DESCRIPTION OF 3D-SHAPE USING A COMPLEX FUNCTION ON THE SPHERE

Dejan V. Vranić and Dietmar Saupe

University of Konstanz, Department of Computer and Information Science
D-78457 Konstanz, Germany
{vranic, saupe}@fmi.uni-konstanz.de

ABSTRACT

We propose a novel feature vector suitable for searching collections of 3D-objects by shape similarity. In this search a polygonal mesh model serves as a query. For each model feature vectors are automatically extracted and stored. Shape similarity between 3D-objects in the search space is determined by finding and ranking nearest neighbors in the feature vector space. Ranked objects are retrieved for inspection, selection, and processing. The feature vector is obtained by forming a complex function on the sphere. Afterwards, we apply the Fast Fourier Transform (FFT) on the sphere and obtain Fourier coefficients for spherical harmonics. The absolute values of the coefficients form the feature vector.

Retrieval efficiency of the new approach is evaluated by constructing precision/recall diagrams and using two different 3D-model databases. We compared the approach with two methods based on real functions on the sphere. Our empirical comparison showed that the complex feature vector performed best.

We also prepared a Web-based retrieval system for testing methods discussed in this paper.

1. INTRODUCTION AND RELATED WORK

As the amount of new information generated in the world rapidly increases, efficient search in collections of structured data, texts, and multimedia objects (e.g., images, video and audio sequences, and 3D-models) becomes more important. Solutions for efficient retrieval of structured information as well as textual documents exist. However, the search for a specific multimedia object is more intricate. Features of multimedia data can be extracted manually (high-level features), automatically (low-level features), and semi-automatically (controlled mapping of low-level to high-level features). Properties like automatic extraction and flexibility in describing objects attract many researchers to explore low-level features of images, videos, 3D-models, etc. Consequently, several techniques for content-based multimedia retrieval (mostly for images) have been reported.

The topic of our research [2,8-11] is automatic generation of descriptions suitable for searching collections of 3D-mesh models by shape similarity. We consider that the most important criteria for defining 3D-shape descriptors are the following:
1. invariance with respect to translation, rotation, scaling, and reflection of a 3D-object,
2. robustness with respect to level-of-detail,
3. efficient feature extraction and search, and
4. multi-resolution (MR) feature representation.

A broader range of criteria for multimedia descriptors in general is defined by the MPEG-7 standard [4]. The MPEG-7 standard also identifies the following processing chain in multimedia retrieval applications.

Feature Extraction	→	Feature Description	→	Description Consumption
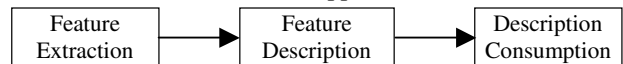
MPEG-7 standardizes only feature description while feature extraction (analysis) and description consumption (application, e.g., search) are left for the competition. Our 3D-model retrieval algorithm complies with the context of MPEG-7 [10].

Usually, before extracting shape features from 3D models a canonical position and orientation is estimated for each object [2,6,8-11]. The pose estimation is used for accomplishing the invariance requirement (1). The most prominent tool for pose estimation is the Principal Component Analysis (PCA, also known as Karhunen-Loeve transform) [7]. Modifications of the discrete PCA used for 3D-model pose estimation are proposed in [2,6,9]. The Continuous PCA (CPCA) suitable for finding the canonical position of a polygonal mesh was introduced in [8]. We stress that all of the (infinitely many) points in the polygons of an object are equally relevant for the CPCA. There are a few exceptions in which pose estimation is not necessary because features are derived from relative properties of a model (e.g., topology [3] and curvature [5]). Typically, extraction of such descriptors is time-consuming.

The requirements (2)-(4) should be fulfilled by descriptor definitions. Descriptors are mostly represented as real-valued vectors [2,5,6,8-11] thereby the l_1 (or l_2) norm can be engaged for nearest neighbor search. Other representations (e.g., graphs [3] or octrees [2]) can be used, as well as new similarity measures can be defined [3,5].

In this paper we present a new 3D-shape descriptor derived from a complex function on the sphere. Using Fourier expansion of this function provides embedded MR feature vectors. We describe our experiments that we designed to evaluate the proposed method. Our Web-based retrieval system serves as a proof-of-concept.

2. COMPLEX FUNCTIONS ON THE SPHERE FOR 3D-SHAPE DESCRIPTORS

In this section we explain how complex functions suitable for describing 3D-shape can be generated. Triangle meshes are commonly used to represent 3D-models. Let $\{T_1,...,T_m\}$ ($T_i \subset \mathbb{R}^3$) be the set of triangles of a mesh, given by vertices (geometry) $\{\mathbf{p}_1,...,\mathbf{p}_n\}$ ($\mathbf{p}_i = (x_i, y_i, z_i) \in \mathbb{R}^3$). An index table with three vertices per triangle (topology) is associated to the mesh. Then the object is $I = \bigcup_{i=1,...,m} T_i$, the point set of all triangles. As described in section 1, we start with pose estimation. In order to determine the canonical coordinate frame, we apply the CPCA [11] to the set I. Our study confirmed that the CPCA performs better than the discrete PCA (including modifications). If a 3D-object is closed and orientable polygonal mesh, then the CPCA can be applied to the model volume. However, only a small fraction of available 3D-models represents solid objects, therefore, we apply the CPCA to the set I.

In the canonical coordinate frame, we define a unit sphere S^2 with the center in the origin (i.e., center of the sphere coincides with the center of mass of the model). Further, we define the function $r(\mathbf{u})$ ($\mathbf{u} \in S^2$)

$$r \quad : \quad S^2 \to \mathbb{C}$$
$$r(\mathbf{u}) = x(\mathbf{u}) + i\, y(\mathbf{u})$$
$$x : S^2 \to [0,+\infty) \in \mathbb{R}, \; y : S^2 \to [0,1] \in \mathbb{R},$$

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

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

The imaginary part of $r(\mathbf{u})$ is defined as follows

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

where $\mathbf{n}(\mathbf{u})$ is the normal vector of the mesh at the point $\mathbf{u}\,x(\mathbf{u})$ ($x(\mathbf{u}) \neq 0$). The function $y(\mathbf{u})$ can also be described as a rendered perspective projection of the model on an enclosing sphere.

We recall that in [9] we took a number of samples $x(\mathbf{u})$ as a feature vector in the spatial domain. This feature vector is sensitive to small perturbations of the object. In [11] we improved the robustness of the feature vector by taking samples of the spherical function $x(\mathbf{u})$ at many points, but characterizing the map by just a few coefficients in the spectral domain. In this paper we engage spherical harmonics to merge two features represented by real functions ($x(\mathbf{u})$ and $y(\mathbf{u})$) by embracing them into a single complex function.

3. "COMPLEX FEATURE VECTOR" WITH SPHERICAL HARMONIC REPRESENTATION

As mentioned in section 2, we apply the spherical Fourier transform to the function $r(\mathbf{u})$. The Fourier transform on the sphere uses the spherical harmonics functions $Y_{l,m}$ to represent any function $r \in L^2(S^2)$ as

$$r = \sum_{l \geq 0} \sum_{|m| \leq l} \hat{r}(l,m) Y_{l,m}$$

where $\hat{r}(l,m)$ denotes a Fourier coefficient. Spherical basis functions are certain products of Legendre functions and complex exponentials. The spherical harmonics are orthogonal with respect to integration over the surface of the unit sphere. The complex Fourier coefficients can be efficiently calculated by a spherical FFT algorithm applied to samples taken at points $\mathbf{u}_{ij}$,

$$\mathbf{u}_{ij} = (\cos\varphi_i \sin\theta_j, \sin\varphi_i \sin\theta_j, \cos\theta_j),$$
$$\varphi_i = 2i\pi / n, \quad \theta_i = (2i+1)/2n, \quad i = 0,...,n-1.$$

The parameter $n \in \mathbb{N}$ is selected sufficiently large. More details about spherical harmonics as well as the corresponding software can be found in [1].

We apply the FFT to the functions $r(\mathbf{u})$, $x(\mathbf{u})$, and $y(\mathbf{u})$, defined in section 2. In the case of real functions, the spherical Fourier (complex) coefficients are pairwise conjugate for $m \neq 0$, i.e.,

$$\hat{r}(l,m) = \overline{\hat{r}(l,-m)} \Rightarrow |\hat{r}(l,m)| = |\overline{\hat{r}(l,-m)}|.$$

An example output of the absolute values of the spherical Fourier coefficients of the complex function $r(\mathbf{u})$, when the underlying 3d-model is non symmetrical, is given here:

	$m=-2$	$m=-1$	$m=0$	$m=1$	$m=2$
$l=0$			0.898		
$l=1$		0.082	0.227	0.036	
$l=2$	0.328	0.023	0.289	0.022	0.366

If the function $x(\mathbf{u})$ is engaged, the spherical harmonic coefficients can be used to reconstruct an approximation of the underlying object at different levels. In Figure 1, the original model is a cube. After sampling at 16384 points $\mathbf{u}_{ij}$ ($n = 128$) the FFT was applied. The reconstructed objects are obtained by taking first k rows of coefficients.

We have chosen to use only absolute values of coefficients in the first k rows as components of our feature vectors. Because of the symmetry we take only coefficients with $m \geq 0$ if the underlying function is real. Hence, the dimensions of the feature vectors are given by

$$\dim = \begin{cases} \sum_1^k l = k(k+1)/2 & \text{real function } (x(\mathbf{u}) \text{ or } y(\mathbf{u})) \\ \sum_1^k (2l+1) = k^2 & \text{complex function } (r(\mathbf{u})) \end{cases}.$$

Note that feature vectors derived in this way contain all feature vectors of the same type of smaller dimensions. Therefore, an embedded MR feature representation is provided. In what follows, we refer to features derived from the functions $x(\mathbf{u})$, $y(\mathbf{u})$ and $r(\mathbf{u})$ as ray-based [8,11], shading-based, and complex feature vector, respectively.

4. RESULTS

We use two different collections of 3D-models for testing, our own collection of 1839 objects and the official MPEG-7 test set of 227 meshes. Our 3D-model database is mostly collected from the Internet (e.g., *www.3dcafe.com* and *www.viewpoint.com*) and on average a model contains 5659 vertices and 10314 triangles. We manually classified models by shape. For example, we have 33 models of cars, 63 airplanes, etc. On average a model from the MPEG-7 3D-database contains 6569 vertices and 8977 triangles. The MPEG-7 collection is classified into 15 categories mostly by semantics (not by shape). The categories are the following: aerodynamic (35 models), balloon (7), building (10), car (17), e1m (9), finger (30), fourlimb (31), letters a-e (10 models each), missile (10), soma (7), and tree (21).

In our empirical study we compared retrieval performance of ray-based, shading-based, and the complex feature vectors. A Web-based retrieval system for demonstrating the three methods above is located at: http://www.informatik.uni-leipzig.de/~vranic/ICME2002/. A retrieval example generated by the on-line system is shown in figure 2. A model of an airplane from the MPEG-7 test set was used as the query, while the l_2 norm was applied to the proposed feature vector of dimension 169. The models are visualized from the same direction in the original coordinate frame (before pose normalization).

Feature extraction of the complex vectors is efficient. For example, on a PC with an 866 MHz Pentium III processor running Windows 2000 the average time for extracting complex feature vectors with n=128 (section 3) was 0.415 seconds for our database and 0.355 seconds for the MPEG-7 models.

In order to compare the effectiveness of feature vectors, we used the classification of databases as ground truth to compute precision/recall diagrams [8,11]. Briefly, *recall* is the proportion of the relevant models actually retrieved and *precision* is proportion of retrieved models that is relevant. All models in the same category are relevant to each other. We also examined how the choice of distance metric (l_1 or l_2) affects the retrieval. A part of our results is depicted in figure 3. For each descriptor we calculated average precision/recall diagrams for all models in classes "car" (in both databases), "airplane" (our database), and "aerodynamic" (MPEG-7). In separate tests, we concluded that the best choice of dimensions is to select the first k=13 rows of coefficients. As expected, the overall retrieval performance was better when we searched in the much smaller collection. The results from figure 3 help us rank the three descriptors by performance, while it is difficult to infer which metric is better. The shading-based feature possessed the lowest retrieval rate, because robustness with respect to level-of-detail was weaker compared to the other two features. The ray-based descriptor performed best when searching the MPEG-7 3D-objects, while the complex descriptor was superior when retrieving models from our database. Note that at small recall values the new descriptor performed best in all cases. Since our database is classified by shape and MPEG-7 models are sorted by semantics (e.g., there are limousines, formulas, and seats in the class "car") we conclude that the new (complex) feature vector shows the best performance for shape representation in the context of 3D-model retrieval.

In the future, we will consider some other definitions of the function $y(\mathbf{u})$ in order to improve the robustness with respect to level-of-details. A weighted complex function of the type $r(\mathbf{u}) = \alpha x(\mathbf{u}) + i\,\beta y(\mathbf{u})$ ($\alpha, \beta \in \mathbb{R}$) will be explored.

5. CONCLUSION

In this paper we have introduced a new 3D-shape descriptor derived from a complex function on the sphere. In order to represent features that depend on angular coordinates, we engage spherical harmonics obtaining embedded multi-resolution 3D-shape feature vectors. In our tests the complex feature vector performed better than the ray-based and shading-based feature vectors.

6. REFERENCES

[1] D.M. Healy, D. Rockmore, P. Kostelec, and S. Moore, "FFTs for the 2-sphere — Improvements and variations," TR2002-419, Department of C. S., Dartmouth College, 2002. Paper and corresponding software, *SpharmonicKit*, are available at: http://www.cs.dartmouth.edu/~geelong/sphere/.

[2] M. Heczko, D. Keim, D. Saupe, and D.V. Vranić, "Verfahren zur Ähnlichkeitssuche auf 3D-Objekten," *Datenbank Spektrum*, vol. 2(2), pp. 54-63, 2002.

[3] M. Hilaga, Y. Shinagawa, T. Komura, and T.L. Kunii, "Topology Matching for Fully Automatic Similarity Estimation of 3D Shapes," *Proc. SIGGRAPH 2001*, pp. 203-212, 2001.

[4] MPEG Home Page: http://mpeg.telecomitalialab.com/

[5] MPEG Video Group, "MPEG-7 Visual part of eXperimetation Model (version 9.0)," Doc. ISO/MPEG N3914, MPEG Pisa Meeting, 2001.

[6] E. Paquet, A. Murching, T. Naveen, A. Tabatabai, and M. Rioux, "Description of Shape Information for 2-D and 3-D Objects," *Sig. Proc.: Image Comm.*, vol. 16, pp. 103-122, 2000.

[7] M. Petrou and P. Bosdogianni, *Image Processing: The Fundamentals,* John Wiley, 1999.

[8] D. Saupe and D.V. Vranić, "3D model retrieval with spherical harmonics and moments," *Proc. DAGM 2001*, Munich, Germany, pp. 392-397, September 2001.

[9] D.V. Vranić and D. Saupe, "3D Model Retrieval," *Proc. SCCG 2000*, Budmerice, Slovakia, pp. 89-93, 2000.

[10] D.V. Vranić and D. Saupe, "A Feature Vector Approach for Retrieval of 3D Objects in the Context of MPEG-7," *Proc. ICAV3D 2001*, Mykonos, Greece, pp. 37-40, 2001.

[11] D.V. Vranić, D. Saupe, and J. Richter, "Tools for 3D-object retrieval: Karhunen-Loeve Transform and spherical harmonics," *IEEE MMSP 2001*, Cannes, France, pp. 293-298, 2001.
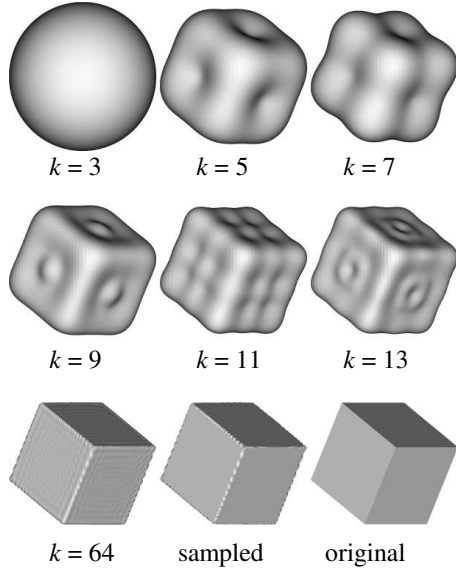
Figure 1. Multi-resolution representation of the function $x(\mathbf{u})$ applied for a cube using k^2 spherical harmonic coefficients.
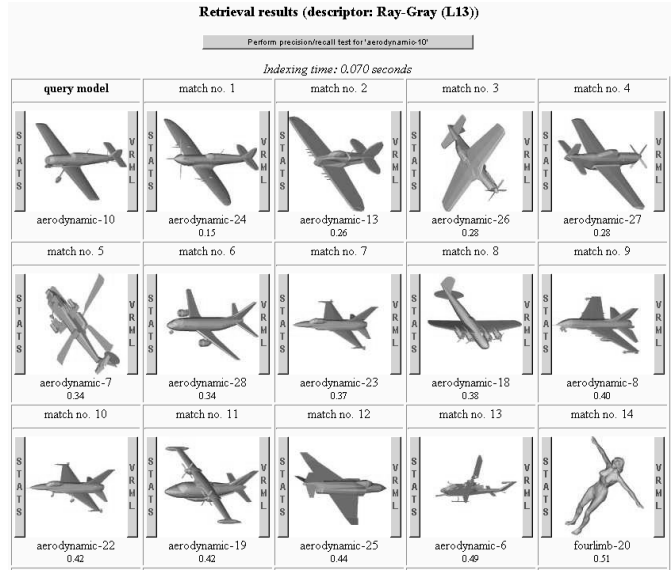
Figure 2. Query for an airplane from the MPEG-7 database using the complex feature vector of dimension 169 and the l_2 distance.
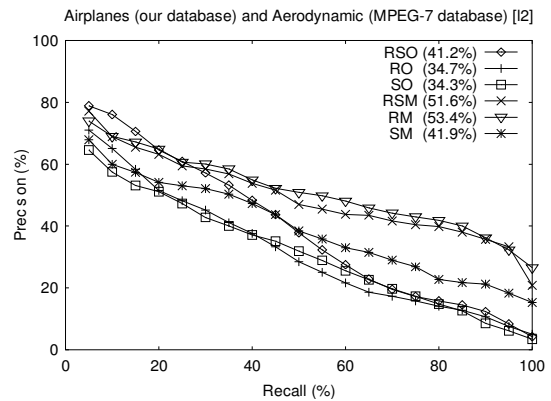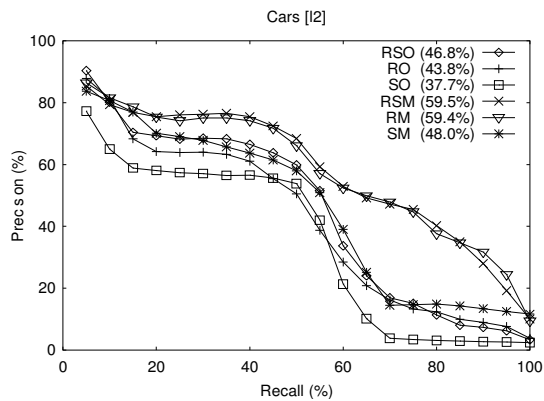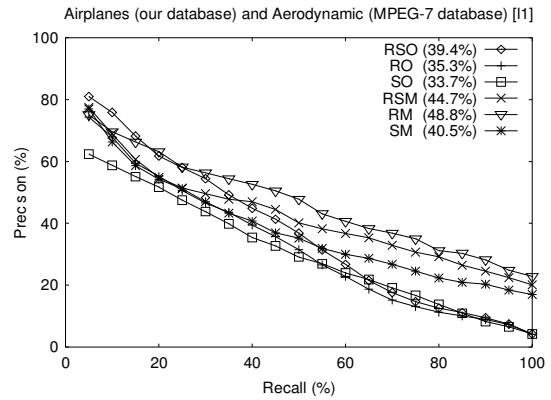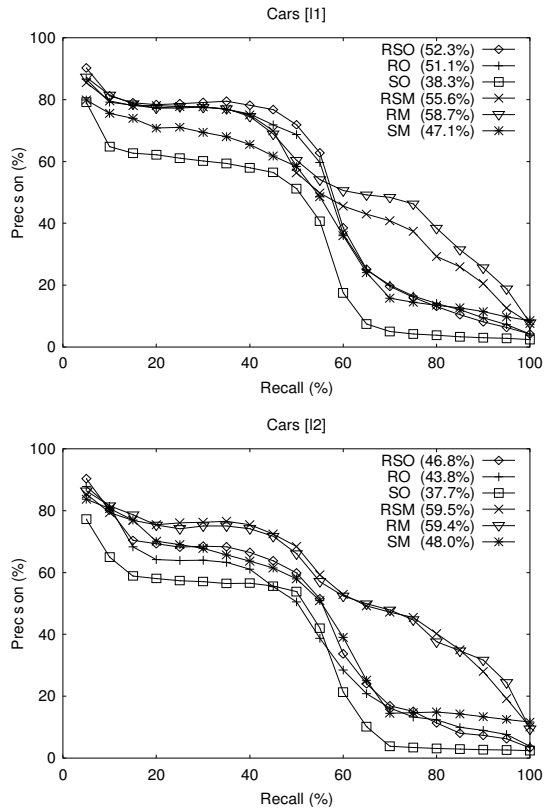
Legend: [l1] denotes the l_1 norm; [l2] denotes the l_2 norm; RSO (resp. RSM) denotes the complex feature vector (FV) applied at our (resp. MPEG-7) database; RO (resp. RM) denotes the ray-based FV applied at our (resp. MPEG-7) database; SO (resp. SM) denotes the shading-based FV applied at our (resp. MPEG-7) database; the mean values of the curves are given in the brackets.

Figure 3. Average precision vs. recall of queries for two categories from both our and MPEG-7 3D-model database using three FVs with spherical harmonic representation, the complex FV of dimension 169 and ray-based and shading-based FV of dimension 91 (always k=13). Both databases contain the category 'cars', the category 'airplanes' exists only in our database, and the category 'aerodynamic' is present only in the MPEG-7 database.