Abstract—In this paper, the combination of different projected shape features is proposed for 3D model retrieval. The projection features include the elevation value (depth), the radial distance, and the angle of a surface mesh. For each of the characteristic values (elevation value, radial distance, and angle value), six projection planes represented as grayscale images will be obtained. The MPEG-7 angular radial transform (ART) is then used to compute the feature vector from each projection plane. Experiments conducted on the Princeton Shape Benchmark (PSB) database have shown that the proposed approach outperforms the state-of-the-art descriptors in terms of the DCG score.

Keywords—3D model retrieval; angle value; radial distance; elevation value; ART.

I. INTRODUCTION

Recent developments in advanced techniques for modeling, digitizing and visualizing of 3D models has made 3D models as plentiful as images and video. Therefore, it is necessary to design a 3D model retrieval system which enables the users to efficiently and effectively search interested 3D models. The primary challenge to a content-based 3D model retrieval system is how to extract the most representative features to discriminate the shapes of various 3D models [1].

Vranic et al. applied Fourier transform to the sphere with spherical harmonics to generate embedded multi-resolution 3D shape features [2]. To be rotation invariant, however, pose normalization must be conducted prior to feature extraction. Therefore, Funkhouser et al. proposed a modified rotation invariant shape descriptor based on the spherical harmonics in which no pose normalization is needed [3].

Some popular features used to represent the 3D models are based on the histograms of geometric statistics [4]-[7]. Ankerst et al. tried to search similar 3D models using shape histograms which characterize the area of intersections of a 3D model with a collection of concentric shells and sectors [4]. The MPEG-7 shape spectrum descriptor (SSD) [5] calculates the histogram of the curvatures of all points on the 3D surface. Osada et al. [6] proposed five features, A3, D1, D2, D3, and D4, to represent 3D models by the probability distributions of some geometric properties computed from a set of randomly selected points located on the surface of the model. However, these features are sensitive to tessellation of 3D polygonal models. Thus, Shih et al. [7] proposed grid D2 (GD2) to improve D2. A 3D model is first decomposed into a voxel grid. The distribution of distances between any two randomly selected valid grids is measured to represent a 3D model.

In general, the 3D models also can be described by its 2D silhouettes from different views [8]-[10]. Users can find similar 3D models by 2D shape features. Super and Lu [8] exploit 2D silhouette contours for 3D object recognition. Curvature and contour scale space are extracted to represent each silhouette. Chen et al. [9] proposed the LightField descriptor (LFD) to represent 3D models. LFD is computed from 10 silhouettes. Each silhouette is represented by a 2D binary image. In fact, 2D silhouettes represented by binary images cannot describe the altitude (depth) information of the 3D model from different views. Thus, Shih et al. [10] proposed the elevation descriptor (ED) to represent the altitude information of a 3D model from six different views.

Kuo and Cheng [11] proposed a 3D shape retrieval system based on principal plane analysis. First, each 3D model is projected onto its principal plane. As a result, each 3D model can be represented by a 2D binary image. The feature vectors are then extracted from the binary shape image. However, using only one 2D binary image cannot effectively represent a complex 3D model. Therefore, Shih et al. [12] proposed the principal plane descriptor (PPD) to describe a 3D model with three 2D binary images by projecting it on the principal, second and third planes. Feature vectors are then extracted from these three binary images for 3D model retrieval.

Papadakis et al. [13] proposed two shape descriptors for 3D model retrieval. The 3D model was first aligned by continuous PCA (CPCA) or normal PCA (NPCA). In CPCA, the traditional one, the principal component is analyzed based on the covariance matrix computed from the coordinate vectors of the vertices, whereas in NPCA the covariance matrix is computed from the unit normal vectors of the mesh surfaces. The spherical harmonics was then applied on the filled 3D model to extract two feature vectors from the CPCA and NPCA aligned models separately. Vranic and Saupe proposed a modified PCA which used the corresponding triangle areas as weighting factors for covariance matrix computation [14]. The directions of 20 vertices on a dodecahedron and the distances computed from the center point to the farthest intersections were used as features to search similar 3D models.

In this paper, the combination of different projected shape features, including the elevation value (depth) [10], the radial distance [15], and the angle value of a surface mesh, will be proposed for 3D model retrieval. The rest of the paper is organized as follows. In Section 2, the proposed 3D model retrieval system will be described. Section 3 gives some
follows:

\[ I^i(x, y) = \theta(x, y, z_{\max}(x, y)), \ -50 \leq x, y \leq 50 \]  
\[ I^j(x, z) = \theta(x, y_{\max}(x, z), z), \ -50 \leq x, z \leq 50 \]  
\[ I^k(y, z) = \theta(x_{\min}(y, z), y, z), \ -50 \leq y, z \leq 50 \]  
\[ I^l(x, y) = \theta(x, y_{\min}(x, y), z), \ -50 \leq x, y \leq 50 \]  
\[ I^m(x, z) = \theta(x_{\max}(x, z), y, z), \ -50 \leq x, z \leq 50 \]  
\[ I^n(y, z) = \theta(x_{\min}(y, z), y, z), \ -50 \leq y, z \leq 50 \]  

where

\[ z_{\max}(x, y) = \max(zV(x, y, z)) \]  
\[ y_{\max}(x, z) = \max(yV(x, y, z)) \]  
\[ x_{\max}(y, z) = \max(xV(x, y, z)) \]  
\[ z_{\min}(x, y) = \min(zV(x, y, z)) \]  
\[ y_{\min}(x, z) = \min(yV(x, y, z)) \]  
\[ x_{\min}(y, z) = \min(xV(x, y, z)) \]

Experimental results to show the effectiveness of the proposed features. Finally, conclusions are given in Section 4.

II. PROPOSED 3D MODEL RETRIEVAL SYSTEM

First, each 3D model is decomposed into a number of voxels. Second, the principal planes method \([12]\) will be used for pose alignment of each 3D model. Third, different features describing variant shape characteristics of each decomposed voxel will be projected onto six viewing planes. Finally, the MPEG-7 ART will be applied to each projection plane to extract the feature values of each 3D model.

A. 3D Model Normalization and Alignment

Given a 3D model, its pose is first aligned by the principal planes method \([12]\). The smallest bounding cube that circumscribes the 3D model is then decomposed into a voxel grid of size 100×100×100. A voxel located at coordinates \((x, y, z)\) will be defined as an opaque voxel, notated as \(Voxel(x, y, z) = 1\), if there is a mesh located within this voxel; otherwise, the voxel is defined as a transparent voxel, notated as \(Voxel(x, y, z) = 0\). To be robust for translation and scaling, the 3D model is transformed such that the model’s mass center becomes \((0, 0, 0)\) and the average distance from all non-zero voxels to the mass center is 25.

Once the pose of a 3D model is aligned, the angle value, radial distance, and elevation (depth) value of each opaque voxel will be projected onto six projection planes from which the feature value will be extracted to represent each 3D model. These six projection planes denote the six different views of the 3D model. The angle value describes the angle between the normal vector \(n\) of the mesh and the ray connecting the mass center of the 3D model and the center point of the mesh (see Fig. 1). The radial distance denotes the distance from the opaque voxel to the mass center of the 3D model (see Fig. 2). The elevation value describes the distance from the opaque voxel to the projection plane (see Fig. 2). These values can capture different shape characteristics (the orientation and location) of each opaque voxel. Each projection plane is represented by a gray level image in which the gray value denotes the angle value, radial distance, or elevation value.

B. Angle Value Projection

The angle value tries to capture the orientation of the model’s surface. For each voxel located at \((x, y, z)\), let \(r\) denote the vector connecting the mass center of the 3D model and the center point of the surface mesh. The angle between the vector \(r\) and the normal vector \(n\) of the mesh serves as one of the characteristics of the surface mesh (see Fig. 1). The cosine of the angle between \(r\) and \(n\) will be treated as the projected angle value of the voxel located at \((x, y, z)\):

\[ \theta(x, y, z) = \frac{n^T r}{\|n\| \|r\|} \times 255 \]  

Let the six angle projection planes be notated as \(I^i\), \(k = 1, 2, \ldots, 6\). Then, the gray value, indicating the projected angle value, of each pixel on these projected images is defined as follows:

![Fig. 1 The angle between the normal vector \(n\) of the surface mesh and the vector \(r\) that connects the mass center of the 3D model and the center point of the surface mesh.](image)

C. Radial Distance Projection

The radial distance tries to capture the location of the model’s surface. For each voxel located at \((x, y, z)\), the radial distance is measured as its distance from the mass center of the 3D model. The radial distance is defined as follows (see Fig. 2):

\[ RD(x, y, z) = \sqrt{x^2 + y^2 + z^2} \]  

Let the six radial distance projection planes be notated as \(I^k\), \(k = 1, 2, \ldots, 6\). Then, the gray value, indicating the projected radial distance, of each pixel on these projected images is defined as follows:

\[ I^i(x, y) = \max(RD(x, y, z)V(x, y, z)), -50 \leq x, y \leq 50 \]  
\[ I^j(x, z) = \max(RD(x, y, z)V(x, y, z)), -50 \leq x, z \leq 50 \]  
\[ I^k(y, z) = \max(RD(x, y, z)V(x, y, z)), -50 \leq y, z \leq 50 \]  
\[ I^l(x, y) = \max(RD(x, y, z)V(x, y, z)), -50 \leq x, y \leq 50 \]  
\[ I^m(x, z) = \max(RD(x, y, z)V(x, y, z)), -50 \leq x, z \leq 50 \]  
\[ I^n(y, z) = \max(RD(x, y, z)V(x, y, z)), -50 \leq y, z \leq 50 \]  

D. Elevation Projection

The elevation value tries to capture the altitude (depth) information of the model’s surface to each viewing (or projection) plane. For each voxel located at \((x, y, z)\), the
The MPEG-7 angular radial transform (ART) is an orthogonal unitary transform. ART consists of a complete set of orthonormal sinusoidal basis functions which are defined on a unit disk in the polar coordinate system. Let \( f(\rho, \theta) \) denote the gray level of the pixel located at \((\rho, \theta)\) on the projection image \( I \). The ART coefficient of the projection image \( I \) can be computed as follows:

\[
F(n, m) = \int \int V_{n,m}(\rho, \theta)f(\rho, \theta)\rho d\rho d\theta
\]

where \( F(n, m) \) is the ART coefficient of order \( n \) and \( m \), \( V_{n,m}(\rho, \theta) \) is the complex ART basis function.

The ART descriptor is formed by the magnitudes of all complex ART coefficients. The default ART descriptor consists of 36 coefficients, \( |F(n, m)| \) for \( 0 \leq n \leq 2 \) and \( 0 \leq m \leq 11 \). In summary, the ART vector extracted from the projection image \( I \) can be represented as follows:

\[
F = \left[ |F(0, 0)|, |F(0, 11)|, |F(1, 0)|, \ldots, |F(1, 11)| \right]
\]

Let \( y^x = [(x^x_1)^T, \ldots, (x^x_6)^T]^T \) denote the feature vectors extracted from the six projection planes indicating the angle value, radial distance, and elevation value of the query model. In the same way, let \( y^z = [(y^z_1)^T, \ldots, (y^z_6)^T]^T \), \( y^y = [(y^y_1)^T, \ldots, (y^y_6)^T]^T \), and \( y^e = [(y^e_1)^T, \ldots, (y^e_6)^T]^T \) denote respectively the feature vectors extracted from the corresponding six projection planes of the matching model in the database. The distance between the query model and the matching model corresponding to the angle value, radial distance, and elevation value are defined as follows:

\[
d^y(x^y, y^z) = \sum_{i=1}^{6} \|x^y_i - y^z_i\| = \sum_{i=1}^{6} \sum_{j=1}^{k} |x^y_i - y^z_i|
\]

\[
d^z(x^z, y^y) = \sum_{i=1}^{6} \|x^z_i - y^y_i\| = \sum_{i=1}^{6} \sum_{j=1}^{k} |x^z_i - y^y_i|
\]

\[
d^e(x^e, y^e) = \sum_{i=1}^{6} \|x^e_i - y^e_i\| = \sum_{i=1}^{6} \sum_{j=1}^{k} |x^e_i - y^e_i|
\]

The overall distance between the input query model and the matching model is defined as the sum of the angle distance, radial distance, and elevation distance:

\[
d(x, y) = d^y(x^y, y^z) + d^z(x^z, y^y) + d^e(x^e, y^e)
\]

The matching models that have the minimum overall distances will be regarded as the retrieved similar models.

### III. Experimental Results

To demonstrate the effectiveness of the proposed method for different 3D models, some experiments have been conducted on the Princeton Shape Benchmark (PSB) database [17]. The PSB database contains 1814 models (161 classes) which are divided into 907 training models (90 classes) and 907 test models (92 classes). The discounted cumulative gain (DCG) [18], will be employed to compare the performance of different approaches. DCG at the \( k \)-th rank is defined as follows:

\[
\text{DCG}_k = \frac{\sum_{i=1}^{k} \frac{2^{y_i} - 1}{\log_2(i+1)}}{\sum_{i=1}^{k} \frac{2^{y_i} - 1}{\log_2(i+1)}}
\]

where \( L_k = 1 \) if the \( k \)-th model in the ranked retrieval list and the query one belong to the same class; otherwise, \( L_k = 0 \). The overall DCG score for a query model is defined as \( \text{DCG}_{\text{max}} \), where \( k_{\text{max}} \) is the total number of models in the database. It is clear that if the models appearing in the head of the retrieval list have the same class label as the query one, the evaluated DCG score will be larger than the DCG score associated with the retrieval result in which models with identical class label to the query one appear in the tail of the retrieval list.

In our experiments, each model in the database will be presented as a query one to measure the DCG score. Table I
compares the retrieval results of the proposed approaches with other state-of-the-art descriptors in terms of DCG score. In this table, ART-A, ART-R, and ART-E denote the approaches that use the ART feature vectors extracted from the projection planes of the angle value, radial distance, and elevation value, respectively. In addition, the combination of these three feature vectors is notated ART-ARE. It can be seen that each of the proposed approaches outperform the other descriptors in terms of DCG score.

### TABLE I. COMPARISON OF THE PROPOSED APPROACHES WITH OTHER DESCRIPTORS IN TERMS OF THE DCG SCORE (%). NOTE THAT THE APPROACHES MARKED WITH * ARE IMPLEMENTED BY AKGUL ET. AL. AND ORIGINALLY APPEARED IN [18].

<table>
<thead>
<tr>
<th>Approach</th>
<th>DCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>SH [3]</td>
<td>58.35</td>
</tr>
<tr>
<td>SSD [5]</td>
<td>48.07</td>
</tr>
<tr>
<td>GD2 [7]</td>
<td>60.91</td>
</tr>
<tr>
<td>LF [9]</td>
<td>64.30</td>
</tr>
<tr>
<td>ED [10]</td>
<td>67.04</td>
</tr>
<tr>
<td>RISH [11]*</td>
<td>58.40</td>
</tr>
<tr>
<td>PPD [12]</td>
<td>65.86</td>
</tr>
<tr>
<td>CRSF [13]</td>
<td>66.80</td>
</tr>
<tr>
<td>DBF [18]</td>
<td>65.90</td>
</tr>
<tr>
<td>DSR-DBF [18]</td>
<td>70.20</td>
</tr>
<tr>
<td>AED [19]</td>
<td>70.29</td>
</tr>
<tr>
<td>DED [20]</td>
<td>66.92</td>
</tr>
<tr>
<td>CED [20]</td>
<td>68.04</td>
</tr>
<tr>
<td>EGI [21]</td>
<td>43.80</td>
</tr>
<tr>
<td>SH-GEDT [22]</td>
<td>58.40</td>
</tr>
<tr>
<td>DBI [23]*</td>
<td>66.30</td>
</tr>
<tr>
<td>DSR [23]*</td>
<td>66.50</td>
</tr>
<tr>
<td>SIL [23]*</td>
<td>59.70</td>
</tr>
<tr>
<td>SWD [24]*</td>
<td>65.40</td>
</tr>
<tr>
<td>3DHT [25]*</td>
<td>57.70</td>
</tr>
<tr>
<td>CAH [26]*</td>
<td>43.30</td>
</tr>
<tr>
<td>REXT [27]*</td>
<td>60.10</td>
</tr>
<tr>
<td>Proposed ART-A</td>
<td>69.98</td>
</tr>
<tr>
<td>Proposed ART-R</td>
<td>71.60</td>
</tr>
<tr>
<td>Proposed ART-E</td>
<td>71.90</td>
</tr>
<tr>
<td>Proposed ART-ARE</td>
<td>73.53</td>
</tr>
</tbody>
</table>

### IV. CONCLUSIONS

In this paper, a new feature descriptor which combines the projected shape features, including the angle value, radial distance, and elevation value of each surface mesh, is proposed for 3D model retrieval. For each of the characteristic values, six projection planes represented as gray-level images will be generated. MPEG-7 angular radial transform (ART) is then used to compute the feature vector from each projection plane. Experiments conducted on Princeton Shape Benchmark (PSB) database have shown that the proposed approaches outperform other state-of-the-art descriptors in terms of DCG score.

### REFERENCES


