3D Shape Descriptors Based on Fourier Transforms

D. Zarpalas [1], P. Daras [2], D. Tzovaras [1] and M. G. Strintzis [1,2] *Fellow, IEEE*

[1]Informatics and Telematics Institute
1st km Thermi-Panorama Road
570 01 Thessaloniki, Greece

[2]Information Processing Laboratory
Electrical and Computer Engineering Department
Aristotle University of Thessaloniki
540 06 Thessaloniki, Greece

Abstract

The present paper proposes a novel combination of Fourier Transforms for the creation of powerful descriptors, which can be used for efficient content-based search and retrieval of 3D models. The proposed method is applied to the volume of each model, which is decomposed into rays and planes on which the 1D and the 2D Fourier Transform are applied respectively. The coefficients produced from these transformations are the input to the Spherical Fourier Transform, resulting to rotation invariant descriptors, which contain significant shape characteristics. Experiments, which were performed using a database of 544 3D objects, proved that the proposed method is superior in terms of precision-recall than most of the methods reported so far.

1 Introduction

The rapid development of modelling, scanning, digitizing and visualizing techniques for 3D objects, has increase the number of the 3D models available on the Internet. This availability has lead the research community in the development of efficient search and retrieval algorithms capable of determining accurately the similarity between 3D objects. Based on that premise many groups worldwide are developing 3D search systems that index a large repository of computer graphics models, supporting 2D and 3D sketching interfaces for shape-based queries [1, 2, 3].

Most of the algorithms proposed so far are based on the query-by-content approach, which has, up to now, been almost universally adopted in the literature. The MPEG-7 shape spectrum descriptor [4] is defined as the histogram of the shape index, calculated over the entire surface of a 3D object. The shape index gives the angular coordinate of a polar representation of the principal curvature vector, and it is implicitly invariant with respect to rotation, translation and scaling. Zhang and Chen [5] use features such as volume-surface ratio, moment invariants and Fourier transform coefficients. They improve the retrieval performance by an active learning phase in which a human annotator assigns attributes such as "airplane", "car", and so on to a number of sample models. Osada *et al.* [6] introduce and compare shape distributions, which measure properties based on distance, angle, area and volume measurements between random surface points. They evaluate the similarity between the objects using a metric that measures distances between distributions. In [7] the descriptors that are extracted from the input model are the geometrical characteristics of the 3D objects such as the angles and edges that describe the outline of the model.

In [8] a 3D content-based retrieval method relying on 3D Zernike moments is presented. These moments are computed as a projection of the function defining the object onto a set of orthonormal functions within the unit sphere; the 3D Zernike polynomials have been introduced in [9]. The descriptors produced are rotation invariant.

In [10], the "Light Field Descriptor Method" (LFD) is proposed. They extract features from the light fields rendered from cameras on a sphere. Then, the similarity between two 3D models is measured by summing up the similarity from all corresponding images of a light field.

In [11], the method "Radialized Spherical Extent Function" (REXT) is based on the Spherical Fast Fourier Transform (SFFT), which is applied to sampled points on the surface of concentric spheres. The sampled points are produced taking distances on radii beginning from the 3D model's center of mass. The method requires a priori registration with principal axes.

In [12], the "Spherical Harmonic Descriptor" (SHD) method also uses the Spherical Fast Fourier Transform. The surface of the 3D models is first voxelized and then sampled points are selected on concentric spheres. The absolute values of the SFFT coefficients calculated on the spheres, form the descriptor vector.

The approach followed in the present paper, involves the decomposition of the digitized model's volume into ray segments and splines (planes). The 1D and 2D Fourier Transform are applied to the rays and splines respectively, creat-

ing new functions, which are used as input to the Spherical Fourier Transform. Then, the magnitude of the produced coefficients is taken into account so as to create descriptor vectors, which are completely invariant to rotation. Further, a method of assigning weights to each descriptor vector is introduced. The weights are significantly improve the effectiveness of the overall method.

The rest of the paper is organized as follows: In Section 2 the proposed approach is described. Experimental results evaluating the efficiency of the proposed method are given in Section 3. Finally, conclusions are drawn in Section 4.

2 Proposed Method

Firstly, the model's volume is digitized by defining voxels and the volumetric binary function $f_b(\boldsymbol{x})$ is created, where $\boldsymbol{x} = (x, y, z)$ and $x, y, z \in [0, 2N]$, which is defined as:

$$f_b(\boldsymbol{x}) = \begin{cases} 1, & \text{when } \boldsymbol{x} \text{ lies within the 3D model's volume,} \\ 0, & \text{otherwise.} \end{cases}$$

Using the center of mass and the distance of the most distant voxel from the center of mass, translation and scaling invariance are easily achieved.

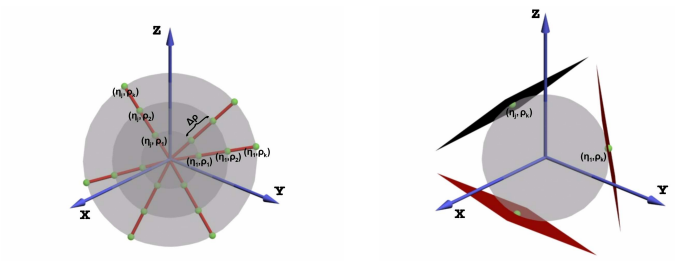
Afterwards, the integer volumetric function of the model $f(\boldsymbol{x})$ is formed by compactifying the information included in $f_b(\boldsymbol{x})$. Every eight neighboring voxels are grouped, forming a bigger one which takes values from 0, when none of the eight voxels lie inside the model's volume, to 8, when all of them lie inside, and $x, y, z \in [0, N]$.

The next step involves the decomposition of $f(\boldsymbol{x})$ into rays and splines. The origin of the rays is the model's center of mass, while the splines are defined on planes normal to these rays. The spatial position of the ray segments is described by the couple $(\boldsymbol{\eta}, \rho)$ where $\boldsymbol{\eta} = [cos\phi sin\theta, sin\phi sin\theta, cos\theta]$ is the unit vector, $A(\boldsymbol{\eta}, \rho) = \{\boldsymbol{x} | \frac{\boldsymbol{x}}{|\boldsymbol{x}|} = \boldsymbol{\eta}, \rho \leq |\boldsymbol{x}| < \rho + \Delta_\rho\}$, where Δ_ρ is the length of the radius segment (Fig. 1a). The same couple is also used to describe the splines: $\Pi(\boldsymbol{\eta}, \rho) = \{\boldsymbol{x} | \boldsymbol{x}^T \cdot \boldsymbol{\eta} = \rho\}$ (Fig. 1b). Each ray segment and spline can be regarded as a 1D and 2D function, $f_1(i)$ and $f_2(i, j)$, respectively. The values of $f_1(i)$ and $f_2(i, j)$ are the values of the voxels that they intersect.

Rays

After the decomposition, $f_1(i)$ is transformed into the frequency domain, applying the 1D Fourier Transform:

$$FT_1(n) = \frac{1}{N} \sum_{i=0}^{N-1} f_1(i) \exp(-\hat{j} \frac{2\pi i k}{N}) \quad (1)$$



(a) (b)

Figure 1: Decomposition.

where $\hat{j} = \sqrt{-1}$, $n = 0, \ldots, N-1$ and N the total number of sampled points on each ray segment. Note that instead of storing the actual coefficient, their amplitudes are calculated and stored.

Splines

Splines are treated in a similar way with rays. For reasons that will be explained in the sequence, the Polar Fourier Transform is computed for each $f_2(i, j)$:

$$FT_2(m, n) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f_2(i, j) \exp(-\hat{j}(\frac{2\pi i m}{N} + \frac{2\pi j n}{N})) \quad (2)$$

where $m, n = 0, \ldots, N-1$. In the DFT shifts in the spatial domain cause corresponding linear shifts in the phase component:

$$FT_2(m, n) \exp[-\hat{j}(am + bn)] \leftrightarrow f_2(i + a, j + b) \quad (3)$$

Thus, the DFT magnitude is invariant to circular translation. Therefore, using discrete polar coordinates, (2) becomes:

$$FT_2(m, n) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f_2(r_{ij}, \xi_{ij}) \exp(-\hat{j}(m r_{ij} + n \xi_{ij})) \quad (4)$$

where

$$r_{ij} = \sqrt{(c_1 i + c_2)^2 + (c_1 j + c_2)^2}$$

$$\xi_{ij} = tan^{-1}(\frac{c_1 j + c2}{c_1 i + c_2}) \quad i, j = 0, \ldots, N-1$$

$$c_1 = \frac{\sqrt{2}}{N-1} \cdot r_{max}, \quad c_2 = -\frac{1}{\sqrt{2}} \cdot r_{max}$$

and rotation is converted to a circular translation of ξ. Again, the harmonic amplitudes $|F_2(m, n)|$ are only considered.

By applying to all $A(\boldsymbol{\eta}, \rho)$ the Fourier Transform and storing the amplitude of only one frequency component, it can be assumed that a new function F is created, whose domain are the points $(\boldsymbol{\eta}, \rho)$ and values

the selected frequency components for each ray segment, $F_n^1(\boldsymbol{\eta},\rho) = FT_1(n;\boldsymbol{\eta},\rho)$, where $FT_1(n;\boldsymbol{\eta},\rho)$ denotes the n-th frequency component of $f_1(\boldsymbol{\eta},\rho)$. In the case of splines $F_{mn}^2(\boldsymbol{\eta},\rho)$ are produced, where $F_{mn}^2(\boldsymbol{\eta},\rho) = F_2(m,n;\boldsymbol{\eta},\rho)$. By restricting to different values of ρ, F_n^1 and F_{mn}^2 can be assumed as a set of concentric spheres. Each sphere will be the input to the Spherical Fourier Transform.

Spherical Fourier Transform

Spherical harmonics [13] are special functions on the unit sphere, generally denoted by $Y_{lm}(\boldsymbol{\eta})$, where $l \geq 0$ and $|m| \leq l$. Any function $g(\boldsymbol{\eta})$ whose domain is the unit sphere, can be expanded as an infinite Fourier series of spherical harmonics:

$$g(\boldsymbol{\eta}_i) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \alpha_{lm} Y_{lm}(\boldsymbol{\eta}_i), \qquad i = 1,\ldots,N_s \quad (5)$$

where N_s is the total number of sampled points on the unit sphere and the expansion coefficients α_{lm} are determined by:

$$\alpha_{lm} = \sum_{i=1}^{N_s} g(\boldsymbol{\eta}_i) Y_{lm}(\boldsymbol{\eta}_i) \frac{4\pi}{N_s} \quad (6)$$

The overall vector length of α_{lm} coefficients with the same l:

$$A(l)^2 = \sum_{m} \alpha_{lm} \quad (7)$$

is preserved under rotation, and this is the reason why the quantities $A(l)$ are known as the rotationally invariant shape descriptors. By using F_n^1 and F_{mn}^2 as input, the results $A_n^1(l)$ and $A_{mn}^2(l)$ are the final descriptor vectors.

Rotation Invariance Conditions

There are two conditions that should be met, in order the whole procedure to result in rotation invariant descriptors. Let us assume that the model is rotated, hence $f(\boldsymbol{x})$ is rotated. This rotation can be assumed as a rotation of $F_n^1(\boldsymbol{\eta},\rho)$ or $F_{mn}^2(\boldsymbol{\eta},\rho)$. This problem is settled by using the Spherical Fourier Transform which results in rotation invariant descriptors. Moreover, the planes that are perpendicular to the axis of rotation, will be rotated around that axis, resulting to a rotated version of their splines. The procedure complies with this condition, since the Polar Fourier is rotation invariant.

3 Experimental Results

As stated in the previous section, using F_n^1 or F_{mn}^2 with different values of n or m,n, will result in different descriptor vector. In our experiments only the low frequencies were considered, therefore only F_n^1 for $n = 0,1,2,3$ and F_{mn}^2 for $m,n = 0,1$ were created. In the case of the Spherical Fourier Transform, the maximum value of l used, was 25.

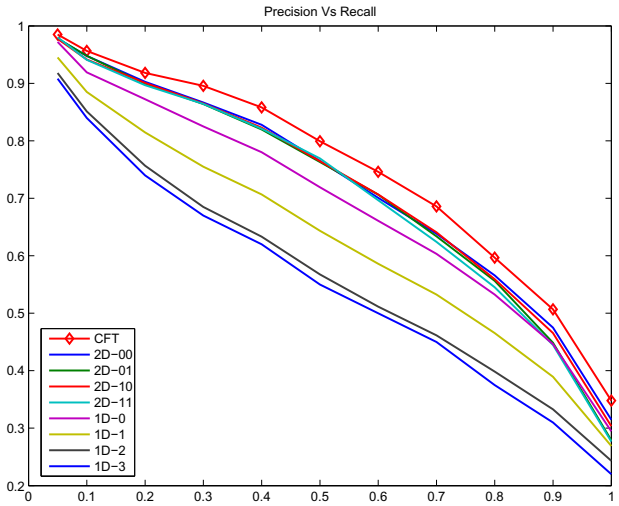


Figure 2: Precision-recall diagram of all descriptor vectors.

In order to compare two models, their L-2 distance was used. Figure 2 illustrates the effectiveness of the 8 produced descriptor vectors, in terms of precision versus recall, using a database which contains 544 3D objects [14]. As it is depicted in Figure 2, the effectiveness of each descriptor vector varies, although the 2D Fourier Transform gives more stable results.

A combination of the aforementioned descriptor vectors was created, which is depicted in Figure 2 as "Combined Fourier Transforms", showing that the combination improves the effectiveness of the method.

In order to combine the descriptor vectors, weights were assigned to each of them, based on their discriminative power. The weight calculation procedure is as follows: Let us assume that a model A belongs to class C_i. All the distances between the descriptor vectors of the objects which belong to same class should be smaller than those between each object of this class and the objects of all other classes. Using this observation, the sum D_{rel} of the distances between A (the descriptors of A) and the rest of the models of class C_i is calculated. Additionally, the sum D_{irrel} of the distances between A and all the models which does not belong to class C_i is calculated. The ratio $\frac{D_{rel}}{D_{irrel}}$ indicates how efficient this descriptor vector is, since a small value of this ratio will indicate that the specific vector has captured distinctive information about the models. The ratios for every descriptor vector are calculated, and they are normalized in order to sum up to 100. These normalized ratios are the weights that are assigned to each descriptor vector. The to-

tal distance between two models A and B can be expressed as:

$$D(A,B) = \sum_{i=0}^{N_f} \alpha_i \cdot D_i(A,B)$$

where N_f is the number of descriptor vectors, $N_f = 8$, α_i is the corresponding weight and $D_i(A,B)$ is the distance between A and B with respect to the i-th descriptor vector.

Figure 3 illustrates the precision-recall diagram of the proposed method and the methods presented in [10, 11, 12]. It is obvious that the proposed method is superior or at least competitive with those most commonly cited in the literature.

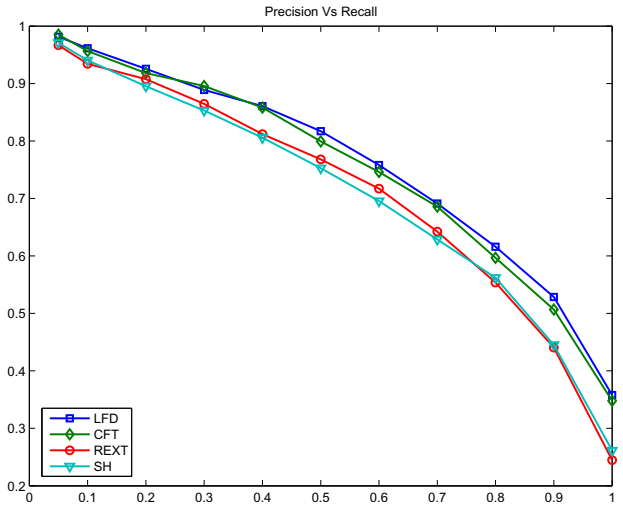


Figure 3: Comparison of the proposed method against the competitive methods in terms of precision-recall.

4 Conclusions

In this paper a new method for efficient content-based search and retrieval of 3D objects was presented. The method proposes the combination of three different Fourier Transforms: the 1D Fourier Transform, the 2D Polar Fourier Transform and the Spherical Fourier Transform. Taking advantage of the rotation invariance ensured by the aforementioned transforms and introducing a new method of assigning proper weights to each of the produced descriptor vectors, a compound descriptor vector is concluded, which is used for the retrieval of 3D objects.

Experiments were performed using a database of 544 3D objects and the results of the proposed method were compared with those of the best known retrieval methods in the literature. The results show clearly that the proposed method outperforms or is competitive with others in terms of precision recall.

References

[1] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton Shape Benchmark", *In Proc. of Shape Modeling International*, Genova, Italy, June 2004.

[2] 3D Search Engine, http://merkur01.inf.uni-konstanz.de/CCCC/

[3] ITI's 3D Search Engine, http://3d-search.iti.gr

[4] M. Bober, "MPEG-7 Visual Shape Descriptors" *IEEE Transactions on Circuits and Systems for Video Technology (CSVT)*, Vol. 11, No. 6, pp. 716-719, Jun. 2001.

[5] C. Zhang, and T. Chen, "Indexing and Retrieval of 3D Models Aided by Active Learning", *In Proc. of ACM Multimedia 2001*, pp.615-616, Ottawa, Canada, 2001.

[6] R. Osada, T. Funkhouser, B. Chazelle and D. Dobkin, "Matching 3D Models with Shape Distributions", *In proc. of Inter. Conf. on Shape Modeling and Applications (SMI2001)*, 154-166, 2001.

[7] I. Kolonias, D. Tzovaras, S. Malassiotis and M. G. Strintzis, "Fast Content-Based Search of VRML Models Based on Shape Descriptors", *IEEE Trans. on Multimedia*, to appear.

[8] M. Novotni and R. Klein, "3D zernike descriptors for content based shape retrieval", *In Proc. of the 8th ACM symposium on Solid modeling and applications*, pp. 216-225, Seattle, Washington, USA, 2003.

[9] N. Canterakis "3D Zernike moments and zernike affine invariants for 3D image analysis and recognition", *In Proc. of the 11th Scandinavian Conf. on Image Analysis*, 1999.

[10] D.Y. Chen, X.P. Tian, Y.T. Shen and M. Ouhyoung, "On Visual Similarity Based 3D Model Retrieval", *Computer Graphics Forum (EUROGRAPHICS'03)*, Vol. 22, No. 3, pp. 223-232, Sept. 2003.

[11] D.Vranic, "An improvement of rotation invariant 3D-shape descriptor based on functions on concentric spheres", *In Proc. of IEEE Int. Conf. on Image Processing (ICIP 2003)*, vol. III, pp. 757760, Barcelona, Spain 2003.

[12] M. Kazhdan, T. Funkhouser and S. Rusinkiewicz, "Rotation Invariant Spherical Harmonic Representation of 3D Shape Descriptors", *Eurographics Symposium on Geometry Processing*, 2003.

[13] D.W.Ritchie, "Parametric Protein Shale Recognition", *PhD Thesis*, University of Aberdeen, 1998.

[14] P. Daras, D. Zarpalas, D. Tzovaras and M. G. Strintzis, "Shape Matching Using the 3D Radon Transform", *In Proc. of 3D Data Processing, Visualization & Transmission (3DPVT 2004)*, Thessaloniki, Greece, Sept 6-9, 2004.