(\theta_i - \frac{\pi}{256}) - \cos(\theta_i + \frac{\pi}{256}))$ . For the feature vector we ignore  $M^{0,0,0}$ , and use  $1 \leq q + r + s \leq m$ . As  $m$  grows from 2 to 6 the dimension of the corresponding feature vectors increases from 9 to 19, 34, 55, and 83 (the dimension is  $(m + 1)(m + 2)(m + 3)/6 - 1$ ).

## 4 Results and conclusion

For our tests we collected a data base of 1829 models which we manually classified. For example, we obtained 52 models of cars, 68 airplanes, 26 bottles, and 28 swords. On average a model contains 5667 vertices and 10505 triangles. We



**Fig. 3.** Precision vs. recall results for four classes (airplanes, cars, bottles, and swords) using three methods, the ray-based feature vectors with spherical harmonics (Rays-SH), with moments (Ray-moments), and a method based on statistical moments [8] (Moments). The dimensions of the feature vectors are shown in the legends in brackets.

used  $n^2 = 128^2 = 16384$  samples  $r(\mathbf{u}_{ij})$  of the spherical function for the computation of the spherical harmonics and the moments. For the nearest neighbor computation in feature vector space we used the  $l_1$ -distance.

The retrieval performance can be expressed in so-called precision-recall diagrams. Briefly, *precision* is the proportion of retrieved models that are relevant (i.e., in the correct class) and *recall* is the proportion of the relevant models actually retrieved. By increasing the number of nearest neighbors in the retrieval the recall value increases while the precision typically decreases. By examining the precision-recall diagrams for different queries (and classes) we obtained a measure of the retrieval performance. For our tests we selected one class of objects (e.g., cars) and used each of the objects in the class as a query model. The precision-recall values for these experiments were averaged and yielded one curve in the corresponding diagram.

In our first test series we studied the dependency of the retrieval performance on the dimensionality of the feature vectors. The class of objects was given by the 68 airplanes. We conclude that both types of ray-based feature vectors yield a better performance when the dimension is increased.

In our second test series we compared the performance of the feature vectors for retrieving 3D models in four classes. In all cases the representation of the ray-based feature vector using spherical harmonics performed best, see Figure 3.

The graphs also include results for feature vectors based on statistical moments from [8], defined as  $M^{q,r,s} = \sum_i S_i x_i^q y_i^r z_i^s$  where the point  $(x_i, y_i, z_i)$  is the centroid of the  $i$ -th triangle and  $S_i$  is the area of that triangle. Due to the normalization the moments with  $q + r + s \leq 1$  are zero and can be omitted in the moment feature vectors. The retrieval performance was tested for several dimensions, and we found that the performance decreased for dimensions larger than 31. As shown in Figure 3 the retrieval performance of these feature vectors was inferior to that produced by the ray-based feature vectors.

To conclude we summarize that the Fast Fourier Transform on the sphere with spherical harmonics provides a natural approach for generating embedded multi-resolution 3D shape feature vectors. In tests using a ray-based feature vector the representation with spherical harmonics performed better than a representation using moments.

## References

1. K. Arbter, W.E. Snyder, H. Burkhardt, G. Herzinger, "Application of affine-invariant Fourier descriptors to recognition of 3-D objects", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12 (1990) 640–647.
2. M. Ankerst, G. Kastenmüller, H.-P. Kriegel, T. Seidl, "3D shape histograms for similarity search and classification in spatial databases", *Proc. 6th Intern. Symp. on Spatial Databases (SSD'99)*, Hong Kong, China, Springer-Verlag, 1999.
3. N. Canterakis, "3D Zernike moments and Zernike affine invariants for 3D image analysis and recognition", *Proc. 11th Intern. Conf. on Image Analysis.*, Kangerlussuaq, Greenland, June 1999.
4. D.M. Healy, D. Rockmore, P. Kostelec, and S. Moore, "FFTs for the 2-sphere — Improvements and variations," *Advances in Applied Mathematics*, (to appear). Preprint and corresponding software, *SpharmonicKit*, are available at: <http://www.cs.dartmouth.edu/~geelong/sphere/>.
5. M. Heczko, D. Keim, D. Saupe, and D.V. Vranić, "A method for similarity search of 3D objects", *Proc. of BTW 2001*, Oldenburg, Germany, pp. 384–401, 2001.
6. MPEG Video Group, "MPEG-7 Visual part of eXperimentation Model (version 9.0)," Doc. ISO/MPEG N3914, Pisa, January, 2001.
7. M. Novotni and R. Klein, "A geometric approach to 3D object comparison," *Proc. of Shape Modelling International (SMI 2001)*, Genova, Italy, 2001, (to appear).
8. E. Paquet, A. Murching, T. Naveen, A. Tabatabai, and M. Rioux, "Description of shape information for 2-D and 3-D objects," *Signal Processing: Image Communication*, 16:103–122, 2000.
9. M.T. Suzuki, T. Kato, and N. Otsu, "A similarity retrieval of 3D polygonal models using rotation invariant shape descriptors," *IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC2000)*, Nashville, Tennessee, pp. 2946–2952, 2000.
10. D.V. Vranić and D. Saupe, "3D model retrieval," *Proc. of Spring Conf. on Comp. Graph. and its Appl. (SCCG2000)*, Budmerice, Slovakia, pp. 89–93, May 2000.
11. D.V. Vranić and D. Saupe, "Tools for 3D-object retrieval: Karhunen-Loeve Transform and spherical harmonics," submitted to *IEEE Workshop on Multimedia Signal Processing (MMSP'2001)*, Cannes, France, Oct. 2001.