TOOLS FOR 3D-OBJECT RETRIEVAL: KARHUNEN-LOEVE TRANSFORM AND SPHERICAL HARMONICS

D.V. Vranić, D. Saupe, and J. Richter
Department of Computer Science, University of Leipzig, Leipzig, Germany
phone +49 (341) 973-2251, fax +49 (341) 973-2252, email saupe@acm.org

Abstract - We present tools for 3D object retrieval in which a model, 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 (pose estimation) in which models are transformed into a canonical coordinate frame. Second, feature vectors are extracted and compared with those derived from normalized models in the search space. Using a metric in the feature vector space nearest neighbors are computed and ranked. Objects thus retrieved are displayed for inspection, selection, and processing. For the pose estimation we introduce a modified Karhunen-Loeve transform that takes into account not only vertices or polygon centroids from the 3D models but all points in the polygons of the objects. Some feature vectors can be regarded as samples of functions on the 2-sphere. We use Fourier expansions of these functions as uniform representations allowing embedded multi-resolution feature vectors. Our implementation demonstrates and visualizes these tools.

INTRODUCTION AND PREVIOUS WORK

Objects in databases traditionally have been accessed using attached information such as textual annotation. Recently, methods for retrieving multimedia documents using audio-visual content as a key are developed and standardized in MPEG-7 [3]. 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, 3, 4, 5, 6, 7]. In this paper we discuss two tools for 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 in unpredictable positions and orientations in 3D-space. The normalization step transforms model into a canonical coordinate frame. The goal of this procedure is that if one chose a different scale, position, rotation, or orientation of an original model, then

the representation in the canonical coordinate frame 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 the same as much as possible. The normalization step ensures that models can be retrieved regardless of the choices their authors have made for their mesh representation.

2. *Feature extraction.* The features capture the 3D shape of the objects. Proposed features range from simple bounding box parameters [5] to complex image-based representations [2]. Usually, the features are stored as vectors with real-valued components and 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 attributed vectors that 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.

There have been several approaches for the normalization step, the most prominent one being the *princple component analysis* (PCA) that produces an affine transformation of space, that is also known as the Karhunen-Loeve or Hotelling transform. The transform is defined by 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 of the coordinate system a rotation is applied so that the largest spread of the transformed points (the variance) 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. Essentially, that is the approach taken in [3]. A serious 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. As a solution to this issue we introduced appropriately choosen vertex weights for the PCA [7], while Paquet et al. [5] used centers of gravity of triangles as vectors for the PCA with weights proportional to triangle areas. Such methods improve retrieval results, see Figure 2.

The shape descriptor in [6] is invariant only with respect to rotations of 90 degrees around coordinate axes. The invariance was attained using a well known general principle. Any feature vector can be made invariant with respect to a finite group of transformations of space by summing or averaging feature vectors computed from all possible transformations of an object.

In [4] the pose estimation is based on moments of solid objects. However, 3D models are not guaranteed to consist of closed surfaces bounding one or more solids, and it would be a difficult and questionable undertaking to enforce objects to be solids by stitching up surfaces with boundaries. Therefore, the approach is suitable only for a small class of 3D models.

In this paper we propose two tools for Steps 1 and 2 of any algorithm following the general layout. For the pose estimation previous we generalize the Karhunen-Loeve transform so that *all* of the (infinitely many) points in the polygons of an object are equally relevant for the transformation. For the feature vectors we notice that a class of them can be regarded as taking samples of functions on the 2-sphere. Using Fourier expansions of these functions provides a new uniform approach which facilitates embedded multi-resolution feature vectors. Our implementation demonstrates and visualizes these tools.

CANONICAL COORDINATE FRAME

In this section we outline the details for our continuous PCA and the associated Karhunen-Loeve transform. We regard a given triangle mesh as consisting of a set of triangles $T = \{T_1, \ldots, T_m\}$, $T_i \subset \mathbb{R}^3$, given by a set of vertices (geometry) $P = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$, $\mathbf{p}_i = (x_i, y_i, z_i) \in \mathbb{R}^3$, and a table with a list of indices of three vertices for each triangle (topology). Then $I = \bigcup_{i=1}^{m} T_i$ is the point set of all triangles, i.e., our given object. Our goal is to derive an affine map $\tau : \mathbb{R}^3 \to \mathbb{R}^3$ in such way that for an arbitrary concatenation σ of translations, rotations, reflections, and scaling the desired invariance property of τ, namely $\tau(I) = \tau(\sigma(I))$ holds where we have set $\sigma(I) := \{\sigma(\mathbf{v}) | \mathbf{v} \in I\}$ and similarly for τ. Let S_i be the area of triangle T_i, $i = 1, \ldots, m$. For simplicity of notation we may assume that the triangles intersect only on subsets of measure zero so that we may write the overall surface in the model as $S := S_1 + \ldots + S_m = \int_I dv$.

The *translation invariance* is accomplished by translating the center of gravity of a model, $\mathbf{c}$, to the origin, i.e., by forming the point set $I_1 := I - \mathbf{c} = \{\mathbf{u} \mid \mathbf{u} = \mathbf{v} - \mathbf{c}, \ \mathbf{v} \in I\}$. To secure the *rotation invariance* we apply the PCA on the set I_1. First, we calculate the covariance 3×3-matrix $M = \frac{1}{S} \int_{I_1} \mathbf{v} \cdot \mathbf{v}^T dv$. The remaining part of this step follows the standard PCA. Since the matrix M is a symmetric real matrix its eigenvalues are real and the eigenvectors orthogonal. We calculate the eigenvalues of M, sort them in decreasing order, compute the corresponding eigenvectors and scale them to Euclidean unit length. We form the rotation matrix R, which has the scaled eigenvectors as rows. Afterwards, we rotate the set I_1 and obtain a new point set $I_2 = R \cdot I_1 = \{\mathbf{v} \mid \mathbf{v} = R \cdot \mathbf{u}, \ \mathbf{u} \in I_1\}$. To ensure the *reflection invariance* we multiply points in I_2 by a diagonal matrix $F = \mathrm{diag}(\mathrm{sign}(f_x), \mathrm{sign}(f_y), \mathrm{sign}(f_z))$, where $f_x = \frac{1}{S} \int_{I_2} \mathrm{sign}(v_x) v_x^2 dv$, $(f_y, f_z$ similar), and $\mathbf{v} = (v_x, v_y, v_z) \in I_2$. *Scaling invariance* is achieved by scaling the set I_2 by the inverse of $s = [(s_x^2 + s_y^2 + s_z^2)/3]^{1/2}$, where s_x, s_y, and s_z denote the average distances of points $\mathbf{v} \in I_2$ from the yz-, xz-, and xy-coordinate hyperplanes, respectively, i.e., $s_x = \frac{1}{S} \int_{I_2} |v_x| dv$ and likewise for s_y, s_z.

Putting all the above together, the affine map τ, defined by $\tau(\mathbf{v}) = s^{-1} \cdot F \cdot R \cdot (\mathbf{v} - \mathbf{c})$ is applied to all points of the original object I. In practice, it suffices to transform only the set of vertices P.

In contrast to the usual application of the PCA we work with sums of integrals over triangles in place of sums over vertices which makes our approach more complete taking into account all points of the model I with equal weight. The calculation of the integrals is only slightly more expensive. Due to space restrictions we omitt the formulas which can easily be derived.

SPHERICAL HARMONIC REPRESENTATION

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

$$\begin{aligned} r &: & S^2 &\to & \mathbb{R} \\ & & \mathbf{u} &\mapsto & \max\{r \geq 0 \mid r\mathbf{u} \in I \cup \{0\}\} \end{aligned}$$

where 0 is the origin. This function $r(\mathbf{u})$ measures the extent of the object in directions given by $\mathbf{u} \in S^2$, compare [7]. Similarly, one may consider a rendered perspective projection of the object on an enclosing sphere as another example (compare [2]). If we can characterize such maps with a small number of parameters then these can be regarded as good candidates for feature vectors in 3D object retrieval. The Fourier transform on the sphere provides a suitable approach that 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 $\mathbf{u}_{ij} = (\cos\varphi_i \cos 2\varphi_j, \cos\varphi_i \sin 2\varphi_j, \sin\varphi_i)$, where $\varphi_k = (2k + 1 - n)\pi/2n$, $k = 0, \ldots, n-1$ and $i, j = 0, \ldots, n-1$. We cannot give more details here and refer to the survey and software in [1]. A example output of the absolute values of the spherical Fourier coefficients (up to $l = 3$) is given here:

			0.37			
		0.020	0.052	0.020		
	0.068	0.012	0.012	0.012	0.068	
0.0052	0.0025	0.0032	0.0026	0.0032	0.0025	0.0052

Feature vectors can be extracted from the first l rows of coefficients. This implies that such a feature vector contains all feature vectors of the same type of smaller dimension, thereby providing a novel embedded multi-resolution approach for 3D shape feature vectors, see also Figure 1.

RESULTS

As a proof of concept we provide results from just two experiments. As feature vectors we used the absolute values from the first 3, 5, and 9 rows of
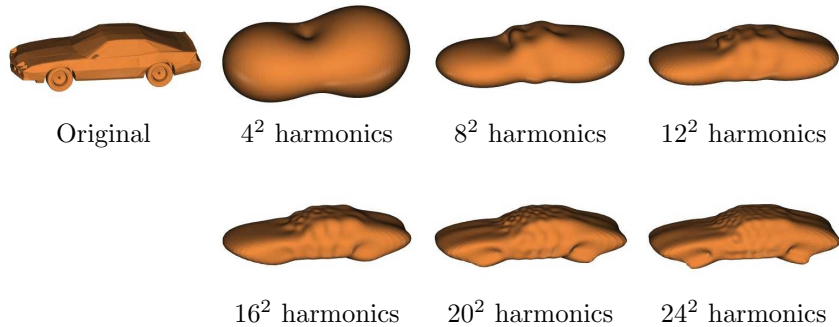
Figure 1: Multi-resolution representation of the function $r(\mathbf{u}) = \max\{r \geq 0 \mid r\mathbf{u} \in I \cup \{0\}\}$ used to derive feature vectors from Fourier coefficients for spherical harmonics.

spherical harmonic coefficients. Since the absolute values in each row are symmetric, we obtained feature vector dimensions 6, 15, and 45 (small, medium and large dimension). The distance between vectors was calculated using the l_2 norm. We also compared the performance using three different PCAs, the simple one based on mesh vertices [3], the one using vertices with weights according to local triangle area [7], and the continuous PCA introduced here.

The 3D model database used for experiments contained around 1900 models. We had manually classified models by shape (e.g., cars, planes, bottles, chairs, etc.) and used this classification in so-called precision/recall tests. Briefly, precision is the proportion of retrieved models that are relevant and recall is the proportion of the relevant models actually retrieved. By examining the precision/recall diagrams for different queries (and classes) we obtained a measure of the retrieval performance. The results in Figure 2 indicate that increasing the dimension of the feature vectors improved the retrieval results. Moreover, on average the KLT in the continuous form was as good or better than the KLT based on weighted vertices, and much better than the plain KLT without weights. For some individual queries we observed that the continuous KLT performed best.

CONCLUSION

In summary we have introduced two tools useful in systems for 3D object retrieval. The continuous Karhunen-Loeve transform ensures that all surface points of the model receive equal weight during pose estimation. The Fast Fourier Transform on the sphere with spherical harmonics provides a natural approach for generating embedded multi-resolution 3D shape feature vectors.
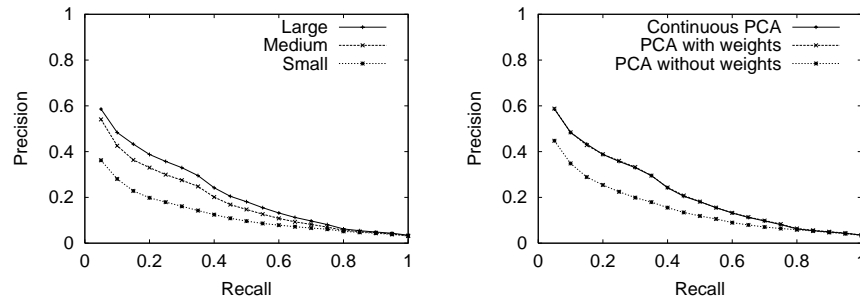
Figure 2: Precision vs. recall of queries averaged over five classes of objects bottles, swords, chairs, cars, and planes). On the left different dimensions are used and on the right different different PCAs were employed for the high dimensional feature vector.

References

[1] 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/g̃eelong/sphere/.

[2] M. Heczko, D. Keim, D. Saupe, and D.V. Vranić, "A method for similarity search of 3D objects" (in German), *Proc. of BTW 2001*, Oldenburg, Germany, pp. 384–401, March 2001.

[3] MPEG Video Group, "MPEG-7 Visual part of eXperi-metation Model (version 8.0)," Doc. ISO/MPEG N3673, La Baule, October, 2000.

[4] M. Novotni and R. Klein, "A geometric approach to 3D object comparison," *Proc. of SMI 2001*, Genova, Italy, 2001, (to appear).

[5] 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.

[6] 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.

[7] D.V. Vranić and D. Saupe, "3D Model Retrieval," *Proc. of Spring Conference on Computer Graphics and its Applications (SCCG2000)*, Budmerice Manor, Slovakia, pp. 89–93, May 2000.