Transform-based Methods for Indexing and Retrieval of 3D Objects

Helin Dutağacı[1], Bülent Sankur[1], Yücel Yemez[2]

[1]*Electrical-Electronics Department,*
Boğaziçi University, Istanbul, Turkey
E-mail: {dutagach, sankur}@boun.edu.tr

[2]*Department of Computer Engineering*
Koç University, Istanbul, Turkey
E-mail: yyemez@ku.edu.tr

Abstract

We compare two transform-based indexing methods for retrieval of 3D objects. We apply 3D Discrete Fourier Transform (DFT) and 3D Radial Cosine Transform (RCT) to the voxelized data of 3D objects. Rotation invariant features are derived from the coefficients of these transforms. Furthermore we compare two different voxel representations, namely, binary denoting object and background space, and continuous after distance transformation. In the binary voxel representation the voxel values are simply set to 1 on the surface of the object and 0 elsewhere. In the continuous-valued representation the space is filled with a function of distance transform. The rotation invariance properties of the DFT and RCT schemes are analyzed. We have conducted retrieval experiments on the Princeton Shape Benchmark and investigated the retrieval performance of the methods using several quality measures.

1. Introduction

Automatic and fast retrieval of three-dimensional objects from large databases is becoming more vital with the increasing number and scope of three-dimensional object models in computer applications such as CAD/CAM, 3D games, virtual reality media, biomedicine, virtual museums, etc. Therefore it is necessary to build efficient indexing schemes that exploit discriminatory shape characteristics of objects of different categories.

** This work was partially supported by TÜBİTAK project 103E038*

In this work we focus on retrieval of objects belonging to general categories, such as cats, tables, airplanes, etc. This type of categorization is subjectively plausible in that it corresponds to what we would picture in our mind while searching an object in the Web. A significant effort has been dedicated in the literature to obtain rotation-invariant features from 3D objects. For example, Zaharia and Preteux **[5]** use shape index histograms to compare the objects. While the shape index, based on principal curvatures, is a powerful object surface attribute, it is computationally tedious and also quite sensitive to noise and resolution level.

Osada et al. **[6]** use various shape functions, such as distance between two arbitrary points on the object surface. The sample distribution of these shape functions become then object signatures. This distribution-based approach is appropriate for shape categorization, for example, used as a pre-classifier, but not for object identification.

An alternative to rotation-invariant features is to obviate the rotation uncertainty. Thus an object can be aligned along its principal axes, e.g., its principal components. Paquet et al. in **[10]** construct "three cords-based histograms" after PCA-based alignment. Ricard et al. **[4]** utilize magnitudes of 3D Angular Radial Transform coefficients applied to the voxelized objects as object descriptors. Since magnitudes of 3D ART coefficients are only invariant to rotations around z-axis, these authors align the object's principal axis with the z-axis prior to computation of ART coefficients. In the same vein, Vranic and Saupe **[7]** take the 3D-DFT of the binary voxel representations. Since the 3D-DFT coefficients are not rotation invariant, DFT is applied after alignment to principal

axes. Vranic and Saupe **[8]** have also experimented spherical harmonics expansion with the PCA aligned objects.

Since PCA alignment may give nonconsistent orientations within a class **[2]**, there have been attempts to extract rotation-invariant features from transform coefficients. Novotni and Klein **[9]** use 3D Zernike moments as descriptors for 3D shape retrieval. Kazhdan et al. **[3]** derive rotation-invariant features from spherical harmonic coefficients. They first rasterize objects in a voxel grid to obtain a 3D binary function, and then compute spherical harmonic invariants of the binary on concentric shells.

In this paper, we investigate and compare two novel 3D indexation methods based, respectively, on Discrete Fourier Transform and on Radial Cosine Transform. First we derive novel rotation-invariant features from DFT coefficients, called "Normalized Spectral Energy (*NSE*)". Secondly, the Angular Radial Transform is simplified and made invariant to yield Radial Cosine functions, to get fast retrieval results. These methods are tested on the Princeton Shape Benchmark **[1],** which provides an object database and its ground-truth categorization.

The paper is organized as follows. In section 2 the two voxel representations we have used are introduced. In section 3 we briefly describe the transform-based rotation invariant descriptors, namely DFT-based and RCT-based descriptors. In section 4, we set up an experiment to test the rotation invariance of DFT-based descriptors. In section 5, we give the metrics for comparison of descriptors. Section 6 is reserved for experimental results. Finally we conclude in section 7.

2. Voxel representations

We render the mesh representation of the object in a 3D voxel grid of size NxNxN, such that the object's center of mass coincides with the center of the 3D grid. The object center, $\mathbf{x}_{center}$, is calculated from the triangular mesh as follows:

$$\mathbf{x}_{center} = \frac{1}{N_t}\sum_t A_t \mathbf{x}_t \ , \qquad (1)$$

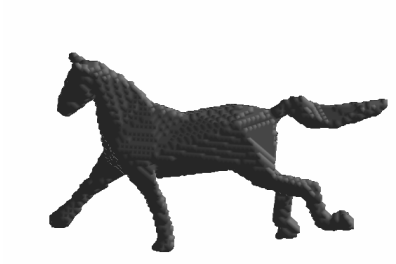
where A_t is the area and $\mathbf{x}_t$ is the center of mass of triangle t, and N_t is the number of triangles in the object.

The object is scaled so that the maximum distance from the center of mass to the surface is N/2. We obtain a 3D binary function $v(\mathbf{x})$ on the voxel grid such that, $v(\mathbf{x})$ is 1 if $\mathbf{x} = [x_1, x_2, x_3]$ is on the object surface, and is 0 otherwise. Figure 1 (a) shows the voxelized form of an object.

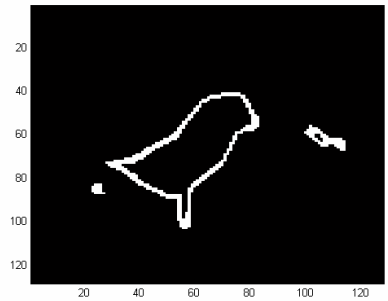
The second voxel representation we use, $v_d(\mathbf{x})$, called inverse distance function (IDF), is a function of 3D distance transform $d(\mathbf{x})$:

$$v_d(\mathbf{x}) = \frac{1}{d(\mathbf{x}) + 1} \qquad (2)$$

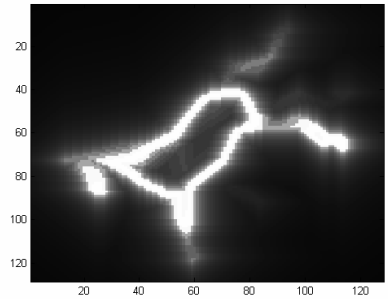
In Eq. 2, $d(\mathbf{x})$ is the minimum L_1 distance of $\mathbf{x}$ to the object surface. We have chosen L_1 distance for rapid distance transform calculation. The 3D function $v_d(\mathbf{x})$ is equal to 1 on the object surface and decreases as the one moves away from the object surface.



(a)



(b)



(c)

Figure 1. (a) Voxelized object; (b) Cross section of the binary function; (c) Cross section of the inverse distance function.

Since $v_d(\mathbf{x})$ decays rapidly to zero towards the corners of the bounding box, one can assume that the range of $v_d(\mathbf{x})$ does not get affected significantly by rotation. In section 4 we demonstrate in fact that the features derived from IDF representation are almost invariant to rotation.

Figure 1 (b) and (c) show the values of the binary function $v(\mathbf{x})$ and the IDF $v_d(\mathbf{x})$ on the same two x_3 slices. IDF is advantageous in that it fills the 3D space so that at any cross section, either planar or spherical, one has more information about the shape content. IDF also provides spatial smoothing so that high-frequency components due to sharp shape details are reduced. The spectral energy for IDF is concentrated at the center in contrast with the larger frequency support in the case of the binary function.

3. Rotation invariant descriptors

3.1. 3D DFT-based descriptors

The 3D DFT of the N-samples of the 3D function $v(\mathbf{x})$ of an object is,

$$V_{DFT}(\mathbf{u}) = \sum_{\mathbf{x}} v(\mathbf{x}) e^{-j2\pi\, \mathbf{x}^T \mathbf{u}/N}, \qquad (3)$$

where $\mathbf{u} = [u_1, u_2, u_3]$ is the frequency vector. If the object is rotated with a rotation matrix $\mathbf{R}$, its binary function becomes is $\tilde{v}(\mathbf{x}) = v(\mathbf{R}\mathbf{x})$. If $\mathbf{x}$ and $\mathbf{u}$ were continuous variables then the DFT of $\tilde{v}(\mathbf{x})$ would be,

$$\tilde{V}_{DFT}(\mathbf{u}) = V(\mathbf{R}\mathbf{u}), \qquad (4)$$

However $\mathbf{R}\mathbf{x}$ and $\mathbf{R}\mathbf{u}$ do not necessarily take integer values, so Eq. 4 must be interpreted as a nearest neighbor vector approximation. It follows that the spectral energy in a sphere centered at the origin of the frequency domain remains constant under rotation. A measure of the spectral energy (*SE*) in a sphere of radius r can be formulated as follows:

$$SE(r) = \sum_{|\mathbf{u}|<r} \left| V_{DFT}(\mathbf{u}) \right| \qquad r = 1,2,..N \qquad (5)$$

Let us define the incremental spectral energy (*ISE*) as the difference of the spectral energies contained within concentric spheres. The incremental spectral energy will then correspond to the total spectral energy in a frequency shell of width 1 at radius r of:

$$ISE(r) = SE(r) - SE(r-1) \qquad (6)$$

Since the shell volume grows proportionally to the radius r^2, we normalize the *ISE* by r^2 and take its square root to balance out large values accumulated in the low-pass shells. The normalized spectral energy (*NSE*) is then used as the DFT-based descriptors of the object:

$$NSE(r) = \sqrt{\frac{ISE(r)}{r^2}} \qquad (7)$$

Apart from being rotation invariant, as demonstrated in Section 4, the *NSE* descriptors provide in a sense a multiresolution representation of the object. *NSE* values at small radii (low-pass region) carry information about the gross shape of the object, while shape details are encoded in the spectral shells at high-frequency radii.

3.2. RCT-based descriptors

The radial cosine transform of the 3D function $v(\mathbf{x})$ is,

$$V_{RCT}(m) = \sum_{\mathbf{x}} v(\mathbf{x}) \phi_m(|\mathbf{x}|) \qquad (8)$$

where $\phi_m(r)$ are radial cosine transform basis functions and are defined as follows:

$$\phi_m(r) = \begin{cases} 1 & if\ r = 0 \\ 2\cos(\pi m r) & otherwise \end{cases} \qquad (9)$$

The RCT coefficients constitute a set of rotation invariant shape descriptors. We will refer these descriptors as $RCT(m)$ for $m = 0,1,2,...M$.

4. Testing rotation invariance of *NSE*

The DFT-based descriptor, namely the normalized spectral energy (*NSE*), has the property of rotation invariance per Eq. 4. However, slight deviation from rotation invariance can occur due to voxelization distortion. **In addition the IDF representation is not totally rotation invariant due to the distance transform values at the corners of the bounding box.** Therefore we test invariance separately on two voxel representations, namely, binary function and IDF. We employ a pair of objects from each of three different categories: Dogs, cats and horses (Figure 2). The objects are intentionally chosen from similar categories to observe the distinguishing power of the *NSE* descriptor as well as to demonstrate its rotation invariance property. The triangular mesh of each object

is arbitrarily rotated in 3D and then it is voxelized. We have 10 rotated versions per object.

The objects are rendered in a 128x128x128 voxel grid to yield the 3D binary function. The DFT-based descriptors of length 64 for each of the 60 objects are obtained and plotted. Figure 4 (a) shows incremental spectral energy (i.e. the total spectral energy in each frequency shell). In Figure 4 (b) and Figure 4 (c) the *ISE* values in the band-pass and high-pass are plotted respectively. It can be observed that rotated versions of the same object have very close *ISE* values. Furthermore the *ISE*s of rotated objects of the same category (for instance Feline1 and Feline 2; or Horse 1 and Horse 2) are close to each other. Note that *ISE*s of cats and dogs are closer to each other than to those of horses; i.e. similar categories yield similar *ISE*s.

Figure 4(a), (b) and (c) demonstrate the multiresolution nature of the DFT-based descriptors. For small frequency indices all rotated objects in the six categories have very similar *ISE* values, since, in a coarser category, these objects can be classified as "quadruped animals". In the band-pass region, the objects' *ISE* values start to be differentiated according to the shape details of their respective categories on a finer scale. As the frequency index increases further, *ISE* values of the rotated versions of the same object also start to differ. This is due to voxelization noise, which shows itself at high frequencies.

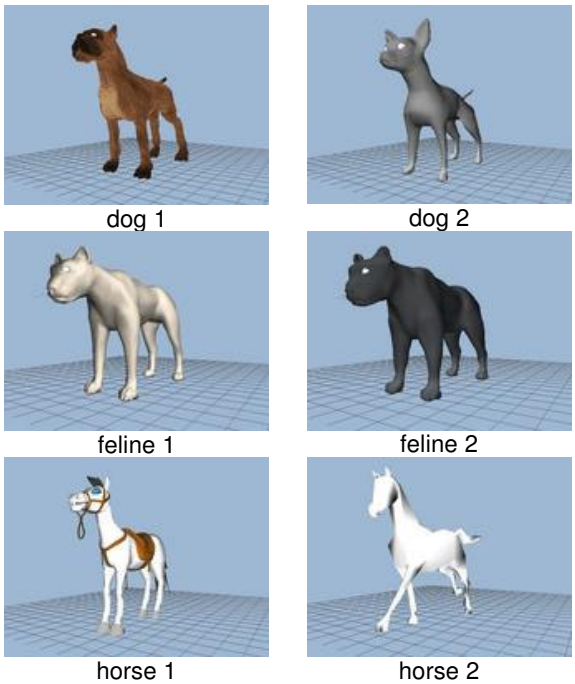


Figure 2. Objects used for testing rotation invariance of DFT-based descriptors.

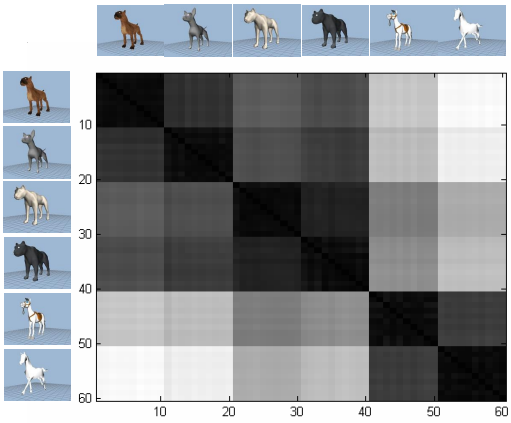


Figure 3. Distance matrix of 60 rotated models of 6 objects. DFT-based descriptors obtained from the IDF voxel representation are used.

When the objects are represented by the IDF instead of the binary function, the resulting incremental spectral energy values become as depicted in Figure 4 (d). Figure 4 (e) and (f) shows the same *ISE* values at band-pass and high-pass. The descriptors are also rotation invariant when IDF is used. The main difference to those obtained with the binary function is that, most of the spectral energy is concentrated at lower frequencies. The total energy per frequency shell saturates at high frequencies as opposed to the increasing total energy for the binary case.

Figure 3 shows the distance matrix of 60 rotated versions of 6 objects. The displayed colors correspond to Euclidean distances between DFT-based features obtained from IDF representation. The identification performance for this small database of 6 objects is 100 per cent when tested over their 60 different rotations.

5. Metrics for retrieval

We have experimented three metrics for matching a query object's descriptor, $\mathbf{w}_q$, against the descriptors of the objects in the database, $\mathbf{w}_d$: The Euclidean distance, the cosine distance and the correlation distance, which are defined in order as:

$$euc(\mathbf{w}_q, \mathbf{w}_d) = \left| \mathbf{w}_q - \mathbf{w}_d \right|, \tag{10}$$

$$cosine(\mathbf{w}_q, \mathbf{w}_d) = 1 - \frac{\mathbf{w}_q^T \mathbf{w}_d}{\left|\mathbf{w}_q\right| \left|\mathbf{w}_d\right|}, \tag{11}$$

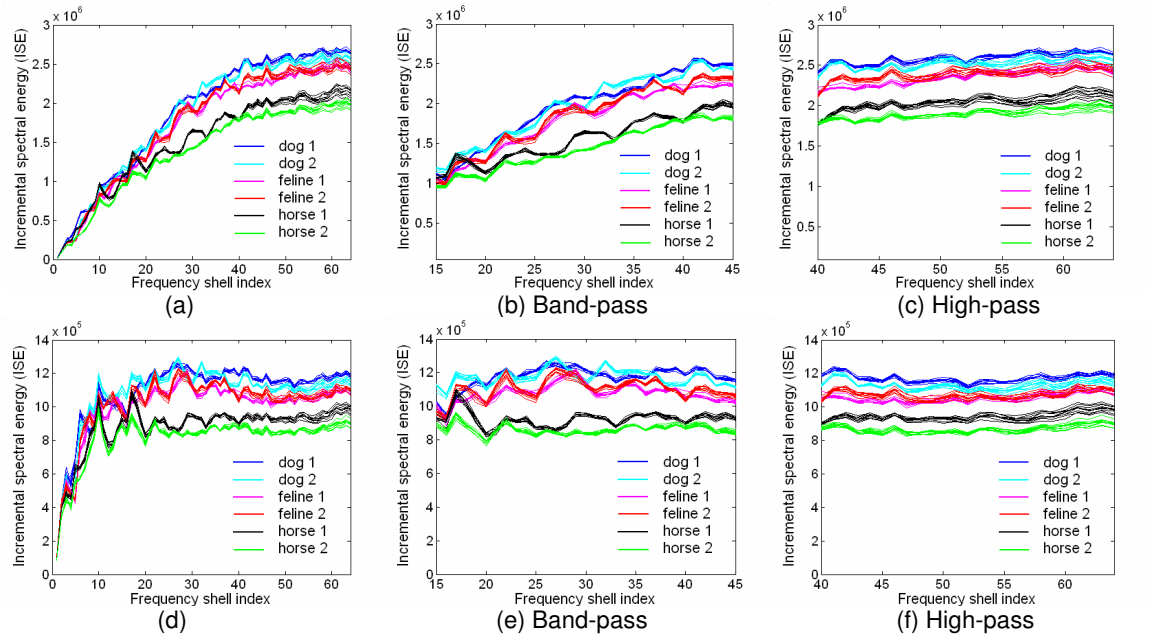


Figure 4. (a) Incremental spectral energy (*ISE*), (binary case); (b) *ISE* at band-pass (binary case); (c) *ISE* at high-pass (binary case); (d) Incremental spectral energy (*ISE*), (IDF case); (e) *ISE* at band-pass (IDF case); (f) *ISE* at high-pass

$$corr(\mathbf{w}_q, \mathbf{w}_d) = 1 - \frac{(\mathbf{w}_q - \hat{\mathbf{w}})^T (\mathbf{w}_d - \hat{\mathbf{w}})}{\left|(\mathbf{w}_q - \hat{\mathbf{w}})\right\|(\mathbf{w}_d - \hat{\mathbf{w}})|}, \quad (12)$$

where $\hat{\mathbf{w}}$ is the mean descriptor of the database.

We have also tested a fourth dissimilarity measure that exploits the three orderings given by the Euclidean, cosine and correlation distances. When a 3D object is queried, its descriptor is compared with those of the objects in the database using the three metrics, and each metric returns a similarity rank. The index of the object with the highest rank sum becomes the identified object. This combined distance measure is given by:

$$\begin{aligned} combined(\mathbf{w}_q, \mathbf{w}_d) = \; & rank[euc(\mathbf{w}_q, \mathbf{w}_d)] \\ & + rank[cosine(\mathbf{w}_q, \mathbf{w}_d)] \\ & + rank[corr(\mathbf{w}_q, \mathbf{w}_d)] \end{aligned} \quad (13)$$

We have also experimented different combinations of city block (L_1), Euclidean (L_2), correlation and cosine distances. Through extensive experimentation we have observed that the combined measure given in Eq. 13 yielded superior results in every case. Therefore we have utilized the combined distance measure in the retrieval experiments.

6. Experimental results

For experimentation, we have used Princeton Shape Benchmark **[1]**, where there are 1814 3D mesh models. 907 of these models are reserved for training system parameters, and 907 of them are provided for testing. Along with the 3D models, a hierarchical classification of the models is provided. The benchmark provides four levels of categorization and the "base level" corresponds to the finest categorization, where the distances between categories are low. For example, the aircrafts are further categorized into fighter jet, commercial, etc. At the base level there are 90 categories in the training set and 92 categories in the test set.

To evaluate the retrieval performance of the systems, precision-recall curves and five scalar values (First tier, second tier, E-measure, discounted cumulative gain and average precision) are used. Let C be the total number of objects in the database that belong to the class of the query object and let C_K be the number of correctly retrieved objects among the K best matches. Then recall is the ratio of C_K to C and precision is the ratio of C_K to K. First and second tier are the precision values when K is equal to $C-1$ and $2C-1$ respectively. Average precision is the

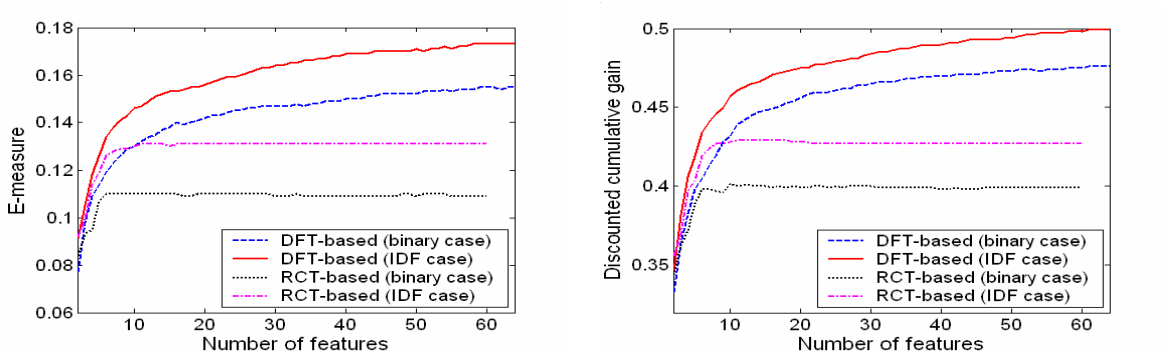


Figure 5. Performance measures versus number of features.

precision averaged over recall values. The E-measure is defined as:

$$E = \frac{2}{1/P_{32} + 1/R_{32}}, \qquad (14)$$

where P_{32} and R_{32} are the precision and recall values when K is equal to 32. Discounted cumulative gain is defined as, 1 + sum 1/lg(i) if the ith best match shape is in the correct class. This sum is then normalized by the maximum possible value **[1]**. These five retrieval evaluation measures are calculated using the software provided by **[1]**.

Figure 5 shows E-measure and discounted cumulative gain as functions of number of features, for the schemes discussed in this paper. These figures are calculated by macro averaging over the classes in the training set of 3D models provided by Princeton Shape Benchmark. It is clear that distance transform-based voxel representation (IDF case) outperforms binary representation for DFT and RCT based methods.

RCT-based features yield the worst results, however the measures saturate at small number of features. We have chosen 16 and 25 as the optimum number of features for binary case and IDF case respectively, although in both cases, the performance measures hardly alter after 10 features. Since they are very easy to calculate and each model is represented with small number of features, RCT-based features can be used for fast pre-classification.

DFT-based scheme gives good results. In both binary and IDF case, the performance measures get their highest values when number of features are 64, i.e. when we utilize *NSE*s of all the frequency shells.

After setting the number of features, we have tested the schemes over the objects in the test set. Table 1 gives the macro-averaged performance measures as well as the number of corresponding features. Zernike and SH columns correspond to results reported for Zernike moments and Spherical Harmonic descriptors, respectively, as in **[9]**.

IDF brings in a great deal of improvement to the performance measures, especially to the RCT-based scheme. DFT-based features give better results than SH-based descriptors in E-measure, discounted cumulative gain and average precision. The size of the

Table 1. Retrieval performance measures on test objects.

	RCT (Binary case)	RCT (IDF case)	DFT (Binary case)	DFT (IDF case)	Zernike [9]	SH [9]
No of features	16	25	64	64	154	32x29
First tier	0.0970	0.1260	0.1690	0.2150	*0.2808**	*0.2675**
Second tier	0.1560	0.1990	0.2590	0.3190	*0.3417**	*0.3239**
E-measure	0.1000	0.1220	0.1570	0.1870	0.1320	0.1237
dcg	0.3740	0.4080	0.4510	0.4970	0.4808	0.4635
Ave. precision	0.2637	0.2993	0.3226	0.3591	0.3028	0.2879

** In [9] first tier and second tier are defined as the precision values when the number of retrieved models, K, is equal to C and $2C$ respectively. This makes first and second tier values higher than those we have calculated by setting, K to $C-1$ and $2C-1$ respectively.*

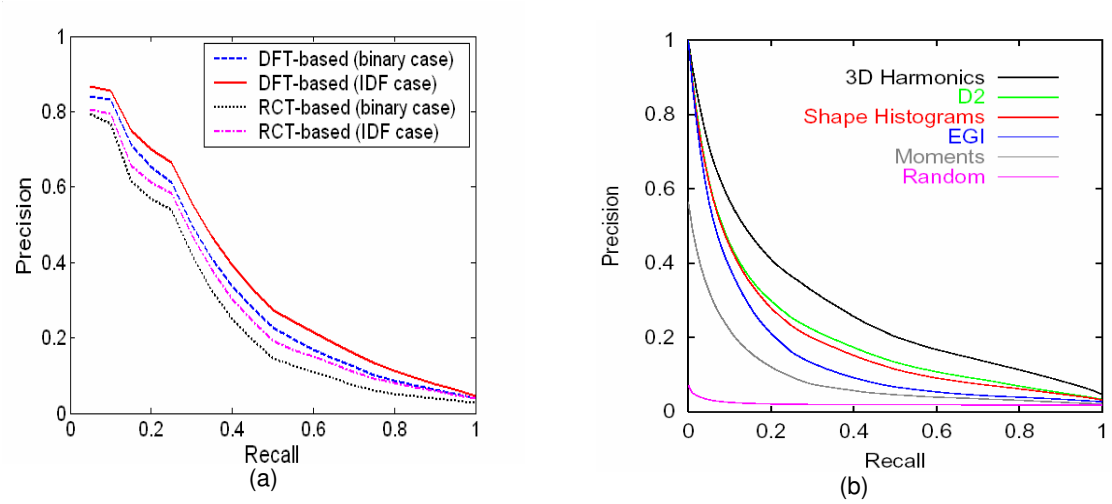


Figure 6. Precision-recall curves

DFT-based descriptor is far smaller than that of SH.

Figure 6 (a) gives the precision-recall curves for the four methods discussed in this paper. Figure 6 (b) is from **[2]** and shows precision-recall curve of SH-based retrieval method. DFT-based features calculated over IDF-based voxel representation gives a precision-recall curve comparable to that of SH-based scheme.

Figure 7 shows retrieval results for "fighter jet". The object with green background is the query object; the objects with blue and red backgrounds correspond to correct and wrong class respectively. All but one objects in the figure correspond to the correct category "fighter jet". Figure 8 shows retrieval results for the query object of class "potted plant".

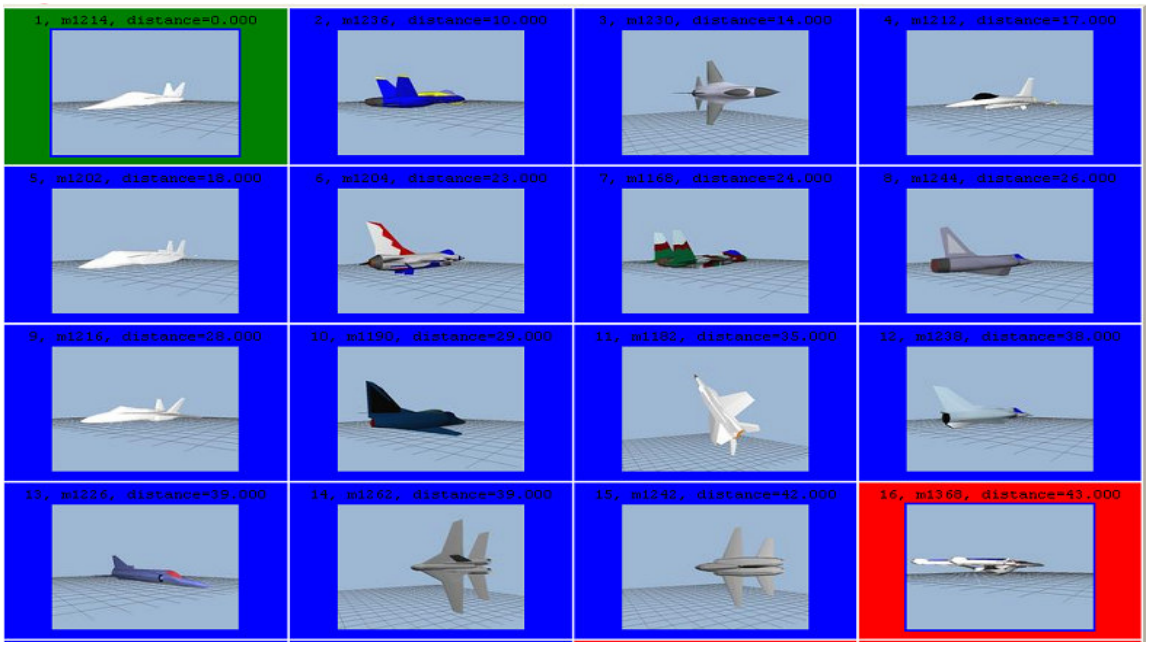


Figure 7. Retrieval results for "fighter jet";
DFT-based descriptors and IDF-based voxel representation are used.

7. Conclusions

In this paper, we have investigated and compared two transformations for 3D object indexing purposes, namely Discrete Fourier Transform and Radial Cosine Transform. We have derived novel shape descriptors from DFT coefficients, and showed that they are invariant to rotation. We have proposed RCT-based descriptors for fast and rough classification of 3D objects. Furthermore we have showed that a distance transform-based voxel representation improves the performance of transform domain shape indexing schemes.

8. References

[1] Princeton Shape Benchmark; 2003. http://shape.cs.princeton.edu/

[2] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, and D. Jacobs, "A Search Engine for 3D Models", *ACM Transactions on Graphics*, 22(1), January 2003, pp. 83-105.

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

[4] J. Ricard, D. Coeurjolly and A. Baskurt, "ART Extension for Description, Indexing and Retrieval of 3D Objects", *17th International Con.e on Pattern Recognition, ICPR 2004*, Cambridge, United Kingdom, 2004.

[5] T.Zaharia and F. Preteux, "Three-dimensional shape-based retrieval within the MPEG-7 framework", *Proceedings SPIE Conference 4304 on Nonlinear Image Processing and Pattern Analysis XII*, San Jose, CA, January 2001, pp. 133-145.

[6] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape Distributions", *ACM Transactions on Graphics*, 21(4), October, 2002, pp. 807-832.

[7] D. V. Vranic and D. Saupe, "3D Shape Descriptor Based on 3D Fourier Transform", *Proceedings of the EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services (ECMCS 2001)*, Budapest, Hungary, September 2001, pp. 271-274.

[8] D. V. Vranic and D. Saupe, "Description of 3D-Shape using a Complex Function on the Sphere", *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2002)*, Lausanne, Switzerland, August 2002, pp. 177-180.

[9] M. Novotni, R. Klein "Shape Retrieval using 3D Zernike Descriptors", *Computer Aided Design*, 2004; 36(11), pp. 1047-1062

[10] E. Paquet, M. Rioux, A. Murching, T. Naveen and A. Tabatabai, "Description of shape information for 2-D and 3-D objects", *Signal Processing: Image Communication*, 16 (2000), pp. 103-1222.

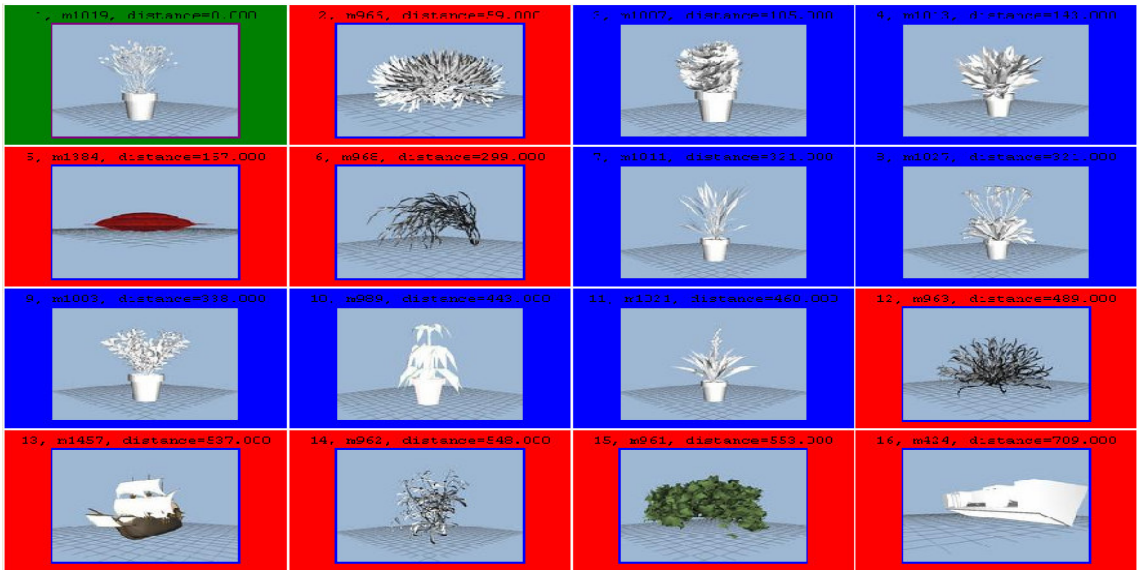


Figure 8. Retrieval results for "potted plant"; DFT-based descriptors and IDF-based voxel representation are used.