

## AN EFFICIENT 3D MODEL RETRIEVAL BASED ON PRINCIPAL AXES ANALYSIS AND FEATURE INTEGRATION

SHENG-FUU LIN\*, CHIN-CHIA WU, CHI-YAO HSU  
and DOU-CHIH HSU

*Department of Electrical Engineering  
National Chiao Tung University, 1001 Ta Hsueh Road  
Hsinchu, Taiwan 300, R.O.C.*

*\*sflin@mail.nctu.edu.tw*

Three-dimensional (3D) model retrieval has gathered great importance in recent years, since the number of available 3D models on the Internet has drastically increased. Many content-based 3D model retrieval approaches have been proposed. Among these methods, visual similarity-based methods have shown higher retrieval accuracy. However, because these methods capture enormous shape features from different viewpoints or locations, a large amount of calculation and comparison is necessary. Furthermore, there is a trade-off between retrieval accuracy and speed. In this paper, a 3D model retrieval method constituting Continuous Principal Component Analysis (CPCA), Fourier descriptors, and Zernike moments is proposed. CPCA is applied to extract significant shape features based on projecting the model along the principal axes. Then, Fourier descriptors and Zernike moments are used to provide shape descriptors with rotation invariants. In addition, a feature integration process combines them. A strategy of similarity measure is proposed to solve the axes switching problem. To conclude, the experimental results show that the approach outperforms SECTORS<sup>2</sup> and D2,<sup>18</sup> and has slightly better retrieval results than Light Field Descriptor (LFD)<sup>6</sup> and spin-image signatures.<sup>3</sup> Moreover, the approach is more efficient and the storage size is much less.

*Keywords:* 3D model retrieval; continuous principal component analysis; Fourier descriptor; Zernike moment.

### 1. Introduction

In recent years, the high development of modeling tools and scanning devices has made the acquisition of three-dimensional (3D) models much easier. A variety of techniques have been developed to reconstruct 3D models, like photometric stereo<sup>13</sup> or a physically based method.<sup>10</sup> Moreover, users can search and download an increasing number of 3D models on the Web directly. To find the appropriate models that users want, a 3D model retrieval method is necessary to search a large variety of 3D models. The traditional approach called text-based search engine usually uses

\*Author for correspondence

keywords in filenames, captions, or context to search for 3D models since it is simple and intuitive.<sup>15</sup> Unfortunately, this method fails in many cases, such as bad annotations, different language definitions, and related keywords too common to find relevant matches. In addition, annotations added by humans are often subjective and must deal with synonyms and semantic meaning similarity. This means the text-based method is too limited and ambiguous to be suitable for applications with a large number of 3D models.

In contrast, using the model shape data itself, called content-based search engine, is a more promising approach for implementing a retrieval system. In recent years, a growing body of research about content-based retrieval has been conducted in many fields, and the vast literature devoted to this topic has been reviewed on several occasions.<sup>5,14,24,28</sup> Hence, the efficiency and accuracy of content-based 3D model retrieval methods are considered important.

The survey paper proposed by Tangelder and Velkamp<sup>24</sup> divided content-based retrieval methods into three broad categories: (1) graph-based methods, (2) feature-based methods, and (3) geometry-based methods. Graph-based methods attempt to extract a geometric meaning from a 3D model using a graph to show the relationship to shape primitives. For example, Reeb graph approaches<sup>4,8</sup> use topological structures of 3D models to estimate the similarity between two models. Graph-based methods are difficult to apply to natural models like animals, but are suitable for models created from computer-aided design (CAD) especially. The skeleton-based approach proposed by Sundar *et al.*<sup>23</sup> is a well-known technique for retrieving 3D models. It captures crucial information about the structure of objects but needs several parameters to be set, such as threshold values for thinning and clustering. Furthermore, the difficulties of skeleton-based methods are discrimination between skeletal representations and automatic skeleton extraction from varying 3D models.

Feature-based methods use global or local features to encode 3D models to feature vectors. The retrieval is conducted by comparing the distance for each feature vector. Generally, many researchers have focused on feature-based methods, because varying 3D models can be represented by lower dimension features. Additionally, these features can handle rotation and even scale invariants. Global feature methods use the distribution of vertices or polygons to determine the similarity between 3D models.<sup>7,9,12,17–19,27</sup> A shape descriptor using a sliced image histogram was proposed by Park.<sup>20</sup> To measure shape distributions of 3D models is a famous feature-based method, and its key idea is to represent the signature of a 3D model as a shape distribution. For example, Funkhouser *et al.*<sup>7</sup> described the spherical harmonics shape descriptor based on voxels, and Vranić and Saupe<sup>27</sup> proposed a description for 3D shapes using a complex function. Osada *et al.*<sup>18</sup> proposed D2 shape distributions using the histogram of distances between a pair of points on the surface for 3D models. Hence, this method is simple, fast, and robust, and can measure the global geometric properties of an object.

Besides the shape distribution approach, Ankerst *et al.*<sup>1,2</sup> proposed a technique using 3D shape histograms for similarity searching. The shape histograms were yielded

from three different models, including shell model, sector model, and combined model. Their experimental results showed that the approach has high classification accuracy and efficient query process. Nevertheless, it still needs to determine the optimal number of bins, which may take much to evaluate the experimental results. Shape-based approaches still have many challenges such as partial matching problem and the appropriate definition of shape descriptors.

On the contrary, the local features represent a 3D model by using neighborhood information on the surface, and the local features are usually provided by 2D histogram images, such as spin-image<sup>11</sup> and mutual angle-distance histogram.<sup>16</sup> Assfalg<sup>3</sup> proposed a content-based retrieval approach by using spin-image signatures, which modifies the original spin image. The spin-image signatures provided an efficient and discriminative way to retrieve a 3D model and used a fuzzy c-means to cluster these signatures. The advantages of this retrieval system are less storage size and short matching time compared with principal component analysis (PCA) spin-images. However, its precision at high recall rate is still less than the light field descriptor (LFD) method proposed by Chen *et al.*<sup>6</sup>

LFD is a visual similarity-based 3D model retrieval approach that uses LFDs to compare similarities among models. The concept is that if two 3D objects are similar, their appearances are also similar from any view angle, thus the LFD represents a model as a collection of feature images rendered from 20 different views uniformly located on a sphere. Among these feature images, only ten are needed to represent a 3D model since the feature images projected from two opposite vertices are identical. Moreover, LFD has shown that it is robust against rigid transformation, noise, and model degeneracy, and its performance was evaluated on base classification testing set of Princeton shape benchmark (PSB).<sup>22</sup> According to the experimental results of PSB, LFD provided the best retrieval precision among 12 shape-matching algorithms, even if it only used two-dimensional (2D) projection as a representation of 3D objects, which is less discriminating than 3D matching. Nevertheless, the generation and comparison time of LFD are higher than the other descriptors, since it needs to find a massive amount of possible corresponding images from the models and database.

The study presented in this paper attempts to overcome these shortcomings of LFD. LFD uses 20 images to generate pose-independent features for comparing 3D models. On the other hand, PCA is a common method to find the principal axes of a 3D model. Furthermore, the model can be transformed to the same reference frame, so the pose should be consistent for each model. Thus, once the principal axes are aligned, only three images are yielded by projecting a 3D model along its principal axes. However, Funkhouser *et al.*<sup>7</sup> suggested that two important problems could drastically affect model retrieval results. The first is the orientation of principal axes, which causes the projected images to have different appearances and rotated images between two models, and the second is the axes switching problem, which causes different image orders between two models. Consequently, these issues are addressed in this paper.

To that end, a PCA-based method for 3D model retrieval using principal axes integrated descriptor (PAID), which employs Fourier descriptors and Zernike moments, is proposed. The fundamental of the method assumes that two similar 3D models have similar principal axes, and their three projected images are similar to each other. Thus, a continuous PCA (CPCA) is adopted to find the precise principal axes of 3D models. Two rotation invariant features, Fourier descriptors, and Zernike moments, minimize the effect of the rotation of principal axes, and provide the descriptions about both shapes and image contents. In addition, a mechanism is advanced to determine the correct combination of image orders to solve the axes switching problem. Furthermore, the feature integration has advantages of both shape and content representations, and brings better retrieval performance. The proposed method provides high accuracy and efficiency in 3D model retrieval. Moreover, the dimensions of features are reduced significantly less than LFD. The experimental results show that the retrieval performance of the proposed method is superior to other competitions.

The organization of the rest of this paper is as follows. The representation of 3D models is discussed in Sec. 2. The proposed retrieval system including online and offline stages is described in Sec. 3. In Sec. 4, the experimental results are presented. Finally, Sec. 5 concludes this paper.

## 2. The Representation of 3D Models

This section mainly discusses the representation of a 3D model. Figure 1 shows the diagram of the proposed feature extraction method. First, this paper uses CPCA to find the principal axes of a 3D model, and three projected images of the model are generated by projecting the model along the principal axes. Next, the images as input apply Fourier descriptors and Zernike moments to obtain shape features. Finally, a feature integration process combines them into an integrated feature called PAID to represent 3D models, which will be discussed in Sec. 3.

### 2.1. CPCA method

PCA involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The original PCA method used points of a 3D model, such as vertices on the surface, to find principal axes. For similar models, however, it may obtain different results because varied sizes of triangles are not considered. To overcome this shortcoming, Vranić *et al.*<sup>25</sup> proposed a weighted PCA method that uses the neighbor triangle area of each vertex to compute the weightings, and then utilized these to find the proper principal axes. Although this method could reduce the effect of non-uniformly distributed vertices, it did not take account of a variety of distributions with varying triangle areas. Due to this problem, Vranić *et al.*<sup>26</sup> presented CPCA to address it.

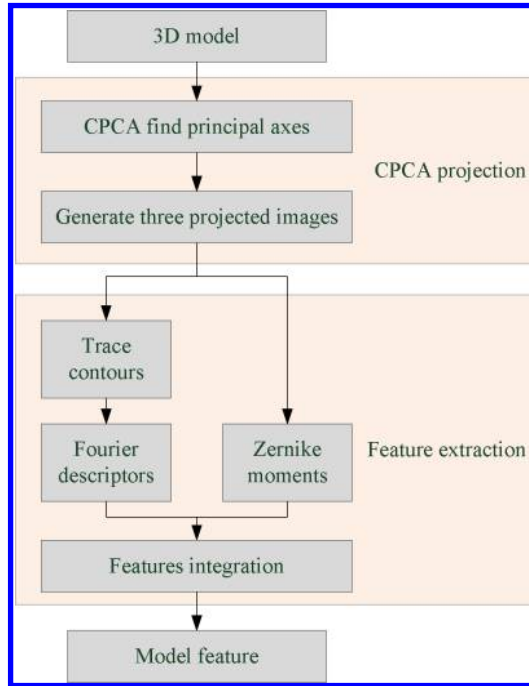


Fig. 1. The proposed feature extraction method.

To take variations in geometry for 3D models into consideration, this paper applies CPCA to compute the precise principal axes of models, illustrated as follows. Regard a given triangle mesh as a set of triangles,  $T = \{t_1, \dots, t_m\}$ ,  $t_i \in R^3$ , and create a table with a list of indices of three vertices for each triangle, and let  $E = \cup_{i=1}^m t_i$  be the point set of all triangles. Let  $A_i$  be the area of triangle  $t_i$ ,  $i = 1, \dots, m$ . Assuming the triangles do not intersect each other so that the overall surface in the model is:

$$A = A_1 + \dots + A_m = \int_E dv. \tag{1}$$

To get the translation invariance, this method formed a point set  $E_1$  by translating the center of gravity to the origin,  $c$ . The point set is described as:

$$E_1 = E - c = \{u | u = v - c, v \in E\}. \tag{2}$$

To ensure rotation invariance, it first computed the covariance  $3 \times 3$  matrix  $C = \frac{1}{A} \int_{E_1} v \cdot v^T dv$ , and calculated the eigenvalues and corresponding eigenvectors of  $C$ . Next, it sorted the eigenvalues in decreasing order and scaled corresponding eigenvectors to Euclidean unit length. The rotation matrix  $H$  is formed by locating the scaled eigenvectors as rows of  $H$ . Then, a new set  $E_2$  can be obtained by using the

rotation matrix  $H$  to rotate  $E_1$ . The operation is shown as:

$$E_2 = HE_1 = \{v | v = Hu, u \in E_1\}. \quad (3)$$

To get the reflection invariance, this method multiplied points in  $E_2$  by a diagonal matrix  $D$ , which is represented as follows:

$$D = \text{diag}(\text{sign}(f_x), \text{sign}(f_y), \text{sign}(f_z)), \quad (4)$$

where  $f_x = \frac{1}{A} \int_{E_2} \text{sign}(v_x) v_x^2 dv$ ,  $f_y$  and  $f_z$  are similar, and  $v = (v_x, v_y, v_z) \in E_2$ .

The scaling invariance is accomplished by scaling the set  $E_2$  by the inverse of  $m = \sqrt{(m_x^2 + m_y^2 + m_z^2)/3}$ , where  $m_x, m_y$ , and  $m_z$  are the average distances of points  $v \in E_2$  from the  $yz$ -,  $xz$ -, and  $xy$ - coordinate hyperplanes, respectively. Finally, the affine map is defined as:

$$\tau(v) = m^{-1}DH(v - c). \quad (5)$$

After applying Eq. (5), the 3D model can be aligned with its principal axes. Subsequently, three feature images are obtained by projecting the triangle surfaces of the 3D model along the principal axes. For instance, Fig. 2 shows two identical airplanes with different poses. After applying the CPCA method, they have the same 2D feature images, as shown in Fig. 3. To this end, the 3D model retrieval problem can be simplified to match the three corresponding feature images.

However, the CPCA method cannot ensure that similar objects have the same orientation of principal axes. This causes the rotation problem of corresponding

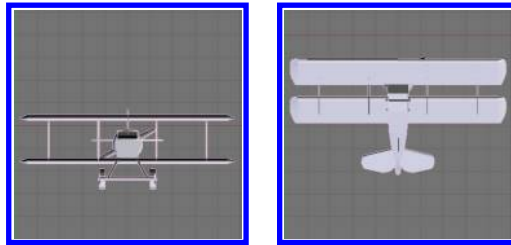


Fig. 2. Two identical airplanes (no. 1119 in PSB) with different poses.

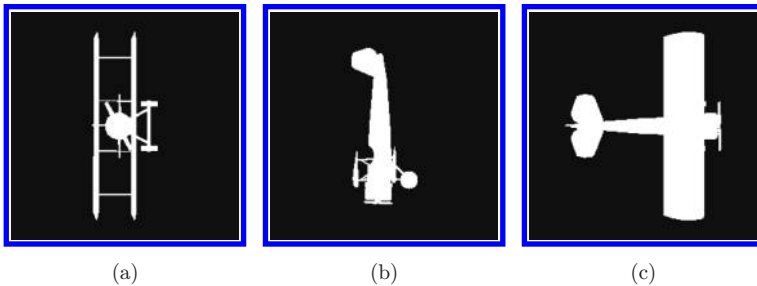


Fig. 3. Three principal projected images of both airplanes in Fig. 2. (a)–(c) are the projected image of first, second, and third principal axes, respectively.

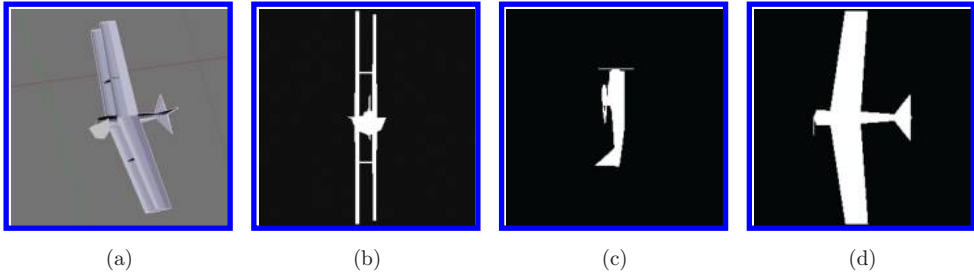


Fig. 4. The 3D model of a single-propeller plane and its three principal projected images. (a) The 3D model. (b) The first projected image. (c) The second projected image. (d) The third projected image.

feature images. For example, Figs. 3(a)–3(c) and Figs. 4(b)–4(d) show three feature images of two similar airplanes, respectively. The feature images corresponded to the second axis (Figs. 3(b) and 4(c)), rotate  $180^\circ$  to each other, thus they cannot be compared directly. Hence, this paper adopts Fourier descriptors and Zernike moments descriptors to compensate for the rotation problem of the feature images.

Besides the image rotation problem, using the CPCA method could cause the axes switching problem, that is, the order of corresponding principal axes is different. Figure 5 illustrates this problem by showing the order of principal projected images. The order of Figs. 5(b)–5(d) and Figs. 5(f)–5(h) is according to the results from applying the CPCA method to these two objects, respectively.

However, to compare images correctly, the corresponding relation should be that Figs. 5(b)–5(d) correspond to Figs. 5(f)–5(h), respectively. If the corresponding feature images are switched, the comparison error increases drastically. To solve this problem, the correct corresponding relation can be found by searching for the minimal distance between two feature vectors from three projected images. This process will be used in a similarity measure procedure and will be discussed in detail in Sec. 3.2.

## 2.2. Fourier descriptors

Fourier descriptors describe the shape in terms of representing the contour of the shape as a periodic function. In other words, Fourier descriptors are used to extract contour features based on contour signatures extracted from binary images. This paper adopts a contour tracing algorithm proposed by Pavlidis<sup>21</sup> to find the contours of three principal projected images. Figure 6 shows an example of tracing the contours from three principal projected images.

Fourier descriptors can be derived from different shape signatures, such as centroid distance, complex coordinates, cumulative angles, and curvature function. Zhang and Lu<sup>29</sup> suggested that the shape signature from centroid distance has the higher retrieval performance than other methods to derive Fourier descriptors. In addition, the centroid distance signature keeps invariant in translation. An example of the contour signature represented by centroid distance is shown in Fig. 7.

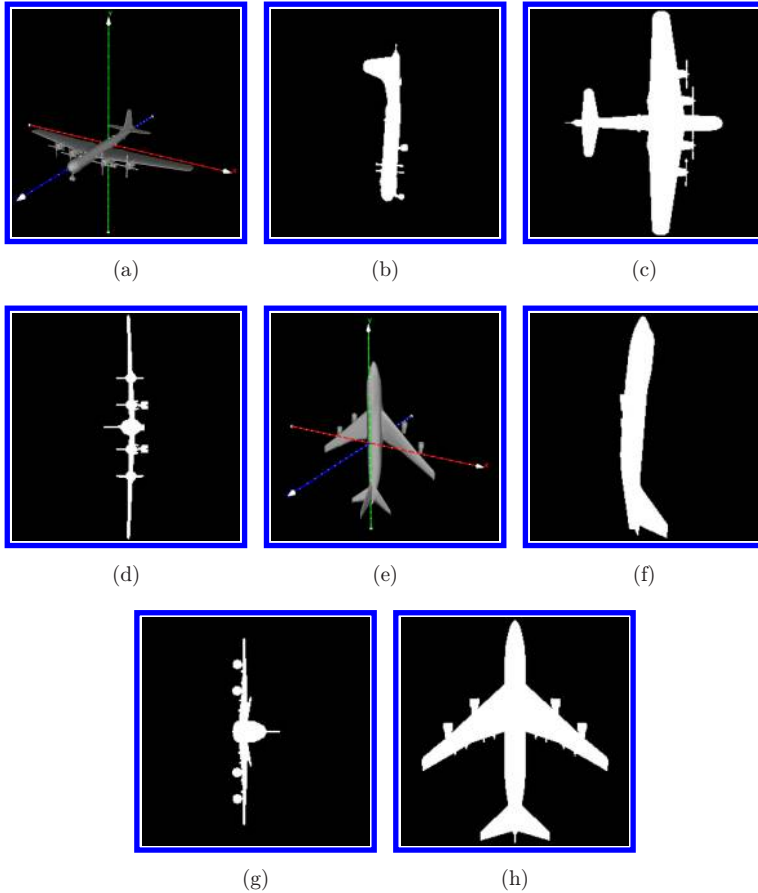


Fig. 5. Two 3D models of planes with different orientations. (a) The 3D model of a four-propeller plane. (b)–(d) are its first, second, and third projected images, respectively. (e) The 3D model of a jet plane. (f)–(h) are its first, second, and third projected images, respectively.

Figure 7(a) shows one of three projected image contours in Fig. 6, and Fig. 7(b) shows the centroid distance of Fig. 7(a).

The coefficients of Fourier descriptors can be expressed as follows. Assuming the distance function is normalized to  $N$  points in the sampling space, and the discrete Fourier transform is defined as follows:

$$F(n) = \frac{1}{N} \sum_{k=0}^{N-1} f(k) \exp\left(\frac{-j2\pi nk}{N}\right), \quad n = 0, 1, \dots, N - 1. \quad (6)$$

To achieve rotation invariance, phase information of the coefficients  $F(n)$  is ignored and only the magnitude  $|F(n)|$  is used. Scaling invariance can be achieved by dividing  $|F(1)|$  for complex-valued coefficients as Eq. (7), or by dividing  $|F(0)|$  for



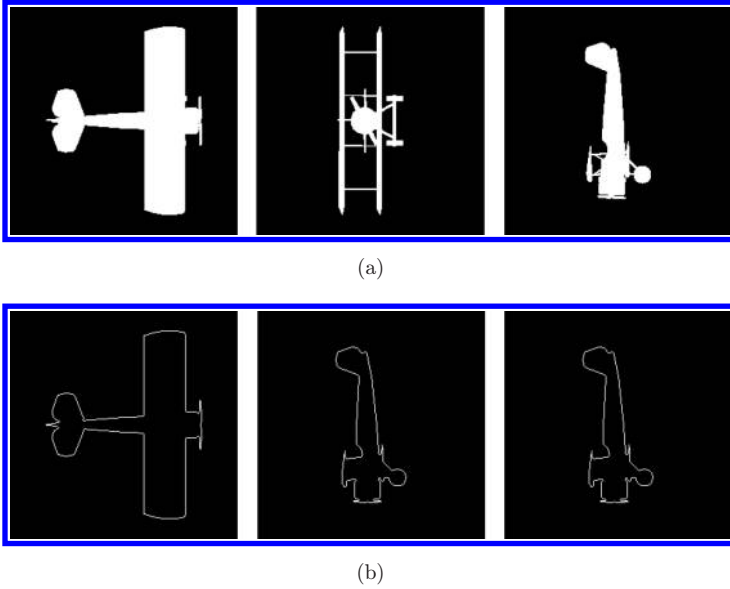


Fig. 6. An example of the projected images and their contours from three principal axes. (a) The projected images. (b) The contours of the projected images.

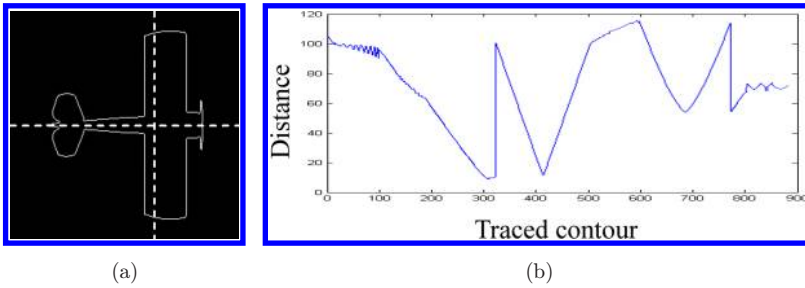


Fig. 7. An example of centroid distance signature. (a) The contour image is one in Fig. 6(b), and the crossed dotted lines mark the centroid of the contour. (b) The centroid distance of the contour in (a).

real-valued coefficients as Eq. (8), and the descriptors can be defined as follows:

$$F_{\text{complex}} = \left[ \frac{|F(2)|}{|F(1)|}, \frac{|F(3)|}{|F(1)|}, \dots, \frac{|F(N-1)|}{|F(1)|} \right]^T, \tag{7}$$

$$F_{\text{real}} = \left[ \frac{|F(1)|}{|F(0)|}, \frac{|F(2)|}{|F(0)|}, \dots, \frac{|F(N/2)|}{|F(0)|} \right]^T. \tag{8}$$

In Eq. (8), only half the coefficients  $F(n)$  are needed to characterize the shape when using real-valued coefficients. To take advantage, real-valued coefficients are used in the proposed method.

### 2.3. Zernike moments

Zernike moments, which are constructed using a set of complex polynomials that form a complete orthogonal basis set defined on the unit disc, are used as a region-based image retrieval tool. Zhang and Lu<sup>30</sup> suggested that it has better performance than Fourier descriptors in 2D image retrieval. Like Fourier descriptors, Zernike moments also have translation, scale and rotation invariance.

Complex Zernike moments are derived from Zernike polynomials, and they are expressed as a 2D Zernike moment:

$$Z_{mn} = \frac{m+1}{\pi} \int_x \int_y f(x, y) \cdot V_{mn}(x, y)^* dydx, \quad x^2 + y^2 \leq 1, \tag{9}$$

where  $(\cdot)^*$  denotes the complex conjugate. Equation (9) can be rewritten in polar coordinate form as:

$$Z_{mn} = \frac{m+1}{\pi} \int_0^{2\pi} \int_0^1 f(r, \theta) \cdot V_{mn}(r, \theta) drd\theta. \tag{10}$$

For easy computation, the set Zernike moments must follow polynomials:

$$V_{mn}(r, \theta) = R_{mn}(r) \exp(jn\theta), \tag{11}$$

and  $R_{mn}(r)$  is defined as:

$$R_{mn}(r) = \sum_{s=0}^{\frac{m-|n|}{2}} (-1)^s \frac{(m-s)!}{s! (\frac{m+|n|}{2} - s)! (\frac{m-|n|}{2} - s)!} r^{m-2s}, \tag{12}$$

where  $n$  and  $m$  are subject to  $|m| - n$  is even and  $|n| \leq m$ . Zernike polynomials are a complete set of complex-valued function orthogonal over the unit disc. For  $f(x, y)$  is real and discrete, the complex Zernike moments of order  $n$  with repetition  $m$  can be derived from modifying Eq. (9):

$$Z_{mn} = \frac{m+1}{\pi} \sum_x \sum_y f(x, y) [VR_{mn}(x, y) + jVI_{nm}(x, y)], \quad x^2 + y^2 \leq 1, \tag{13}$$

where  $VR_{mn}(x, y)$  is the real part of  $V_{mn}(r, \theta)$ , and  $VI_{nm}(x, y)$  is the imaginary part of  $V_{mn}(r, \theta)$ .

This paper chooses the magnitude of the Zernike moments to keep the rotation invariant. In addition, every component is divided by the magnitude of  $Z_{00}$ , but not including  $Z_{00}$  itself, to keep the scale invariant. Thus, the representation can be shown as follows:

$$Z = \left[ \frac{|Z_{11}|}{|Z_{00}|}, \frac{|Z_{20}|}{|Z_{00}|}, \frac{|Z_{22}|}{|Z_{00}|}, \dots, \frac{|Z_{\infty\infty}|}{|Z_{00}|} \right]^T. \tag{14}$$

### 3. The Proposed Retrieval System

This chapter describes the proposed 3D model retrieval system that contains an offline procedure for feature database creation and an online procedure for query retrieval. Figure 8 shows the two procedures of the proposed method. Three main parts, CPCA projection, feature integration, and similarity measure, will be explained in detail to show how this retrieval system works.

#### 3.1. Offline procedure

The offline procedure is to create the retrieval database. Two main parts in the procedure are principal axes projection and feature extraction, and they are described in the following two sections.

##### 3.1.1. CPCA projection

The purpose of CPCA projection is to align a 3D model with the referenced frame and obtaining the principal images. Two steps of a CPCA projection procedure for a 3D model are described as follows:

- (1) Using the CPCA to find principal axes of 3D models. It mainly utilizes Eq. (5) to align every 3D model data with their principal axes. Use CPCA to find the principal axes of the model, and then align the model with the referenced frame.
- (2) Project the 3D model geometry data along its three principal axes separately to get three feature images. The feature images are always binary images.

Ideally, if objects have similar shape, they would have similar feature images. However, from the previous section, only using the CPCA projection method would

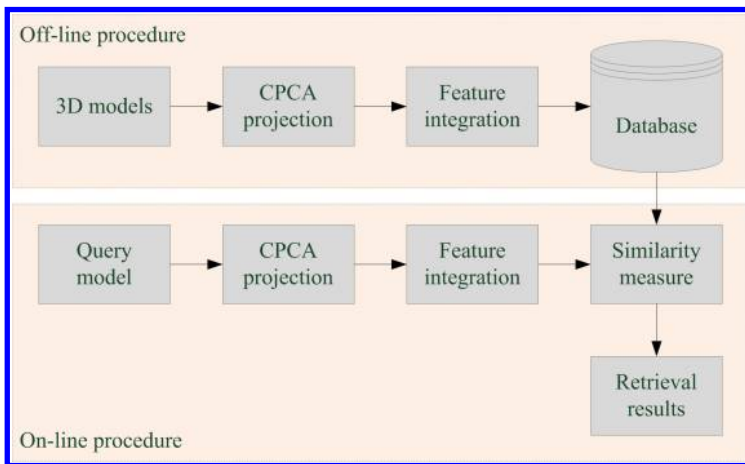


Fig. 8. The retrieval process of the proposed PAID method.

cause the image rotation problem and the switching problem. Thus, these are discussed in the following sections.

### 3.1.2. Feature integration

The purpose of the feature integration procedure is to find the shape features of the feature images of the 3D model. The shape features for the proposed method comprise Fourier descriptors and Zernike moments. The details are summarized as follows.

First, the steps of extracting Fourier descriptors for a 3D model are as follows:

- (1) Use a contour tracing approach proposed by Pavlidis<sup>21</sup> to find the contours of the feature images.
- (2) For each contour, calculate the distance from boundary points to the centroid of the contour to yield a distance function.
- (3) The discrete Fourier transform is applied to find the coefficients for Fourier descriptors, as described in Sec. 2.2.

After calculating the Fourier descriptors, the retrieval problem can be reduced from feature images to feature vectors. For Fourier descriptors of a 2D feature image, it only finds one feature vector  $F(k)$ ,  $k = 0, \dots, 10$  because the feature vector in the front and back sides are the same. Therefore, this paper only takes the front side of feature vectors and omits the back side of feature vectors. Hence, 11 ( $F(0)$  to  $F(10)$ ) coefficients represent a feature image.

The second is extracting Zernike moments. The precision of shape representation depends on the number of moments truncated from the expansion, and the first 25 moments up to order 8 are used in this paper. The number of moments in each order is displayed in Table 1. The Zernike moments descriptor has two feature vectors in a feature image. One is the front side feature  $Z_f(k)$  and the other is the back side feature  $Z_b(k)$ , where  $k = 0, \dots, 24$  for both features. Thus, 50 Zernike moment coefficients represent a feature image.

Both methods are suitable for describing shapes in specific cases. Shape boundaries can be described by Fourier descriptors, and shape inner content can be represented

Table 1. The number of Zernike moments in each order.

$m$	Moments	No. of Moments
0	$Z_{00}$	1
1	$Z_{11}$	1
2	$Z_{20}, Z_{22}$	2
3	$Z_{31}, Z_{33}$	2
4	$Z_{40}, Z_{42}, Z_{44}$	3
5	$Z_{51}, Z_{53}, Z_{55}$	3
6	$Z_{60}, Z_{62}, Z_{64}, Z_{66}$	4
7	$Z_{71}, Z_{73}, Z_{75}, Z_{77}$	4
8	$Z_{80}, Z_{82}, Z_{84}, Z_{86}, Z_{88}$	5

by Zernike moments. To take advantage of both methods, an integrated feature consisting of Fourier descriptor and Zernike moments is proposed. The integrated feature  $I$  is defined as:

$$I = \{w_F F(0), \dots, w_F F(10), w_Z Z_f(0), \dots, w_Z Z_f(24), w_Z Z_b(0), \dots, w_Z Z_b(24)\}, \tag{15}$$

where  $w_F$  and  $w_Z$  are weightings adjusted according to different databases. Higher  $w_F$  controls the retrieval method using more contour information, and higher  $w_Z$  means retrieval using content-based information. The default values of  $w_F$  and  $w_Z$  are 0.5.

In short, since each model has three feature images, the combined feature, called PAID, of one 3D model contains 183 coefficients, 33 coefficients for Fourier descriptors and 150 coefficients for Zernike moments. In our implementation, each coefficient is quantized to an 8-bit value so that reduce the storage size and accelerate the similarity calculation.

### 3.2. Online procedure

The online procedure is to provide a quick retrieval for a query model, which compares the PAID of the queried one with all other 3D models in the database. First, the PAID of a query model are brought by applying CPCA projection and feature extraction mentioned in the offline procedure. Next, the similarity measurement is achieved by comparing the PAID within two models. Finally, similar 3D models, which have higher similarities to the query model, are outputs as the result of the retrieval.

To deal with the axes switching problem mentioned in Sec. 2.1, a strategy is employed in the proposed method. This strategy is based on a simple assumption: the distance between the feature pair of the correct axis correspondence will be shortest. For example, suppose  $(P_1, P_2, P_3)$  and  $(Q_1, Q_2, Q_3)$  are the PAID features of two similar models. The subscript denotes the axis number. If there is no switching of axes, the comparison of two models is to calculate the distance between every feature pair, that is,  $(P_1, Q_1)$ ,  $(P_2, Q_2)$ , and  $(P_3, Q_3)$ . However, in the case of incorrect axis correspondence, the feature pair is composed of two features as  $(P_1, Q_2)$  or  $(P_3, Q_2)$ . This will lead to the undesirable result of comparing two models. Thus, the proposed similarity measurement finds the minimum distance between two features to determine the correct axis correspondence. The algorithm is illustrated as follows.

The distance between two models marked  $q$  and  $p$ , with PAID denoted as  $I^{(q)}$  and  $I^{(p)}$ , is calculated as Eq. (16), and  $l(\cdot, \cdot)$  is L1 distance of two features. The subscript of  $I$  means the axis number.

$$d_{\text{PAID}}(q, p) = \min(l(I_1^{(q)}, I_a^{(p)}) + l(I_2^{(q)}, I_b^{(p)}) + l(I_3^{(q)}, I_c^{(p)})), \tag{16}$$

$$\text{for } \begin{cases} a = 1, 2, 3, \\ b = 1, 2, 3 \text{ and } b \neq a, \\ c = 1, 2, 3, \quad c \neq a \text{ and } c \neq b. \end{cases}$$

According to this calculation, the result of the correspondences of the two models is consistent with human perception. Finally, to obtain an intuitive similarity score, let the distance be normalized by the longest distance of retrieval for a query model  $q$  as shown in the following:

$$s(q, m) = 1 - \frac{d_{\text{PAID}}(q, m)}{\max_r d_{\text{PAID}}(q, r)}, \quad (17)$$

where  $m$  and  $r$  both denote a model in the database. The larger similarity score represents more similarity between two objects. A score of 1 represents two identical objects.

## 4. Experimental Results

The results presented in this section are divided into three parts. We will use PSB<sup>22</sup> database and a few smaller datasets to demonstrate the performance and details of the proposed algorithms. The PSB database and several standard measures are introduced at first. Second, the performance of 3D model retrieval for the proposed method and other approaches are compared in detail.

### 4.1. Test model database

To compare different retrieval approaches, it is essential to use comprehensive databases that contain models classified in different categories. PSB is the well-known benchmark for 3D model retrieval. It contains 1814 objects in general categories like animal, building, vehicle, and so on, and is divided into training and testing two sets. The training set that has 907 objects classified in 90 categories is used to create the ground-truth of the retrieval algorithm. The other that includes 907 different objects in 92 classes is used for computing the retrieval performances.

The experiments are conducted on the PSB database. First, the ground-truth database is built by using the training set. Next, each classified object in the testing set is used as a query object in the experiments. The object belonging to the same class in the ground-truth database is considered the relevant object.

### 4.2. Evaluation methods

The performance of model retrieval is evaluated by computing quantitative statistics of match results. According to the characteristics of different methods, measurement results will show retrieval effectiveness under certain conditions. Five commonly used evaluation methods discussed in Ref. 22 are used to measure the quality of model retrieval, and listed as follows.

- Nearest neighbor (NN)
- First-tier
- Second-tier

- E-measure
- Discounted cumulative gain (DCG)

All scores of these evaluation methods are in range [0, 1]. Higher scores represent better results. It is important that DCG takes the position of relevant models into account. Therefore, DCG appears to be an effective and stable measure for implementing evaluations.

### 4.3. Retrieval performances and discussions

Four other competing methods are selected to verify the performance of our PAID method, namely LFD,<sup>6</sup> spin-image signatures (SIS),<sup>3</sup> shape histogram (SECTORS),<sup>2</sup> and D2 shape distribution.<sup>18</sup> These methods are briefly introduced here.

LFD is a descriptor of a model as a collection of images rendered from the sampled positions on the vertices of a dodecahedron. To recognize a 3D model from 2D images, LFD considers all rotations on all vertices of a dodecahedron. That is, 100 images should be computed to represent a model. The dimension of an image is  $256 \times 256$  pixels in the implementation. SIS transforms spin-images of a 3D model into six signature clusters. Each center of the cluster is represented by an 18-dimensional signature. SECTORS is a spherical descriptor, which is computed on a  $64 \times 64$  spherical grid and then represented by its harmonic coefficients up to order 16. D2 is a histogram of distances between pairs of points on the surface, which is represented with 64 bins. In the experiments, the parameters in our method are summarized as follows:

- The type of the principal axis image is a binary image.
- The dimension of 2D images is  $256 \times 256$ .
- The number of coefficients of Fourier descriptors for each image is 11.
- The number of coefficients of Zernike moments for each image is 50.

Table 2 illustrates the overall performances of all five methods including our PAID method on the PSB testing dataset. The PAID and LFD methods achieve the best results. Except for NN, our method shows the best results among all methods and the same performance in DCG as LFD. The SIS method also carries out good

Table 2. The overall performances of model retrieval from the PSB database.

Method	NN (%)	First-Tier (%)	Second-Tier (%)	E-Measure (Factor)	DCG (Factor)	Storage Size (Bytes)	Average Timing	
							Generation (Sec.)	Comparison (msec.)
PAID	64.8	38.7	49.4	0.287	0.643	183	1.13	0.004
LFD	65.7	38.0	48.7	0.280	0.643	4700	4.32	1.280
SIS	63.2	33.7	45.1	0.261	0.624	432	28.92	0.011
SECTORS	54.6	26.7	35.0	0.209	0.545	552	0.95	0.013
D2	31.1	15.8	23.5	0.139	0.434	136	1.20	0.002

performance which is close to LFD. However, PAID method is more efficient than LFD in both feature generation and model comparison because only three images are applied. The last three columns in Table 2 summarize storage size, generation time and comparison time. The comparison time is between two models. Algorithms are executed on a PC with Pentium 4 2.4 GHz CPU and 2 GB RAM. Compared with LFD, the PAID method has several benefits in terms of computation costs and storage size. Our method spends nearly the same time as that of SECTORS and D2 to generate model features, and needs only 1/25 of storage size in LFD and 1/3 in SECTORS. Overall, the proposed PAID method is a significantly better technique.

Figure 9 depicts the average generation time over all models in PSB spent by the five methods. Generally, except for the LFD method, other four methods need less than 1.5 seconds from low-resolution and medium-resolution models (less than 4096 vertices). The LFD method takes around eight times longer in this case. For high-resolution models (more than 16,384 vertices), it can be observed that the time costs of the PAID, SECTOR and D2 are similar, and the LFD requires more than twice as much time as these methods. In addition, Fig. 9(b) shows the generation time comparison in log space. It is clear that the SIS method performs as good as the SECTOR method at low resolutions. However, the generation time at high resolutions increases dramatically, because the computation of spin-images is very heavy. This limitation causes the SIS method is only suitable for low-resolution and medium-resolution models.

To show more detailed retrieval comparisons, four commonly used datasets, which are animal, furniture, common tool, and vehicle in PSB, are chosen. Each dataset includes several classes. The retrieval performances are measured by precision recall curve and DCG histogram. Moreover, the performance is compared with four other competing methods.

Tables 3–6 show the retrieval statistics using several evaluation measures for animal, furniture, common tool, and vehicle dataset, respectively. The LFD method

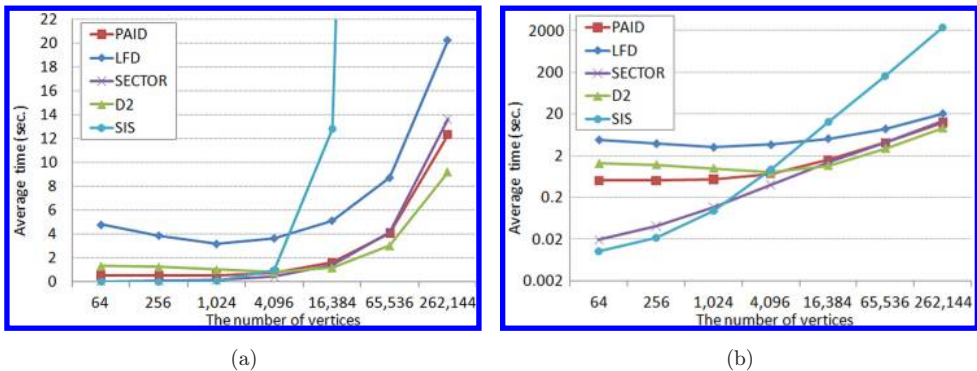


Fig. 9. The comparison of average generation time in different numbers of vertices. The time scale is represented in (a) linear and (b) log space.



Table 3. Comparing four algorithms for the test with animal dataset.

Method	NN (%)	First-Tier (%)	Second-Tier (%)	E-Measure (Factor)	DCG (Factor)
PAID	66.7	51.0	67.6	0.287	0.736
LFD	68.3	45.7	62.3	0.272	0.720
SIS	69.6	46.1	64.3	0.274	0.726
SECTORS	57.1	40.5	51.6	0.249	0.649
D2	34.9	28.0	48.9	0.266	0.566

Table 4. Comparing four algorithms for the test with furniture dataset.

Method	NN (%)	First-Tier (%)	Second-Tier (%)	E-Measure (Factor)	DCG (Factor)
PAID	77.1	62.5	75.3	0.350	0.815
LFD	80.0	52.5	69.1	0.347	0.779
SIS	82.4	64.8	76.9	0.352	0.819
SECTORS	65.7	39.7	54.2	0.289	0.685
D2	48.6	33.1	49.4	0.293	0.611

Table 5. Comparing four algorithms for the test with tool dataset.

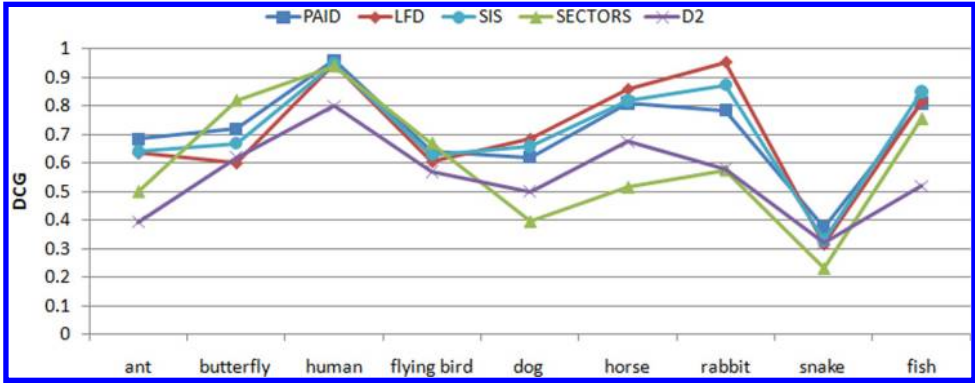
Method	NN (%)	First-Tier (%)	Second-Tier (%)	E-Measure (Factor)	DCG (Factor)
PAID	71.0	55.7	73.8	0.283	0.765
LFD	65.6	49.1	67.4	0.278	0.730
SIS	64.7	50.2	64.1	0.272	0.738
SECTORS	53.8	46.1	62.2	0.252	0.684
D2	43.0	33.1	48.7	0.23	0.577

Table 6. Comparing four algorithms for the test with vehicle dataset.

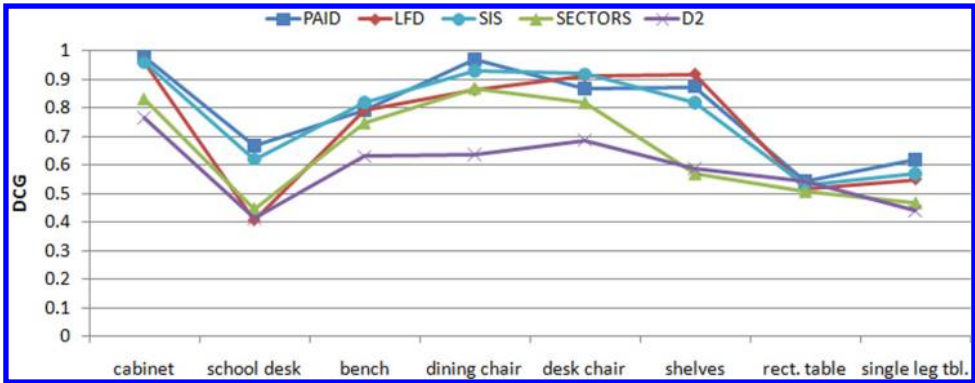
Method	NN (%)	First-Tier (%)	Second-Tier (%)	E-Measure (Factor)	DCG (Factor)
PAID	83.6	66.5	81.7	0.327	0.858
LFD	85.1	63.9	84.9	0.335	0.852
SIS	83.9	64.3	84.2	0.333	0.859
SECTORS	67.2	44.6	60.1	0.292	0.728
D2	47.8	38.8	60.8	0.285	0.662

provides better retrieval precision on NN evaluation. In this experiment, only the LFD method provides better retrieval precision on NN evaluation. This outcome is expected in that LFD uses ten images to represent 20 viewpoints. In contrast, the proposed PAID method yields better retrieval precision on all other evaluations against other methods. The primary factor for this performance is that the projected images of principal axes can tolerate the small shape difference.

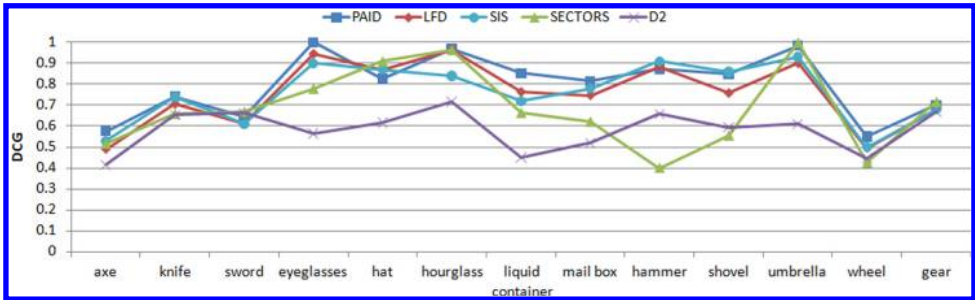
Figure 10 shows the averaged DCG scores of each classification for all datasets in the experiment. The PAID method performs better retrieval precision than the LFD and SIS methods in about 70% of classifications. Nevertheless, a small part of the results compares poorly with the results from other methods. Retrieval performance



(a)

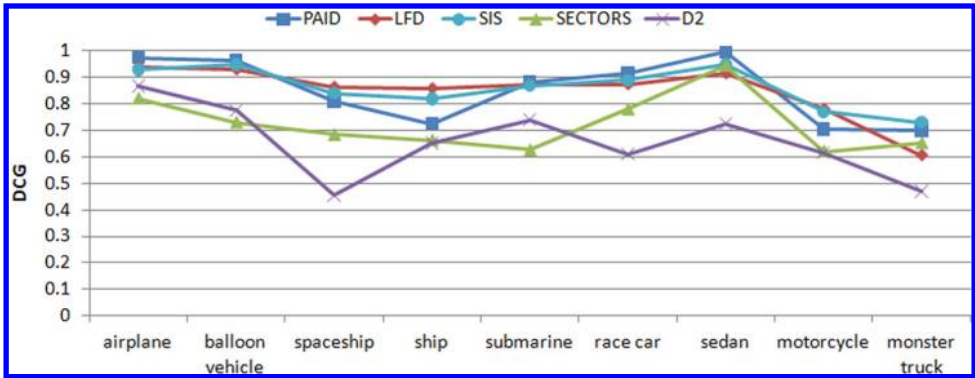


(b)



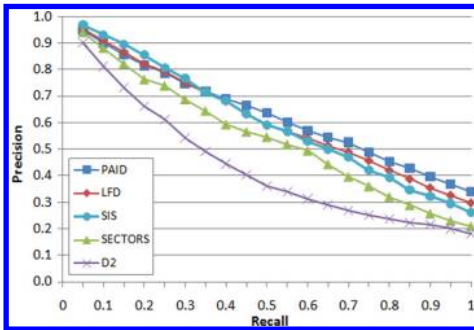
(c)

Fig. 10. The histograms of DCG for retrieval performance on four test datasets. (a) Animal, (b) furniture, (c) common tool, and (d) transportation dataset.

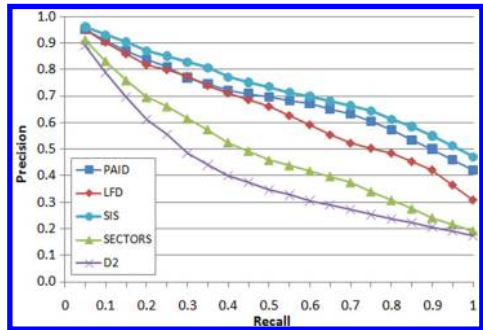


(d)

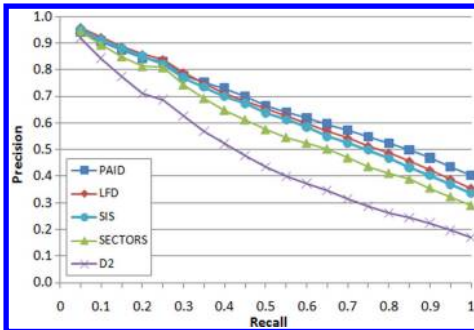
Fig. 10. (Continued)



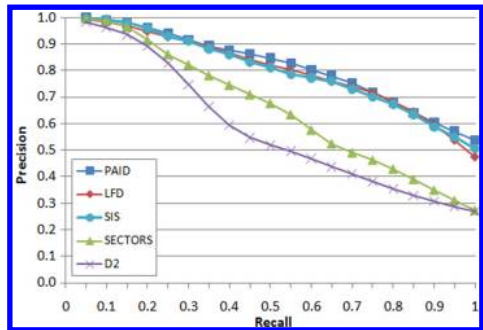
(a)



(b)



(c)



(d)

Fig. 11. Precision-recall curves on four test datasets. (a) Animal, (b) furniture, (c) common tool, and (d) transportation dataset.

will have different degrees of accuracy in different classifications. The major causes of this phenomenon are as follows.

First, the objects in the same class usually have a number of different poses and many variations in shape, which increases the difficulty in obtaining an accurate

retrieval result. For example, the snake has variable shape in the class, and therefore does not have a uniform shape. As Fig. 10(a) shows, the performances for all compared methods in snake model retrieval are not good.

Second, the principal axes of objects do not align well in some classifications because some objects have different proportions of the distribution. For example, the shape of a dining chair usually leads to oblique principal axes. However, this effect can be reduced by using the descriptor with rotation invariance. The result shown in Fig. 10(b) suggests that the proposed method can overcome the problem of oblique principal axes.

Figure 11 shows the precision-recall plot showing the averaged retrieval results achieved for each dataset. From the results, the proposed PAID method indicates very good precision-recall followed closely by the LFD and SIS approaches. Furthermore, the PAID method gets better precision than LFD as in higher recall rate. It shows that the PAID method is suitable for multiple object retrieval. The performances of SECTORS and D2 fall behind our method. In summary, the proposed PAID method has good performance compared with other methods both in terms of object retrieval in the entire testing dataset or in a specific class.

## 5. Conclusion and Future Works

In this paper, an approach called PAID based on contours and image features is proposed for 3D model retrieval. The approach uses Fourier descriptors and Zernike moments extracted from feature images to match among 3D models. Furthermore, the feature images are generated by projecting models along their principal axes found by CPCA. These methods lead to the proposed approach, which is robust against translation, rotation, and scaling for 3D model retrieval. Using only three feature images for each 3D model decreases the dimension of features, and increases the efficiency of model retrieval. In summary, the experimental results show that our approach outperforms SECTORS and D2, and has slightly better retrieval results than LFD and SIS. Moreover, the approach is more efficient than LFD and SIS, and the storage size is much less.

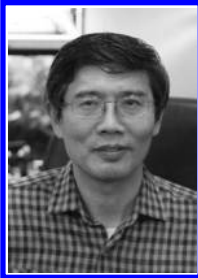
Among many topics to be explored in future work, important ones are listed as follows. First, to reduce the effect of different poses, a better pose evaluation should be considered. Second, some local features are needed to investigate the increasing ability to distinguish details. In addition, colors and textures can be included to add more visual features to improve overall accuracy. Finally, partial matching that can be used in many applications, like object detection and environment recognition, is an important and difficult problem in future research.

## References

1. M. Ankerst, G. Kastenmüller, H.-P. Kriegel and T. Seidl, Nearest neighbor classification in 3D protein databases, *Proc. ISMB* (1999), pp. 34–43.

2. M. Ankerst, G. Kastenmüller, H.-P. Kriegel and T. Seidl, 3D shape histogram for similarity search and classification in spatial databases, *Proc. 6th Int. Sympo. Spatial Databases* (1999), pp. 207–226.
3. J. Assfalg, M. Bertini, A. D. Bimbo and P. Pala, Content-based retrieval of 3-D objects using spin image signatures, *IEEE Trans. Multimed.* **9**(3) (2007) 589–599.
4. D. Bespalov, W. C. Regli and A. Shokoufandeh, Reeb graph based shape retrieval for CAD, *Proc. DETC 03* (2003) Vol. 3.
5. B. Bustos, D. A. Keim, D. Saupe, T. Schreck and D. V. Vranić, Feature-based similarity search in 3D object databases, *ACM Comput. Surv.* **37**(4) (2005) 345–387.
6. D. Y. Chen, M. Ouhyoung, X. P. Tian and Y. T. Shen, On visual similarity based 3D model retrieval, *Comput. Graph. Forum* **22**(3) (2003) 223–232.
7. T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin and D. Jacobs, A search engine for 3D models, *ACM Trans. Graph.* **22**(1) (2003) 83–105.
8. M. Hilaga, Y. Shinagawa, T. Kohmura and T. L. Kunii, Topology matching for fully automatic similarity estimation of 3D shapes, *Proc. 28th Ann. Conf. Int. Conf. Computer Graphics and Interactive Techniques* (2001), pp. 203–212.
9. C. Y. Ip, D. Lapadat, L. Sieger and W. C. Regli, Using shape distributions to compare solid models, *Proc. 7th ACM Symp. Solid Modeling and Application*, Saarbrücken, Germany (2002), pp. 273–280.
10. A. C. Jalba and J. B. Roerdink, Efficient surface reconstruction from noisy data using regularized membrane potentials, *IEEE Trans. Imag. Process.* **18**(5) (2009) 1119–1134.
11. A. Johnson and M. Herbert, Using spin images for efficient object recognition in cluttered 3D scenes, *IEEE Trans. Patt. Anal.* (1999), pp. 433–449.
12. I. Kolonias, D. Tzovaras, S. Malassiotis and M. G. Strintzis, Fast content-based search of VRML models based on shape descriptions, *IEEE Trans. Multimed.* **7**(1) (2005) 114–126.
13. S. Lee, P. S. P. Wang, S. N. Yanushkevich and S. Lee, Noniterative 3D face reconstruction based on photometric stereo, *Int. J. Patt. Recogn. Artif. Intell.* **22**(3) (2008) 389–410.
14. M. Mandot and K. Venuqopalan, A survey of surface retrieval techniques for 3D models, *First Int. Conf. Emerging Trends in Engineering and Technology*, 16–18 July 2008, pp. 48–51.
15. P. Min, A comparison of text and shape matching for retrieval of online 3D models, with statistical significance testing, *Proc. European Conf. Digital Libraries* (2004), pp. 209–220.
16. R. Ohbuchi, T. Minamitani and T. Takei, Shape-similarity search of 3D models by using enhanced shape functions, *Int. J. Comput. Appl. Technol.* **23**(2) (2005) 70–85.
17. R. Osada, T. Funkhouser, B. Chazelle and D. Dobkin, Matching 3D models with shape distributions, *Shape Modeling International* (2001), pp. 154–166.
18. R. Osada, T. Funkhouser, B. Chazelle and D. Dobkin, Shape distributions, *ACM Trans. Graphics* **21**(4) (2002) 807–832.
19. E. Paquet, M. Rioux, A. Murching, T. Naveen and A. Tabatabai, Description of shape information for 2-D and 3-D objects, *Sign. Process.: Imag. Commun.* **16**(1–2) (2000) 103–122.
20. Y. Park, Y. Yun and J. Choi, A new shape descriptor using sliced image histogram for 3D model retrieval, *IEEE Trans. Consum. Electron.* **55**(1) (2009) 240–247.
21. T. Pavlidis, *Algorithms for Graphic and Image Processing* (Computer Science Press, 1982).
22. P. Shilane, P. Min, M. Kazhdan and T. Funkhouser, The Princeton shape benchmark, *Proc. 6th Int. Conf. Shape Modeling and Applications* (2004), pp. 167–178.
23. H. Sundar, D. Silver, N. Gagvani and S. J. Dickinson, Skeleton based shape matching and retrieval, *Proc. Shape Modelling Int.* (2003), pp. 130–139.

24. J. H. Tangelder and R. C. Veltkamp, A survey of content based 3D shape retrieval methods, *Multimed. Tools and Appl.* **39**(3) (2008) 441–471.
25. D. V. Vranić and D. Saupe, 3D model retrieval, *Proc. Spring Conf. Computer Graphics and its Applications* (Budmerice, Slovakia, May 2000), pp. 89–93.
26. D. V. Vranić, D. Saupe and J. Richter, Tools for 3D-object retrieval: Karhunen–Loeve transform and spherical harmonics, *Proc. IEEE 4th Workshop on Multimedia Signal Processing* (Cannes, France, Oct. 2001), pp. 293–298.
27. D. V. Vranić and D. Saupe, Description of 3D-shape using a complex function on the sphere, *Proc. IEEE Int. Conf. Multimedia and Expo* (2002), pp. 177–180.
28. Y. Yang, H. Lin and Y. Zhang, Content-based 3-D model retrieval: A survey, *IEEE Trans. Syst., Man Cybern.* **37**(6) (2007) 1081–1098.
29. D. S. Zhang and G. Lu, A comparative study of Fourier descriptor for shape representation and retrieval, *Proc. 5th Asian Conf. Computer Vision* (Melbourne, Australia, 22–25 Jan. 2002), pp. 646–651.
30. D. Zhang and G. Lu, An integrated approach to shape based image retrieval, *Proc. 5th Asian Conf. Computer Vision* (Melbourne, Australia, 22–25 Jan. 2002), pp. 652–657.



**Sheng-Fuu Lin** received the B.S. and M.S. degrees in mathematics from National Taiwan Normal University in 1976 and 1979, respectively, the M.S. degree in computer science from the University of Maryland, College Park, in 1985, and the Ph.D. degree in electrical engineering from the University of Illinois, Champaign, in 1988. Since 1988, he has been on the faculty of the Department of Electrical Engineering at National Chiao Tung University, Hsinchu, Taiwan, where he is currently a professor.

His research interests include image processing, image recognition, fuzzy theory, automatic target recognition, and scheduling.



**Chi-Yao Hsu** received his B.S. degree from the Department of Electrical Engineering from National Taiwan Ocean University, Taiwan, in 2001 and his M.S. degree from the Department of Electrical Engineering from National Central University, Taiwan, in 2003. He is currently pursuing the Ph.D. degree in the Department of Electrical Engineering at National Chiao Tung University, Hsinchu, Taiwan.

His research interests lie in the areas of neural networks, fuzzy systems, evolutionary algorithms, pattern recognition, and computer vision.



**Chin-Chia Wu** received his B.S. and M.S. degrees in mechanical engineering from National Taiwan Ocean University in 2002 and 2004, respectively. He is currently pursuing the Ph.D. degree in the Department of Electrical Engineering at National Chiao Tung University,

Hsinchu, Taiwan.

His current research interests include image processing, pattern recognition, 3D model retrieval, 3D surface matching and 3D recognition.



**Dou-Chih Hsu** received his B.S. degree from the Department of Power Mechanical Engineering at National Tsing Hua University in 2007, and the M.S. degree from the Institute of Electrical Control Engineering at National Chiao Tung University in 2009.

His research interests include pattern recognition and 3D model retrieval.