

AN IMPROVEMENT OF ROTATION INVARIANT 3D-SHAPE DESCRIPTOR BASED ON FUNCTIONS ON CONCENTRIC SPHERES

Dejan V. Vranić*

University of Konstanz, Department of Computer and Information Science
D-78457 Konstanz, Germany

ABSTRACT

In this paper, we consider 3D-shape descriptors generated by using functions on a sphere. The descriptors are engaged for retrieving polygonal mesh models. Invariance of descriptors with respect to rotation of a model can be achieved either by using the Principle Component Analysis (PCA) or defining features in which the invariance exists. The contribution of the paper is twofold: firstly, we define a new rotation invariant feature vector based on functions on concentric spheres, that outperforms a recently proposed descriptor; secondly, we compare the two approaches for achieving rotation invariance as well as options to use a single function or several functions on concentric spheres to generate feature vectors.

We conclude that descriptors, which use the PCA, outperform others, while capturing the internal structure of a 3D-model with functions on concentric spheres can improve retrieval effectiveness.

1. INTRODUCTION

Similarity-based retrieval of 3D-mesh models is becoming a topic that attracts more and more researchers. Only a few years ago solutions for 3D-object search were related to a classical computer vision problem, i.e., all information about shape was obtained from 2D-images of the object. The increasing number of polygonal mesh models available on the Internet changes the situation. Consequently, various techniques for describing 3D-shape of polygonal meshes have been reported.

Usually, 3D-model retrieval algorithms consist of three steps [7]: pose estimation (normalization), feature extraction, and similarity search. During the normalization, a 3D-object is translated, rotated, scaled, and flipped (reflected) into a canonical position and orientation. Because of feature definitions, some retrieval techniques [3,6] do not require normalization, while other methods [1,5] require only certain stages of the pose estimation step (e.g., translation and scaling). In this paper, we compare the ray-based spherical harmonic (RH) feature

vector [7], which engages the Continuous PCA (CPCA) [7] in the normalization step, to the voxel-based spherical harmonic (VH) descriptor, obtained by summing up squares of magnitudes of certain spherical harmonic coefficients of functions on concentric spheres [1,5]. The summing up was performed in order to obtain a rotation invariant descriptor. Besides, we define two variants of a new feature vector based on the ray-based feature, and compare retrieval performance of all four descriptors.

After our evaluation, we concluded that the feature vector proposed in [7], as well as both feature vectors proposed in this paper, outperformed the descriptor proposed in [1,5].

2. RELATED WORK

In this section, we briefly present the methods proposed in [1,5,7]. The basic idea in forming any of the considered descriptors is to apply the Spherical Fourier Transform (SFT) to a function on a sphere, i.e., to approximate the function using spherical harmonics [2].

Let $r = r(\theta, \varphi)$ be a function on a sphere, $r \in L^2(S^2)$. Function r can be approximated using spherical harmonic basis functions $Y_{l,m}$ as

$$r \cong \sum_{l \geq 0} \sum_{|m| \leq l} \hat{r}(l,m) Y_{l,m}, \quad (1)$$

where $\hat{r}(l,m)$ denotes a complex Fourier coefficient (for more details see [2]). As far as we know, the idea to apply the SFT to the 3D-model retrieval process was introduced in [7].

The RH descriptor is extracted in the canonical coordinate frame, i.e., each 3D-model is normalized first. The function $r = r(\mathbf{u})$ measures the extent of the object in directions given by \mathbf{u} ,

$$\mathbf{u}(\theta, \varphi) = (\cos \varphi \sin \theta, \sin \varphi \sin \theta, \cos \theta) \in S^2. \quad (2)$$

After applying the SFT, we calculate the absolute values $|\hat{r}(l,m)|$ ($0 \leq l < l_{\max}$, $0 \leq m \leq l$) to form the corresponding feature vector of dimension $\dim = (l_{\max} + 1) l_{\max} / 2$. An embedded multi-resolution representation (MRR) is a very desirable characteristic of this descriptor. For instance, if we choose $l_{\max} = 21$, then the dimension of the vector is 231 and all lower-dimensional vectors are contained in it.

The normalization step of the approach presented in [1,5] consists of translation and scale, while the definition of descriptor secures rotation and reflection invariance.

* This work has been supported by the Deutsche Forschungsgemeinschaft (www.dfg.de), Project SA 449/10-1, within the strategic research initiative "Distributed Processing and Delivery of Digital Documents" (V3D2), SPP 1041.

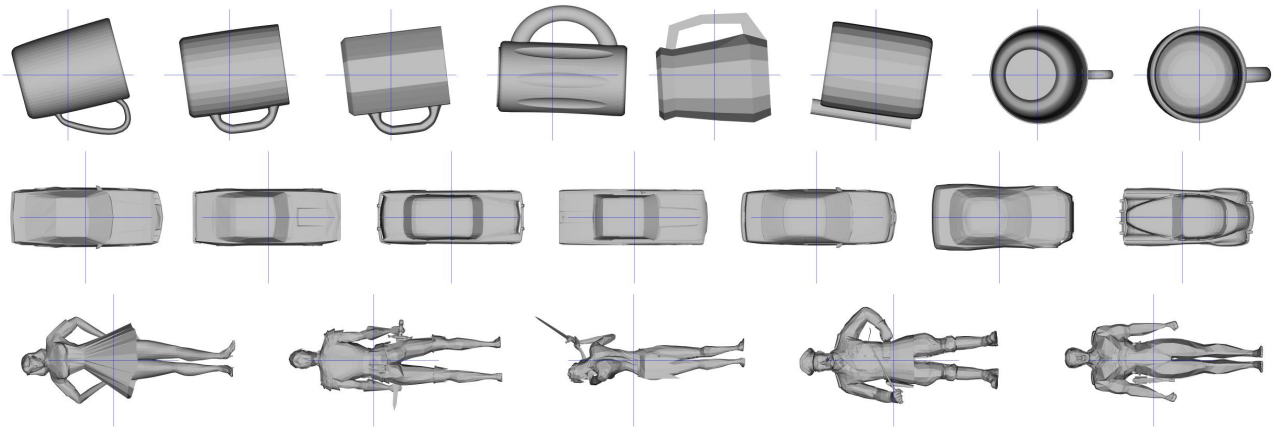


Figure 1. Examples of pose estimation using the Continuous Principle Component Analysis [7]. The models are visualized from the positive side of the z -axis, while the x -axis travels to the right side.

The SFT is also used, but now for R real functions on concentric spheres. At first, a model is translated so that its center of mass coincides with the coordinate origin. Then, the model is voxelized into a $2R \times 2R \times 2R$ grid so that the voxel values are 0 or 1. The model is scaled so that the average distance from non-zero voxels to the origin is approximately $R/2$. Further, the voxel grid is decomposed into R functions on spheres with radii 1 through R . The value of the function on the sphere with radius r_i at the point (θ, φ) is equal to the value of voxel containing the point $r_i \mathbf{u}(\theta, \varphi)$ (2). After sampling the functions, the SFT is applied R times. Finally, in order to have the rotation invariance avoiding the PCA [4] properties of spherical functions $Y_{l,m}$ are used. Namely, the subspace X_m spanned by $\{ Y_{l,m} \mid -l \leq m \leq l \}$ is invariant with respect to rotations of the sphere.

Therefore, the norms

$$\sqrt{\sum_{m \leq l} |\hat{r}_i(l, m)|^2}, \quad 0 \leq l < l_{\max}, \quad 1 \leq i \leq R, \quad (3)$$

are taken to be vector components providing the rotation invariance, but sacrificing the information contained in each single coefficient. The dimension of the vector is $\dim = l_{\max} R$. In [1,5] it is recommended to take $l_{\max} = 16$ and $R = 32$, i.e., $\dim = 512$.

In sections 4 and 5, we define a new descriptor that can be regarded as an improvement of the VH descriptor, and compare methods for obtaining rotation invariance.

3. SHOULD THE PCA BE AVOIDED?

Several approaches for 3D-model retrieval in which the pose estimation was not necessary have been reported [3,6]. Topology matching [3] is an interesting and intricate technique, which matches graph representations of 3D-objects. However, the method is suitable only for certain types of models. Time consuming extraction and matching are also drawbacks. The MPEG-7 shape

spectrum descriptor [6] works well in the case of articulated modifications of models, while its major drawback is a high sensitivity with respect to the levels of detail or different tessellations of an object

In our work, we defined several descriptors and all of them require the normalization step [7]. Translation invariance is fixed by moving the center of gravity of the mesh model to the origin. Rotation invariance is obtained using the PCA [4]. However, we do not apply the PCA to a discrete set of points. Instead, we consider the union of all polygons of the mesh with infinitely many points. The *Continuous* PCA (CPCA) was proposed in [7]. Scaling and reflection invariance are secured by calculating appropriate parameters [7]. We decided to use the CPCA, because it is very efficient in many cases (e.g., Figure 1, the last two rows), it is not time-consuming, and it can be applied even if the mesh model is not orientable or a closed polygonal surface. The CPCA shows certain weaknesses. For instance, cups in Figure 1 are not ideally aligned. The reason for the deviations is not the non-robustness of the CPCA to outliers [1,5], but rather the obvious differences between some parts of the objects. Indeed, robustness w.r.t. outliers is demonstrated in the case of the models of humans (Figure 1, last row). We notice that in some cases scaling and reflection of the models are non-uniform. The choice to take the center of mass as the origin [1,5,7] is questionable in particular when we define functions on a sphere with the center in the origin. For example, the third model in the last row in Figure 1 is well rotated, but the coordinate origin deviates from the other models as well as the scale. Future work will address these problems.

We compared the retrieval performance of the feature vectors that avoid the PCA [1,5,6] to our method [7], and concluded that, in spite of the weaknesses of the PCA, our method performed the best. Only a fraction of our experiments is presented in section 5.

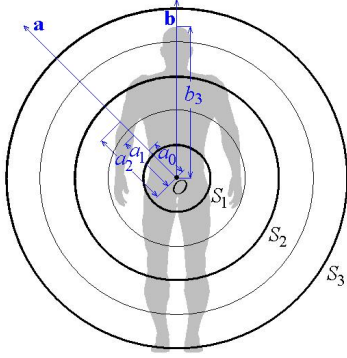


Figure 2. Determination of function values.

4. NEW DESCRIPTORS

The feature vector presented in [1,5] was derived from the voxelized model. The motive for the voxelization was to achieve a better robustness w.r.t. variances of the polygonal surface. However, we feel that many fine details are lost in the voxel grid. Therefore, we propose to translate the center of mass of a model to the origin, scale the model using the scaling factor proposed in [7], cast rays from the origin in many directions $\mathbf{u}(\theta, \varphi)$ [7], find all points of intersection with the polygonal mesh, and define several functions on the sphere using the intersection points. Definition of function values is depicted in Figure 2. Let \mathbf{a} and \mathbf{b} be rays (cast from the origin O) intersecting the mesh at three and one points, respectively. The distances from the intersection points to the origin are a_0, a_1, a_2 , and b_3 . Let f_1, f_2 , and f_3 be the functions on the spheres S_1, S_2 , and S_3 , respectively. For each intersection point we determine the closest sphere and set the corresponding value of the function on that sphere. In the given example, we set $f_1(\mathbf{a}) = a_1, f_2(\mathbf{a}) = a_2, f_3(\mathbf{a}) = 0, f_1(\mathbf{b}) = 0, f_2(\mathbf{b}) = 0$, and $f_3(\mathbf{b}) = b_3$. If two intersection points lying on the same directional vector are closest to the same sphere, then the longer distance determines the function value ($a_1 > a_0$). In practice, we take R concentric spheres to define the functions and 16384 directional vectors \mathbf{u} (2) as in [7]. Centers of all spheres lie at the origin. Radii of the spheres take values $t/R, 2t/R, \dots, t$, where t is an empirically determined constant (we usually set $t = 8$). Using the constant value of parameter t rather than the radius of bounding sphere increases robustness w.r.t. outliers. After the sampling procedure the SFT is applied to each function producing R arrays of complex coefficients.

A new descriptor (RH1) can be defined if we apply (3) to the obtained coefficients. The dimension of the feature vector is $dim = l_{max}R$. We suggest to take $l_{max} = 16$ and $R = 8$. In order to check if we can improve the RH descriptor by using several functions on concentric spheres instead of a single one, we created a similar feature vector (RH2). The RH2 is extracted using the complete normalization step. For each of the R arrays of SFT coefficients, the absolute values of the coefficients are used to form the vector (as in [7]), providing the

embedded MRR for a fixed R . The dimension of the vector is $dim = R(l_{max} + 1)l_{max}/2$ and all lower dimensional vectors (with the same R) are embedded. We suggest to set $R = 8$. Concentric spheres capture information about the internal structure of a model, while only the extent was measured in [7].

5. EVALUATION OF THE DESCRIPTORS

We use two different collections of 3D-models for testing: our own collection of 1841 objects and the MPEG-7 test set of 227 meshes. On average a model contains 5761 vertices and 10166 triangles. In order to reduce subjectivity in creating ground truth, both collections have two different categorizations. Firstly, we roughly classified our set of 3D-models obtaining only a few well-defined categories, e.g., airplanes, bottles, cars, missiles, swords, etc. The classification was mostly semantic rather than shape based. The second categorization was done by our colleagues, without our influence, and it is strictly shape based, e.g., limousines and convertible cars are not in the same class. In total, 416 models are classified into 54 classes. The original classification of the MPEG-7 set is mostly semantic, but non-consistent as well. Therefore, we reclassified the MPEG-7 collection by shape-similarity.

In our empirical study we used precision-recall diagrams to compare retrieval performance of the descriptors considered in sections 2 and 4. In Figure 3, the average precision-recall curves for all classifications of models are shown. The l_1 norm was used as the distance metric. As expected, the precision is higher when we search in the MPEG-7 collection, which is much smaller. The RH2 and RH descriptors showed the best performance, which implies that avoiding the PCA by using (3) is a poor trade-off. The RH1 feature clearly outperformed the VH descriptor. This is a confirmation that our approach to have many samples of a model, and let the SFT reduce the variance (high-frequency noise), is better than filtering the noise by voxelizing the model. This result was expected, because of the relation between the resolution of voxelization R and retrieval efficiency of the VH descriptor. Namely, if R is low, then the voxel grid is a rough representation of the model. If we increase R , then the vector dimension increases reducing the discriminant power of the descriptor. We also observe that the RH2 slightly improved the RH descriptor. Hence, capturing information about the internal structure of a model can increase retrieval effectiveness.

We also tested the l_2 norm as the distance metric. The RH, RH1, and RH2 performed similar as in Figure 3, while the precision of the VH descriptor dropped approximately between 6-8%. Having in mind the definitions of descriptors, there is no meaning in testing the l_∞ metric, which is simply not suitable in this case. The Bull-Eye percentage score, adopted by the MPEG-7, as well as the R-Precision score gave a similar ordering as the precision-recall diagrams.

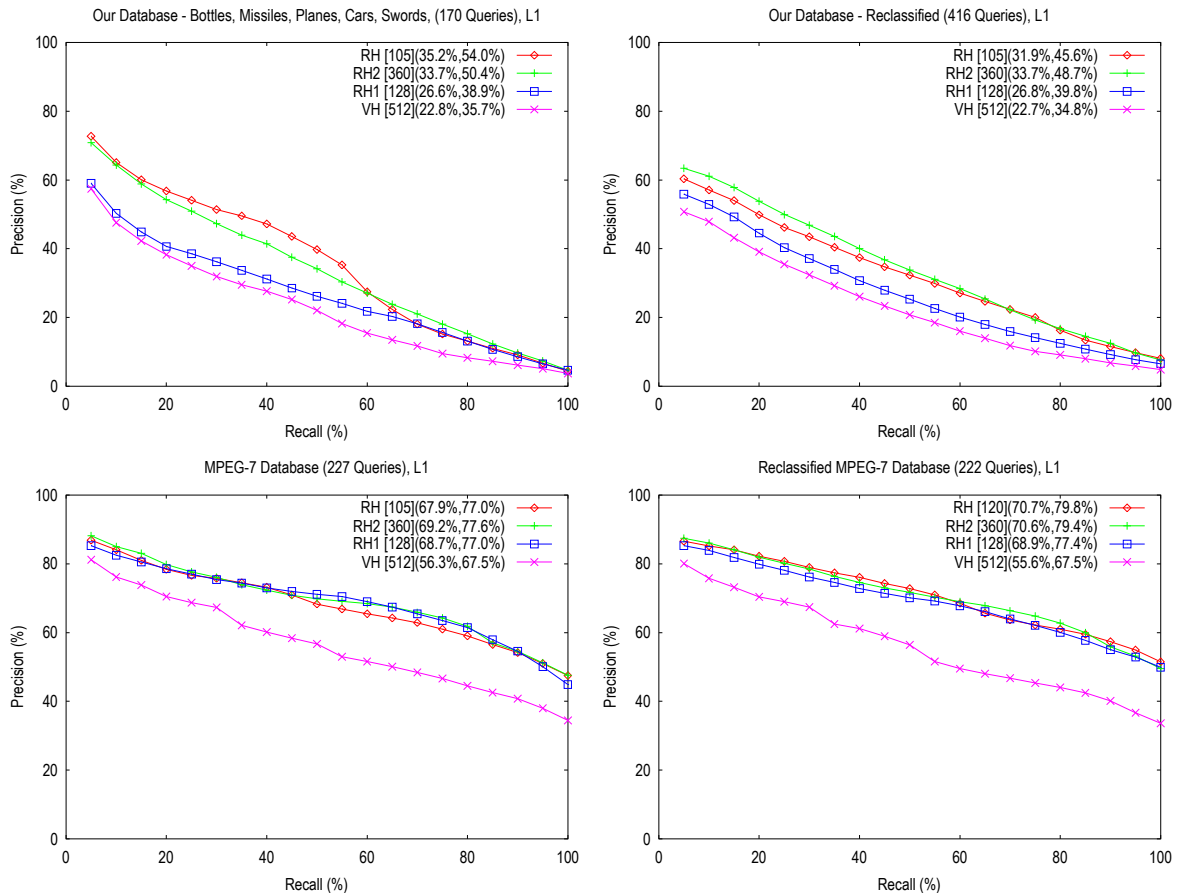


Figure 3. Average precision vs. recall of two databases with two different classifications. For each descriptor, the vector dimension, the average precision for $5\% \leq \text{Recall} \leq 100\%$, and the average precision for $5\% \leq \text{Recall} \leq 50\%$, are given in the brackets.

Our ray-casting algorithm for intersecting a 3D-mesh model is efficient. On a PC with an 1.4 GHz AMD processor running Windows 2000 the average times for extracting the tested RH, RH1, and RH2 feature vectors are 78ms, 268ms, and 255ms, respectively. The extraction time of the VH descriptor is longer because of the voxelization (743ms). The complete normalization needs 36ms, on average.

We also prepared a Web-based retrieval system for demonstrating the four methods, which is located at:

<http://www.informatik.uni-leipzig.de/~vranic/ICIP2003/>.

6. CONCLUSION

We explored the idea to use properties of spherical harmonics for defining rotation invariant 3D-shape feature vectors avoiding the PCA. Besides, since some features of 3D-objects can be seen as functions on spheres, we defined a descriptor capturing the internal structure of a model.

Results show that the descriptors, which use the PCA, outperform the others, while considering the internal structure of a model by using functions on concentric spheres increases retrieval effectiveness.

7. REFERENCES

- [1] T. Fankhouser, et al., "A Search Engine for 3D Models," *ACM Transactions on Graphics*, **22**(1), ACM Press, New York, pp. 83-105, January 2003.
- [2] 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/>.
- [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] Jolliffe, I.T., *Principal Component Analysis*, Springer, 1986.
- [5] M. Kazhdan and T. Funkhouser, "Harmonic 3D Shape Matching," *SIGGRAPH 2002*, Technical Sketch, 2002.
- [6] MPEG Video Group, "MPEG-7 Visual part of eXperimentation Model (version 9.0)," Doc. ISO/MPEG N3914, MPEG Pisa Meeting, 2001.
- [7] 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.