

# PANORAMA: A 3D Shape Descriptor Based on Panoramic Views for Unsupervised 3D Object Retrieval

Panagiotis Papadakis · Ioannis Pratikakis ·  
Theoharis Theoharis · Stavros Perantonis

Received: 17 October 2008 / Accepted: 22 July 2009  
© Springer Science+Business Media, LLC 2009

**Abstract** We present a novel 3D shape descriptor that uses a set of panoramic views of a 3D object which describe the position and orientation of the object's surface in 3D space. We obtain a panoramic view of a 3D object by projecting it to the lateral surface of a cylinder parallel to one of its three principal axes and centered at the centroid of the object. The object is projected to three perpendicular cylinders, each one aligned with one of its principal axes in order to capture the global shape of the object. For each projection we compute the corresponding 2D Discrete Fourier Transform as well as 2D Discrete Wavelet Transform. We further increase the retrieval performance by employing a local (unsupervised) relevance feedback technique that shifts the descriptor of an object closer to its cluster centroid in feature space. The effectiveness of the proposed 3D object retrieval methodology is demonstrated via an extensive consistent evaluation in standard benchmarks that clearly shows better performance against state-of-the-art 3D object retrieval methods.

**Keywords** 3D object retrieval · Cylindrical projection · Panorama · Local relevance feedback · Unsupervised

---

P. Papadakis (✉) · I. Pratikakis · S. Perantonis  
Computational Intelligence Laboratory, Institute of Informatics  
and Telecommunications, National Center for Scientific Research  
"Demokritos", Athens, Greece  
e-mail: [ppapadak@iit.demokritos.gr](mailto:ppapadak@iit.demokritos.gr)

I. Pratikakis  
e-mail: [ipratika@iit.demokritos.gr](mailto:ipratika@iit.demokritos.gr)

S. Perantonis  
e-mail: [sper@iit.demokritos.gr](mailto:sper@iit.demokritos.gr)

T. Theoharis  
Computer Graphics Laboratory, Department of Informatics  
and Telecommunications, University of Athens, Athens, Greece  
e-mail: [theotheo@di.uoa.gr](mailto:theotheo@di.uoa.gr)

## 1 Introduction

Content-based 3D object retrieval is an active research field that has attracted a significant amount of interest in recent years. This is due to the increasing availability of 3D objects through public or proprietary databases, as even novice users can create 3D models from scratch with user-friendly modeling interfaces such as Teddy (Igarashi et al. 1999), Google SketchUp (Google SketchUp 2009), ShapeShop (Schmidt et al. 2005) and others (Olsen et al. 1999). Applications such as CAD, computer games development, archaeology, bioinformatics, etc. are also playing a major role in the proliferation of 3D objects. A second source of 3D objects is 3D scanners, which are constantly gaining users as their price drops.

Already, 3D object search engines have been developed for commercial or research purposes that offer searching using 3D object queries or keyword-based queries. The former approach, which is an instance of content-based retrieval (CBR) alleviates various limitations that are encountered in keyword-based retrieval which become prohibitive as the number of 3D objects increases.

In CBR, each 3D object is represented by a *shape-descriptor* which is used to measure the similarity between two objects. The shape descriptor should capture the discriminative features of a 3D model, have a compact size and permit fast extraction and comparison time.

In this paper, we propose a novel 3D shape descriptor that exhibits top performance by using a set of features which are extracted from a set of panoramic views of a 3D object. The proposed descriptor is called **PANORAMA** which stands for PANoramic Object Representation for Accurate Model Attributing. The panoramic views are used to capture the position of the model's surface in 3D space as well as its orientation. We obtain a panoramic view of a 3D object

by projecting it to the lateral surface of a cylinder aligned with one of the object's three principal axes and centered at the centroid of the object. The object's principal axes are determined during the rotation normalization step. The object is projected onto three perpendicular cylinders, each one aligned with one of its principal axes in order to capture the global shape of the object. For each projection, we compute the corresponding 2D Discrete Fourier Transform as well as 2D Discrete Wavelet Transform. To further enhance the retrieval performance, we employ a local relevance feedback (LRF) technique (Papadakis et al. 2008b) that shifts the descriptor of an object closer to its cluster centroid in feature space. The employed LRF technique is unsupervised and assumes that the  $k$  nearest neighbors of an object belong to the same class, without requiring any feedback from the user. We demonstrate the increased performance of the PANORAMA descriptor by comparing it to other state-of-the-art approaches in standard benchmarks and promote the application of the LRF technique by showing the performance gain that is obtained when it is combined with the PANORAMA descriptor.

The remainder of the paper is organized as follows: In Sect. 2, we provide an overview of the related work in the area of content-based 3D object retrieval (Sect. 2.1) and discuss previous work that incorporates relevance feedback (Section 2.2). In Sect. 3, we give the detailed description for the extraction of the proposed 3D shape descriptor and in Sect. 4, we describe the procedure of the employed LRF technique. The results of an extensive consistent evaluation of the proposed methodology are presented in Sect. 5 and conclusions are drawn in Sect. 6.

## 2 Related Work

### 2.1 Shape Descriptors

In this section, we provide an overview of the related work in the area of 3D shape descriptors for generic 3D object retrieval. 3D object retrieval methodologies that rely on supervision are beyond the scope of this review, since this paper focuses mainly on enhancing the effectiveness of retrieval by using discriminative shape features in an unsupervised context.

Content-based 3D object retrieval methods may be classified into two main categories according to the spatial dimensionality of the information used, namely 2D, 3D and their combination. In the following sections, we review the state-of-the-art for each category.

#### 2.1.1 Methods Based on 2D Representations

In this category, shape descriptors are generated from images-projections which may be contours, silhouettes,

depth buffers or other kinds of 2D representations. Thus, similarity is measured using 2D shape matching techniques. Surprisingly, extended state-of-the-art reviews such as Shilane et al. (2004) and Bustos et al. (2005) show that descriptors belonging to this class exhibit better overall retrieval performance compared to descriptors that belong to the second class.

Chen et al. (2003) proposed the Light Field descriptor, which is comprised of Zernike moments and Fourier coefficients computed on a set of projections taken from the vertices of a dodecahedron. Vranic (2004) proposed a shape descriptor where features are extracted from depth buffers produced by six projections of the object, one for each side of a cube which encloses the object. In the same work, the Silhouette-based descriptor is proposed which uses the silhouettes produced by the three projections taken from the Cartesian planes. In Passalis et al. (2006), proposed PTK, a depth buffer based descriptor which uses parallel projections to capture the object's thickness and an alignment scheme that is based on symmetry. Shih et al. (2007) proposed the elevation descriptor where six depth buffers (elevations) are computed from the faces of the 3D object's bounding box and each buffer is described by a set of concentric circular areas that give the sum of pixel values within the corresponding areas. Ohbuchi et al. (2003) proposed the Multiple Orientation Depth Fourier Transform (MODFT) descriptor where the model is projected from 42 viewpoints to cover all possible view aspects. Each depth buffer is then transformed to the  $r - \theta$  domain and the Fourier transform is applied. To compare two objects, all possible pairs of coefficients are compared which inevitably increases comparison time. Zarpalas et al. (2007) introduced a 3D shape descriptor called the spherical trace transform, which is the generalization of the 2D trace transform. In this method, a variety of 2D features are computed for a set of planes intersecting the volume of a 3D model. A newly proposed method is the depth line descriptor proposed by Chaouch and Verroust-Blondet (2007) where a 3D object is projected to the faces of its bounding box giving 6 depth buffers. Each depth buffer is then decomposed into a set of horizontal and vertical depth lines that are converted to state sequences which describe the change in depth at neighboring pixels.

#### 2.1.2 Methods Based on 3D Representations

In this category, shape descriptors are extracted from 3D shape representations and the similarity is measured using appropriate representations in the spatial domain or in the spectral domain. A set of subcategories can be identified here, namely, statistical, graph-based and spherical function-based descriptors.

Statistical descriptors use histograms to capture the distributions of shape features. They are compact and fast to

compute but they have very limited discrimination ability since they fail to capture local details that are characteristic of the object's shape. In the Shape Histograms descriptor proposed by Ankerst et al. (1999), 3D space is divided into concentric shells, sectors, or both and for each part, the model's shape distribution is computed giving a sum of histograms bins. The shape distributions descriptor proposed by Osada et al. (2001, 2002) measures a set of shape characteristics for a random set of points belonging to the object, using appropriate shape functions, e.g. the D2 function which measures the distance between two random surface points. Ohbuchi et al. (2005) proposed enhanced shape functions, namely the (absolute) angle distance histogram for inconsistently oriented meshes, which are extensions of the D2 shape distribution. Zaharia and Petreux (2001) presented the 3D shape spectrum descriptor which is the histogram that describes the angular representation of the first and second principal curvature along the surface of the 3D object. In Sundar et al. (2003b) make use of 3D shape contexts which are histograms each one corresponding to a surface point and capturing the distribution of the relative coordinates of the remaining surface points.

Graph-based methods use hierarchical structures to represent 3D objects and the similarity is measured using graph-matching techniques. These methods are suited to intra-class search, i.e. searching within very similar objects at different poses (articulations) but they have limited discrimination ability in generic object retrieval. Hilaga et al. (2001) introduced the multi-resolution Reeb graph, which represents a 3D object's topology and skeletal structure at various levels of detail. In Zhang et al. (2005) consider the use of medial surfaces to compute an equivalent directed acyclic graph of an object. In the work of Sundar et al. (2003a), the 3D object passes through a thinning process producing a set of skeletal points, which finally form a directed acyclic graph by applying the minimum spanning tree algorithm. Cornea et al. (2005) propose the use of curve skeletons produced by the application of the generalized distance field to the volume of the 3D object and similarity is measured using the earth mover's distance. The P3DS descriptor developed by Kim et al. (2004) uses an attributed relational graph whose nodes correspond to parts of the object that are represented using ellipsoids and the similarity is computed by employing the earth mover's distance.

A plurality of methods exists that use spherical functions to parameterize the shape of a 3D object. These methods exhibit good discrimination ability in general but most of them cannot capture shape features uniformly. This happens when the longitude-latitude parameterization is adopted that results in non-uniform sampling between the poles of the spherical function. Vranic (2004) proposed the Ray-based descriptor which characterizes a 3D object by a spherical extent function capturing the furthest intersection points of

the model's surface with rays emanating from the origin. Spherical harmonics or moments can be used to represent the spherical extent function. A generalization of the previous approach (Vranic 2004) uses several spherical extent functions of different radii. The GEDT descriptor proposed by Kazhdan et al. (2003) is a volumetric representation of the Gaussian Euclidean Distance Transform of a 3D object, expressed by norms of spherical harmonic frequencies. In Papadakis et al. (2007), the CRSP descriptor was proposed which uses the Continuous PCA (CPCA) along with Normals PCA (NPCA) to alleviate the rotation invariance problem and describes a 3D object using a volumetric spherical-function based representation expressed by spherical harmonics. Yu et al. (2003) used spherical functions to describe the topology and concavity of the surface of a 3D object and the amount of effort required to transform it to its bounding sphere. Generalizing from 2D to 3D, Novotni and Klein (2003) presented the 3D Zernike descriptor, Daras et al. (2006) introduced the generalized radon transform, Ricard et al. (2005) developed the 3D ART descriptor by generalizing the 2D angular radial transform and Zaharia and Preteux (2002) proposed the C3DHTD descriptor by generalizing the 2D Hough Transform.

### 2.1.3 Hybrid Methods

Besides the previous categories, combinations of different methods have been considered in order to enhance the overall performance, which comprise a third category.

Vranic (2005) developed a hybrid descriptor called DESIRE, that consists of the Silhouette, Ray and Depth buffer based descriptors, which are combined linearly by fixed weights. The approach of Bustos et al. (2004) assumes that the classification of a particular dataset is given, in order to estimate the expected performance of the individual shape descriptors for the submitted query and automatically weigh the contribution of each method. However, in the general case, the classification of a 3D model dataset is not fixed since the content of a 3D model dataset is not static. In the context of partial shape matching, Funkhouser and Shilane (2006) use the predicted distinction performance of a set of descriptors based on a preceding training stage and perform a priority driven search in the space of feature correspondences to determine the best match of features between a pair of models. The disadvantages of this approach is its time complexity which is prohibitive for online interaction as well as the storage requirements for the descriptors of all the models in the database. Based on the idea of combining features obtained from 2D and 3D representations, Song and Golshani (2003) developed a descriptor that described an object by obtaining a set of orthogonal projections from different viewpoints and by measuring the curvature of the object's surface. Similar in spirit, Papadakis et al. (2008a)

proposed a hybrid descriptor formed by combining features extracted from a depth-buffer and spherical-function based representation, with enhanced translation and rotation invariance properties. The advantage of this method over similar approaches is the top discriminative power along with minimum space and time requirements.

## 2.2 Relevance Feedback in 3D Object Retrieval

In order to enable the machine to retrieve information through adapting to individual categorization criteria, relevance feedback (RF) was introduced as a means to involve the user in the retrieval process and guide the retrieval system towards the target. Relevance feedback was first used to improve text retrieval (Rocchio 1971), later on successfully employed in image retrieval systems and lately in a few 3D object retrieval systems. It is the information that is acquired from the user's interaction with the retrieval system about the relevance of a subset of the retrieved results. Further information on relevance feedback methods can be found in Ruthven and Lalmas (2003), Crucianu et al. (2004), Zhou and Huang (2001) and Papadakis et al. (2008b).

*Local relevance feedback* (LRF), also known as *pseudo* or *blind relevance feedback*, is different from the conventional approach in that the user does not actually provide any feedback at all. Instead, the required training data are obtained based only on the unsupervised retrieval result. The procedure comprises two steps. First, the user submits a query to the system which uses a set of low-level features to produce a ranked list of results which is not displayed to the user. Second, the system reconfigures itself by only using the top  $m$  matches of the list, based on the assumption that most likely they are relevant to the user's query.

LRF was first employed in the context of text retrieval, in order to extend the keywords comprising the query with related words from the top ranked retrieved documents. Apart from a few studies that incorporated RF in 3D object retrieval (Elad et al. 2001; Bang and Chen 2002; Atmosukarto et al. 2005; Lou et al. 2003; Leifman et al. 2005; Akbar et al. 2006; Novotni et al. 2005), LRF has only lately been examined in Papadakis et al. (2008b).

## 3 Computation of the PANORAMA Descriptor

In this section, we first describe the steps for the computation of the proposed descriptor (PANORAMA), namely: (i) pose normalization (Sect. 3.1), (ii) extraction of the panoramic views (Sect. 3.2) and (iii) feature extraction (Sect. 3.3). Finally, in Sect. 3.4 we describe a weighing scheme that is applied to the features and the procedure for comparing two PANORAMA descriptors.

### 3.1 Pose Normalization

Prior to the extraction of the PANORAMA descriptor, we must first normalize the pose of a 3D object, since the translation, rotation and scale characteristics should not influence the measure of similarity between objects.

To normalize the translation of a 3D model we compute its centroid using CPCA (Vranic 2004). In CPCA, the centroid of a 3D mesh model is computed as the average of its triangle centroids where every triangle is weighed proportionally to its surface area. We translate the model so that its centroid coincides with the origin and translation invariance is achieved as the centroids of all 3D models coincide.

To normalize for rotation, we use CPCA and NPCA (Papadakis et al. 2007) in order to align the principal axes of a 3D model with the coordinate axes. First, we align the 3D model using CPCA to determine its principal axes using the model's spatial surface distribution and then we use NPCA to determine its principal axes using the surface orientation distribution. Both methods use Principal Component Analysis (PCA) to compute the principal axes of the 3D model. The difference between the two methods lies in the input data that are used for the computation of the covariance matrix. In particular, in CPCA the surface area coordinates are used whereas in NPCA the surface orientation coordinates are used which are obtained from the triangles' normal vectors. The detailed description regarding the formulation of CPCA and NPCA can be found in Vranic (2004) and in our previous work (Papadakis et al. 2007), respectively.

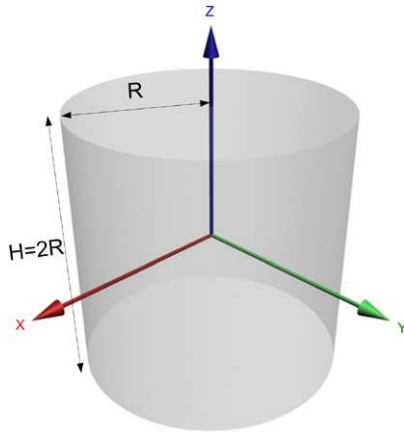
Thus, we obtain two alternative aligned versions of the 3D model, which are separately used to extract two sets of features that are integrated into a single feature vector (see Sect. 3.4).

The PANORAMA shape descriptor is rendered scale invariant, by normalizing the corresponding features to the unit  $L_1$  norm. As will be later described in Sects. 3.3.1 and 3.3.2, the features used by the PANORAMA descriptor are obtained from the 2D Discrete Fourier Transform and 2D Discrete Wavelet Transform. The corresponding coefficients are proportional to the object's scale, therefore by normalizing the coefficients to their unit  $L_1$  norm we are in fact normalizing all objects to the same scale.

### 3.2 Extraction of Panoramic Views

After the normalization of a 3D model's pose, the next step is to acquire a set of panoramic views.

To obtain a panoramic view, we project the model to the lateral surface of a cylinder of radius  $R$  and height  $H = 2R$ , centered at the origin with its axis parallel to one of the coordinate axes (see Fig. 1). We set the value of  $R$  to  $3 * d_{mean}$  where  $d_{mean}$  is the mean distance of the model's surface from its centroid. For each model, the value of  $d_{mean}$  is determined using the diagonal elements of the covariance matrix



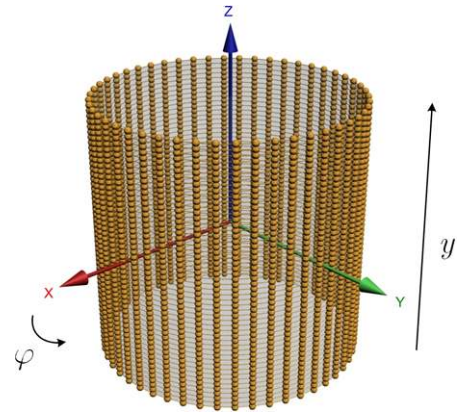
**Fig. 1** The cylinder used for acquiring the projection of a 3D model

used in CPCA that give the average distances of the model’s surface from the coordinate axes. Setting  $R = 3 * d_{mean}$  does not imply that a 3D model will necessarily lie completely inside the cylinder. However, this approach is better than using a bounding cylinder, that may contain the model in small scale in the presence of outlying parts of the model. The empirical value  $3 * d_{mean}$  enables the majority of 3D models to lie completely inside the cylinder, while having a suitable scale to produce a descriptive projection.

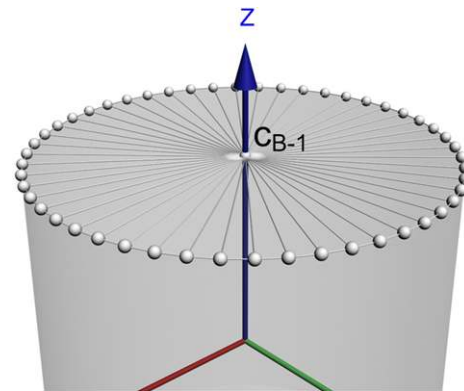
In the following, we parameterize the lateral surface of the cylinder using a set of points  $s(\varphi, y)$  where  $\varphi \in [0, 2\pi]$  is the angle in the  $xy$  plane,  $y \in [0, H]$  and we sample the  $\varphi$  and  $y$  coordinates at rates  $2B$  and  $B$ , respectively (we set  $B = 64$ ). We sample the  $\varphi$  dimension at twice the rate of the  $y$  dimension to account for the difference in length between the perimeter of the cylinder’s lateral surface and its height. Although the perimeter of the cylinder’s lateral surface is  $\pi \simeq 3$  times its height, we set the sampling rates at  $2B$  and  $B$ , respectively, since it was experimentally found as the best trade-off between effectiveness and efficiency. Thus we obtain the set of points  $s(\varphi_u, y_v)$  where  $\varphi_u = u * 2\pi / (2B)$ ,  $y_v = v * H / B$ ,  $u \in [0, 2B - 1]$  and  $v \in [0, B - 1]$ . These points are shown in Fig. 2.

The next step is to determine the value at each point  $s(\varphi_u, y_v)$ . The computation is carried out iteratively for  $v = 0, 1, \dots, B - 1$ , each time considering the set of coplanar  $s(\varphi_u, y_v)$  points i.e. a cross section of the cylinder at height  $y_v$  and for each cross section we cast rays from its center  $c_v$  in the  $\varphi_u$  directions. In Fig. 3, we show the points  $s(\varphi_u, y_v)$  of the top-most cross section ( $v = B - 1$ ) of the projection cylinder along with the corresponding rays emanating from the center of the cross section.

The cylindrical projections are used to capture two characteristics of a 3D model’s surface; (i) the position of the model’s surface in 3D space and (ii) the orientation of the model’s surface. To capture these characteristics we use two kinds of cylindrical projections  $s_1(\varphi_u, y_v)$  and  $s_2(\varphi_u, y_v)$ .



**Fig. 2** The discretization of the lateral surface of the projection cylinder (points in orange) to the set of points  $s(\varphi_u, y_v)$



**Fig. 3** The top-most cross section of the cylinder along with the corresponding rays emanating from the center of the cross section  $c_{B-1}$

By default, the initial value of a point  $s_k(\varphi_u, y_v)$ ,  $k \in \{1, 2\}$  is set to zero.

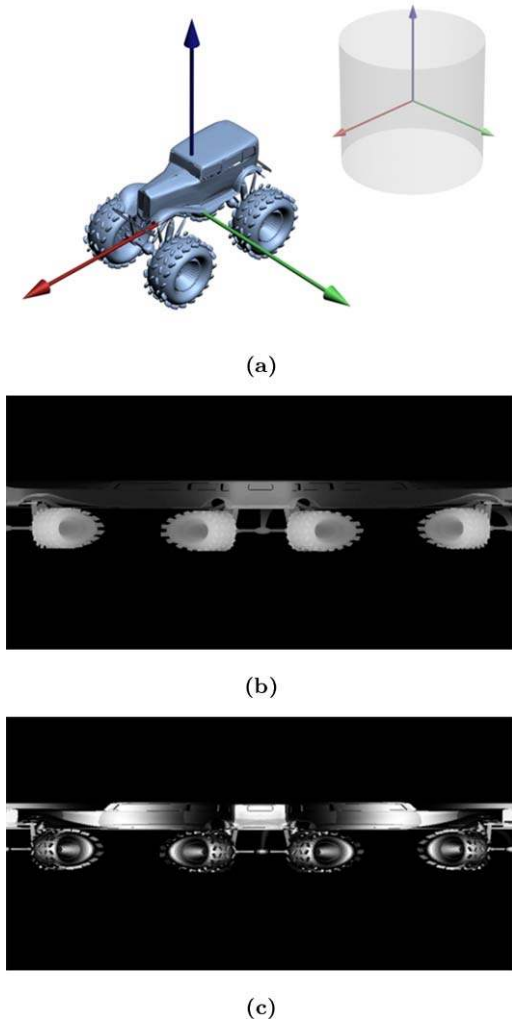
To capture the position of the model’s surface, for each cross section at height  $y_v$  we compute the distances from  $c_v$  of the intersections of the model’s surface with the rays at each direction  $\varphi_u$ .

Let  $pos(\varphi_u, y_v)$  denote the distance of the furthest from  $c_v$  point of intersection between the ray emanating from  $c_v$  in the  $\varphi_u$  direction and the model’s surface, then:

$$s_1(\varphi_u, y_v) = pos(\varphi_u, y_v) \tag{1}$$

Thus the value of a point  $s_1(\varphi_u, y_v)$  lies in the range  $[0, R]$ , where  $R$  denotes the radius of the cylinder.

A cylindrical projection can be viewed as a 2D gray-scale image where pixels correspond to the  $s_k(\varphi_u, y_v)$  intersection points in a manner reminiscent of cylindrical texture mapping (see Theoharis et al. 2008) and their values are mapped to the  $[0, 1]$  space. In Fig. 4(a), we show an example 3D model along with a projection cylinder aligned with the  $z$ -axis. In Fig. 4(b) the unfolded visual representation of its corresponding cylindrical projection  $s_1(\varphi_u, y_v)$  is given.



**Fig. 4** (a) Pose normalized 3D model, (b) the unfolded cylindrical projection of the 3D model capturing the position of the surface (c) the unfolded cylindrical projection of the 3D model capturing the orientation of the surface

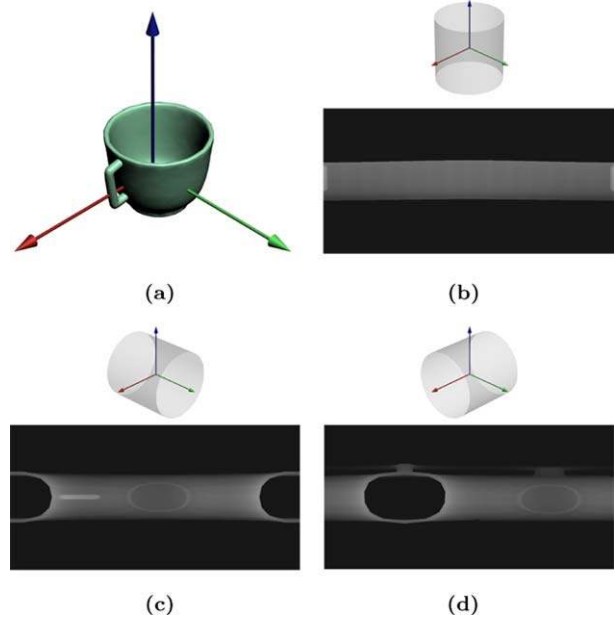
To capture the orientation of the model’s surface, for each cross section at height  $y_v$  we compute the intersections of the model’s surface with the rays at each direction  $\varphi_u$  and measure the angle between a ray and the normal vector of the triangle that is intersected. To determine the value of a point  $s_2(\varphi_u, y_v)$  we use the cosine of the angle between the ray and the normal vector of the furthest from  $c_v$  intersected triangle of the model’s surface.

Let  $ang(\varphi_u, y_v)$  denote the aforementioned angle then the values of the  $s_2(\varphi_u, y_v)$  points are given by:

$$s_2(\varphi_u, y_v) = |\cos(ang(\varphi_u, y_v))|^n \tag{2}$$

In Fig. 4(c) the unfolded visual representation of the cylindrical projection  $s_2(\varphi_u, y_v)$  is given for the model shown in Fig. 4(a).

We take the  $n$ th power of  $|\cos(ang(\varphi_u, y_v))|$ , where  $n \geq 2$ , since this setting enhances the contrast of the pro-



**Fig. 5** (a) A 3D model of a cup and (b)–(d) the corresponding cylindrical projections  $s_{1,t}(\varphi_u, y_v)$  using three cylinders each one aligned with the  $z$ ,  $y$  and  $x$  coordinate axis, respectively

duced cylindrical projection which was experimentally found to be more discriminative. Setting  $n$  to a value in the range [4,6] gives the best results. Also, taking the absolute value of the cosine is necessary to deal with inconsistently oriented triangles along the object’s surface. We do not enhance the contrast of the  $s_1(\varphi_u, y_v)$  projection since it was found to produce less discriminative features.

Although the cylindrical projection captures a large part of the shape of a model, a single cylindrical projection may not be able to capture concave parts. In Fig. 5 we show a typical example of this. Figure 5(a) shows the 3D model of a cup and Fig. 5(b) shows the produced cylindrical projection when using a projection cylinder that is aligned with the  $z$ -axis. As can be observed, this projection cannot capture the interior part of the cup.

To alleviate this problem, we compute cylindrical projections from differently oriented cylinders, i.e. cylinders that are aligned with all coordinate axes in order to obtain a more informative 3D model representation. Thus, we project a 3D model to the lateral surfaces of three cylinders, each one aligned with a different coordinate axis as shown in Fig. 5(b)–(d), which produces three sets of points  $s_{k,t}(\varphi_u, y_v)$  for  $k \in \{1, 2\}$ , where  $t \in \{x, y, z\}$  denotes the axis that the cylinder is aligned with. For the cylindrical projections that are aligned with the  $x$  and  $y$  axes, the  $\varphi_u$  variable is measured at the  $yz$  and  $zx$  planes respectively while all other notations remain the same.

### 3.3 Feature Representation of Panoramic Views

In this section, we detail the procedure for the generation of a set of features that describe a panoramic view. Toward this goal, we use the 2D Discrete Fourier Transform (Sect. 3.3.1) and 2D Discrete Wavelet Transform (Sect. 3.3.2).

#### 3.3.1 2D Discrete Fourier Transform

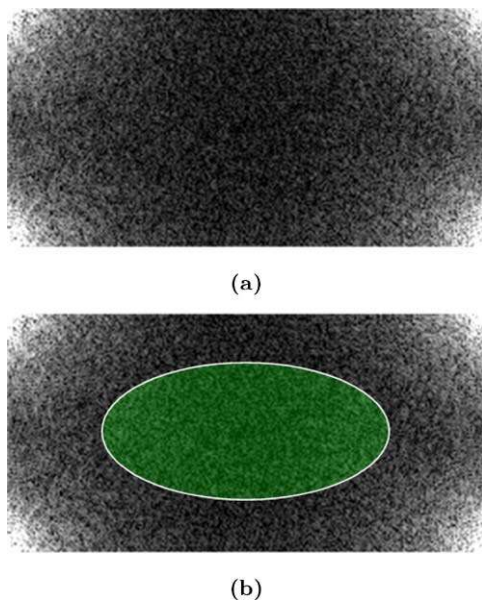
For each cylindrical projection  $s_{k,t}(\varphi_u, y_v)$  ( $k \in \{1, 2\}$  and  $t \in \{x, y, z\}$ ), we compute the corresponding 2D Discrete Fourier Transform (DFT), which is given by:

$$F_{k,t}(m, n) = \sum_{u=0}^{2B-1} \sum_{v=0}^{B-1} s_{k,t}(\varphi_u, y_v) \cdot e^{-2j\pi(\frac{mu}{2B} + \frac{nv}{B})} \quad (3)$$

where  $m \in [0, 2B - 1]$  and  $n \in [0, B - 1]$ .

Since  $s_{k,t}(\varphi_u, y_v)$  is a real-valued function, the Hermitian symmetry property holds for the Fourier coefficients, i.e.  $F_{k,t}(u, v) = F_{k,t}^*(2B - u, B - v)$ , where  $F^*$  denotes the complex conjugate. Hence, the size of the non-redundant information is a set of  $(B + 1) * (\frac{B}{2} + 1)$  Fourier coefficients for each projection. Next, we store the absolute values of the real and imaginary part of each coefficient and normalize the coefficients to the unit  $L_1$  norm, which ensures scaling invariance as explained in Sect. 3.1.

In practice, most of the energy of the Fourier coefficients is concentrated on the four corners of the image of the transform, as can be seen in Fig. 6(a). Therefore, we only keep a subset of the full set of Fourier coefficients, i.e. those containing most of the energy. This can be straightforwardly



**Fig. 6** (a) A typical 2D Fourier transform of a cylindrical projection; (b) The Fourier coefficients that lie inside the area of the ellipsoid are discarded to reduce dimensionality

done by considering an ellipse positioned at the center of the Fourier image and discarding all the coefficients that lie on the interior of the ellipse, as shown in Fig. 6(b). The ratio of the width to the height of the ellipse is equal to the ratio of the width to the height of the Fourier image and the size of the ellipse is set so that the low energy coefficients that are discarded correspond to approximately 35% of the total number of coefficients per projection.

After the completion of all previous operations, the resulting coefficients are denoted by  $\tilde{F}_{k,t}$ . Thus, the final feature set  $\mathbf{s}_F$  of Fourier coefficients for a particular aligned version of a 3D object is denoted by:

$$\mathbf{s}_F = (\tilde{F}_{1,x}, \tilde{F}_{2,x}, \tilde{F}_{1,y}, \tilde{F}_{2,y}, \tilde{F}_{1,z}, \tilde{F}_{2,z}) \quad (4)$$

#### 3.3.2 2D Discrete Wavelet Transform

For each cylindrical projection  $s_{k,t}(\varphi_u, y_v)$  ( $k \in \{1, 2\}$  and  $t \in \{x, y, z\}$ ), we compute the corresponding 2D Discrete Wavelet Transform (DWT), which is given by:

$$W_{k,t}^\varphi(j_0, m, n) = \frac{1}{\sqrt{2B \cdot B}} \cdot \sum_{u=0}^{2B-1} \sum_{v=0}^{B-1} s_{k,t}(\varphi_u, y_v) \cdot \varphi_{j_0,m,n}(u, v), \quad (5)$$

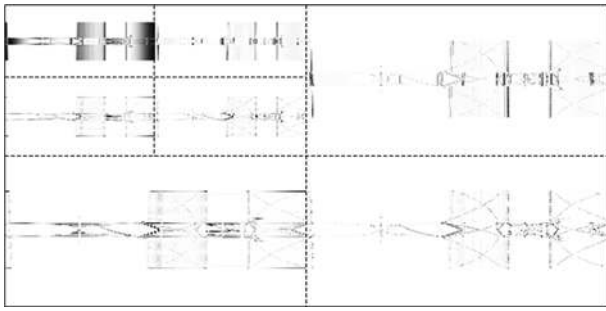
$$W_{k,t}^\psi(j, m, n) = \frac{1}{\sqrt{2B \cdot B}} \cdot \sum_{u=0}^{2B-1} \sum_{v=0}^{B-1} s_{k,t}(\varphi_u, y_v) \cdot \psi_{j,m,n}(u, v) \quad (6)$$

where  $m \in [0, 2B - 1]$ ,  $n \in [0, B - 1]$ ,  $j \geq j_0$  denotes the scale of the multi-level DWT,  $j_0$  is the starting scale and  $\varphi_{j_0,m,n}(u, v)$ ,  $\psi_{j,m,n}(u, v)$  denotes the scaling and wavelet function, respectively. The  $W_{k,t}^\varphi(j_0, m, n)$  approximation-scaling coefficients correspond to the low-pass subband of the transform at the starting scale  $j_0$ . The  $W_{k,t}^\psi(j, m, n)$  detail-wavelet coefficients correspond to the vertical, horizontal and diagonal subbands. We take the absolute values of the coefficients and normalize to their  $L_1$  norm, which are now denoted as  $\tilde{W}_{k,t}^\varphi(j_0, m, n)$  and  $\tilde{W}_{k,t}^\psi(j, m, n)$ .

In Fig. 7, we show the image of a 2-level wavelet transformation of a cylindrical projection. The transform is shown in negative colors to better point out the detail coefficients.

In our implementation, we computed the DWT of each cylindrical projection up to the last level. In particular, since the dimensions of a cylindrical projection are  $(2B) \cdot (B)$ , the total number of levels of the DWT are  $\log_2 B$  and thus  $j = 0, 1, \dots, \log_2 B - 1$ .

To compute the DWT we have used the Haar and Coiflet (C6) filters (basis functions), as they attained the best overall performance. Therefore, two distinctive DWTs are computed, the first using the Haar and the second using the Coiflet basis functions.



**Fig. 7** 2-level wavelet transformation of a cylindrical projection of an airplane (the image is shown in negative colors)

Instead of using directly the coefficients of the DWT as shape features, we compute a set of standard statistic image features that are listed below:

**i. Mean**

$$\mu = \frac{1}{n_x n_y} \sum_{x=1}^{n_x} \sum_{y=1}^{n_y} I(x, y) \tag{7}$$

**ii. Standard deviation**

$$\sigma = \sqrt{\frac{1}{n_x n_y} \sum_{x=1}^{n_x} \sum_{y=1}^{n_y} (I(x, y) - \mu)^2} \tag{8}$$

**iii. Skewness**

$$\beta = \frac{\frac{1}{n_x n_y} \sum_{x=1}^{n_x} \sum_{y=1}^{n_y} (I(x, y) - \mu)^3}{\sigma^3} \tag{9}$$

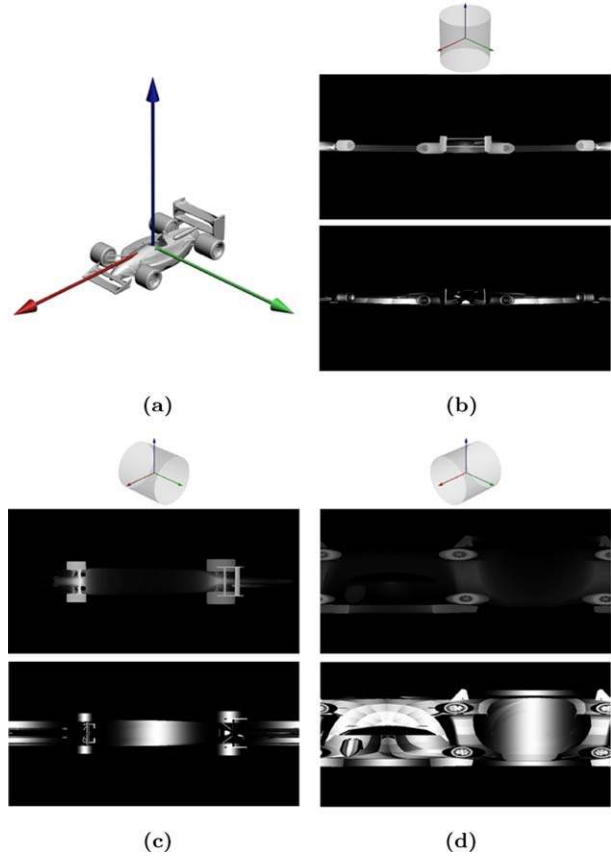
where  $n_x, n_y$  are the dimensions of an image  $I(x, y)$ .

In our case, the above features are computed for each sub-image of the DWT, i.e. for each subband in every level of the DWT. Therefore,  $I(x, y)$  is replaced by the  $\tilde{W}_{k,t}^\varphi(j_0, m, n)$  and  $\tilde{W}_{k,t}^\psi(j, m, n)$  coefficients accordingly. Since we have  $\log_2 B$  levels in the DWT, the total number of subbands per cylindrical projection is  $(3 * \log_2 B + 1)$ . Hence, since we use two wavelet basis functions and three statistic image features, we obtain a total of  $2 * 3 * (3 * \log_2 B + 1)$  features per cylindrical projection which are denoted as  $\tilde{V}_{k,t}$  and the final feature set  $\mathbf{s}_W$  for a particular aligned version of a 3D object is denoted as:

$$\mathbf{s}_W = (\tilde{V}_{1,x}, \tilde{V}_{2,x}, \tilde{V}_{1,y}, \tilde{V}_{2,y}, \tilde{V}_{1,z}, \tilde{V}_{2,z}) \tag{10}$$

**3.4 Features Weighing and Matching**

The features that were generated for each panoramic view are weighed by a factor  $w_t$ , according to the orientation of the cylinder ( $x, y$  or  $z$ ) that was used to acquire the cylindrical projection. We apply this weighing based on the observation that not all cylindrical projections capture



**Fig. 8** (a) A 3D model of a car and (b)–(d) the corresponding cylindrical projections  $s_{1,t}(\varphi_u, y_v)$  and  $s_{2,t}(\varphi_u, y_v)$  for  $t = z, t = y$  and  $t = x$ , respectively

the same amount of information about the model’s shape. The  $t$ -projection cylinder that is parallel to the  $t$  coordinate axis corresponds to one of the principal axes of the 3D model that were determined in the rotation normalization step. The amount of information that is captured by the  $t$ -cylindrical projection is directly related to the principal axes of the model’s surface that are encoded, as is demonstrated in Fig. 8 for an example 3D model of a race car. In this example, the first, second and third principal axis of the object’s surface is aligned with the  $x, y$  and  $z$  axis, respectively, therefore the most informative cylindrical projection corresponds to the  $x$ -projection cylinder (Fig. 8(d)), followed by the  $y$ -projection cylinder (Fig. 8(c)) and the least informative  $z$ -projection cylinder (Fig. 8(b)). We set the factors  $w_t$  to fixed values that were determined experimentally as  $w_x = 0.51, w_y = 0.31$  and  $w_z = 0.18$ .

To compute the distance between the same aligned version (CPCA or NPCA) of two 3D objects, we compute the distances between the corresponding sets of features. We denote the features set of a particular aligned version of an object by:

$$\mathbf{pl} = (\mathbf{s}_{F,l}, \mathbf{s}_{W,l}), l \in \{cpca, npca\} \tag{11}$$



and the full PANORAMA descriptor of 3D object by:

$$\mathbf{P} = (\mathbf{p}_{cpca}, \mathbf{p}_{npca}) \quad (12)$$

Considering the features sets  $\mathbf{p}_l, \hat{\mathbf{p}}_l$  of two 3D objects of the same aligned version  $l$ , the distance is computed by:

$$d_l(\mathbf{p}_l, \hat{\mathbf{p}}_l) = L_1(\mathbf{s}_{F,l}, \hat{\mathbf{s}}_{F,l}) + D_{can}(\mathbf{s}_{W,l}, \hat{\mathbf{s}}_{W,l}) \quad (13)$$

where  $L_1(\cdot)$ ,  $D_{can}(\cdot)$  denotes the Manhattan and the Canberra distance (Kokare et al. 2003), respectively, each one normalized to the  $[0, 1]$  space.

The overall similarity between two 3D objects is measured by computing the distance between the sets of features of the same aligned version and the comparison that gives the minimum distance between the two alignments, sets the final score. Thus the overall distance  $D(\mathbf{P}, \hat{\mathbf{P}})$  between two PANORAMA descriptors  $\mathbf{P}$  and  $\hat{\mathbf{P}}$  is given by:

$$D(\mathbf{P}, \hat{\mathbf{P}}) = \min_l d_l(\mathbf{p}_l, \hat{\mathbf{p}}_l) \quad (14)$$

#### 4 Local Relevance Feedback (LRF) Technique

In this section, we outline the major points of the employed LRF method whose detailed description can be found in Papadakis et al. (2008b). The method comprises two stages, the on-line and the off-line stage.

During the off-line stage, the descriptor of each stored object is updated using its  $k$  nearest neighbors in feature space. These are assumed to belong to the same class and the updated descriptor is the average of the original descriptor and a weighed centroid of the  $k$  nearest neighbors.

During the on-line stage, a user submits a query to the system which finds within the updated feature space its  $k$  nearest neighbors and updates its descriptor according to rule that was applied during the off-line stage. The updated descriptor of the query is then used to measure its similarity against every object of the database and display the results.

The relevance assumption for the  $k$  nearest neighbors is not always valid as irrelevant objects that may belong to the  $k$  nearest neighbors will be mistaken as relevant. This is known as *query drift* and implies the scenario where the retrieval system is misled by the irrelevant data and drawn away from the user's target. However, if the features that are used to compare two objects are discriminative enough to cluster most objects that belong to the same class, then the relevance assumption will be valid in the majority of cases and the average performance will be better.

In Papadakis et al. (2008b) it was shown that LRF increased the performance of the CRSP descriptor (Papadakis et al. 2007). The PANORAMA descriptor is far more discriminative therefore reducing the negative effect of the *query drift* phenomenon and rendering LRF more applicable.

The number  $k$  of nearest neighbors that are considered as relevant to a query is determined by the expected recall of the employed shape features near the neighborhood of a query. In this case, setting  $k = 4$  gave the best results which amounts to approximately 12% of the objects of a class on average, for the datasets used in Sect. 5.

## 5 Results

We next present the results of an extensive evaluation of the proposed PANORAMA descriptor and LRF technique. We tested performance using the following standard 3D model datasets:

- (i) The classified dataset of CCCC (Vranic 2004).
- (ii) The dataset of the National Institute of Standards and Technology (NIST, Fang et al. 2008).
- (iii) The dataset of the Watertight Models track of SHape REtrieval Contest 2007 (WM-SHREC) (Veltkamp and ter Haar 2007).
- (iv) The MPEG-7 dataset (<http://www.chiariglione.org/mpeg/>).
- (v) The test dataset of the Princeton Shape Benchmark (PSB) (Shilane et al. 2004).
- (vi) The dataset of the Engineering Shape Benchmark (ESB) (Jayanti et al. 2006).

Table 1 gives the characteristics of these datasets.

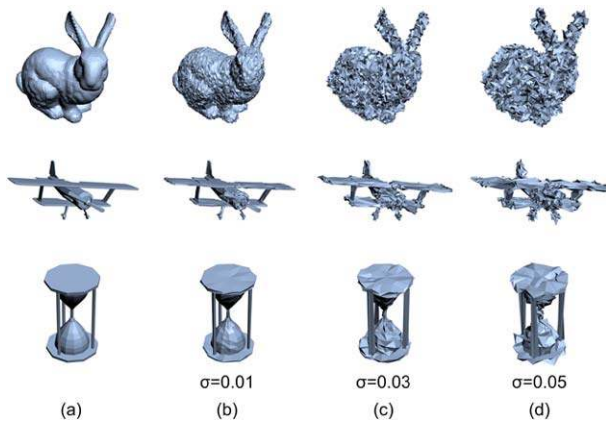
To evaluate the performance we use precision-recall diagrams. Recall is the ratio of relevant to the query retrieved models to the total number of relevant models while precision is the ratio of relevant to the query retrieved models to the number of retrieved models. The evaluations were performed by using each model of a dataset as a query on the remaining set of models and computing the average precision-recall performance overall models.

### 5.1 Robustness

In this section we test the robustness of the PANORAMA descriptor under the presence of noise. We have experimented with various degrees of Gaussian noise added along

**Table 1** Characteristics of the evaluation datasets

Dataset	#models	#classes	Type
CCCC	472	55	generic
NIST	800	40	generic
WM-SHREC	400	20	generic
MPEG-7	1300	135	generic
PSB	907	92	generic
ESB	866	48	CAD



**Fig. 9** (a) Example 3D objects and (b)–(d) the effect of different degrees of additive Gaussian noise on their surface

**Table 2** Effect of noise in the determination of principal axes for the CPCA and NPCA rotation normalization methods within the PSB dataset

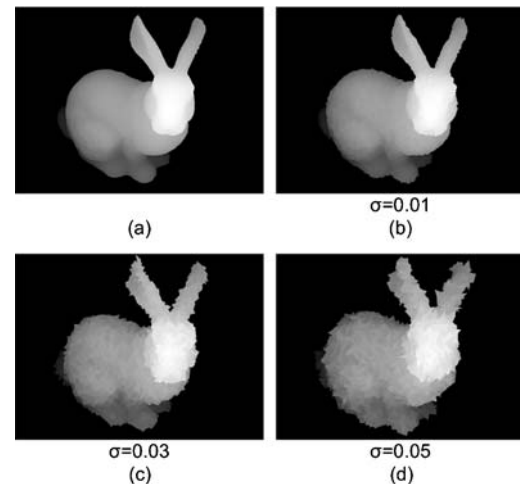
Alignment method	$\sigma = 0.01$	$\sigma = 0.03$	$\sigma = 0.05$
CPCA	6°	6.6°	7.4°
NPCA	10.4°	14°	16.8°

the surface of each 3D object as shown in Fig. 9. Apparently, adding Gaussian noise with  $\sigma > 0.01$  has a destructive effect on the object's surface and most of the geometric details are lost. We believe that noisy 3D objects such as the ones shown in Fig. 9(c)–(d) are rarely encountered and practically useless, thus we are more interested in the robustness of the PANORAMA descriptor with respect to levels of noise that compare to the examples shown in Fig. 9(b).

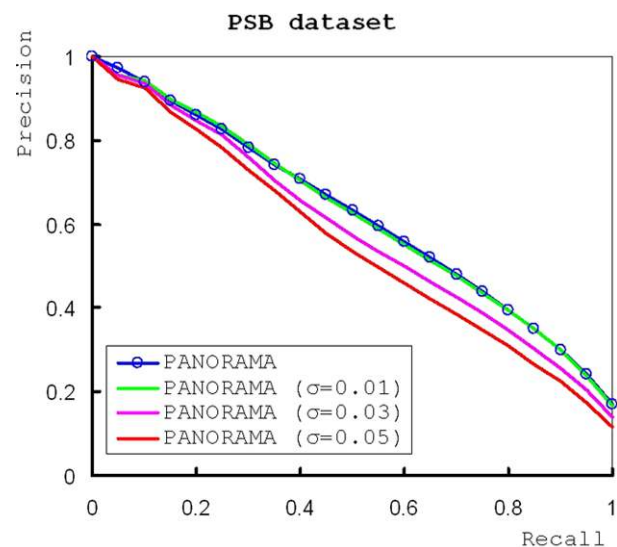
Since the PANORAMA descriptor is based on normalizing the rotation of a 3D object, we measured the effect of noise in the determination of the object's principal axes for CPCA and NPCA. The results of this experiment are given in Table 2 where we show the average angular perturbation of the object's principal axes after the addition of noise within the PSB dataset.

It is easy to understand that a certain change in the coordinates of the vertices that comprise a polygon has a greater relative impact on the orientation of the polygon's normal vector. This can be demonstrated in Fig. 10 where we show the depth buffers that are obtained after adding the same amounts of Gaussian noise for the bunny 3D object of Fig. 9(a). Apparently, the effect of noise is more clearly noticed in Fig. 9 in contrast to Fig. 10 where only the depth is used. Thus, the fact that the NPCA alignment method is affected by noise to a greater degree than CPCA is a reasonable result. Hence, we can say that CPCA is more robust than NPCA with respect to noise.

Nevertheless, a perturbation of the object's principal axes after the addition of noise does not necessarily mean that the



**Fig. 10** Demonstration of the effect of different amounts of additive Gaussian noise on a depth buffer



**Fig. 11** Demonstration of the effect of different amounts of additive Gaussian noise on the retrieval performance of the PANORAMA descriptor

alignment of 3D objects is worse and leads to reduced performance in 3D object retrieval. For example, if the principal axes of 3D objects are perturbed toward the same directions for objects of the same class, then we expect to attain similar retrieval performance.

In the sequel, we evaluate the performance of the PANORAMA descriptor before and after the addition of different levels of Gaussian noise along the surface of 3D objects as was performed previously. However, we should note that the performance of the descriptor is not only affected by the perturbed principal axes but also by the alternated surfaces of 3D objects after the addition of noise which results in different cylindrical projections as well as features. In Fig. 11, we demonstrate the performance in terms of pre-

recision and recall within the PSB dataset. This experiment demonstrates that the PANORAMA descriptor is very robust with respect to reasonable levels of noise, as in the case where  $\sigma = 0.01$ . In this case, we can see that the performance of PANORAMA is totally unaffected while for greater levels of noise ( $\sigma = 0.03$  and  $\sigma = 0.05$ ) the performance decreases gracefully compared to the destructive effect of noise that is demonstrated in Fig. 9.

## 5.2 Justification of PANORAMA settings

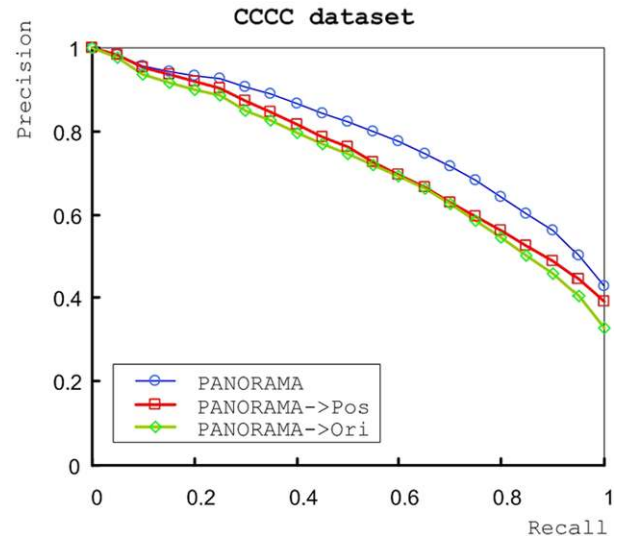
In this section we demonstrate the increase in retrieval performance that is achieved due to specific settings of the PANORAMA descriptor. In particular, in Fig. 12(a) we compare the performance when using  $s_{1,t}(\varphi_u, y_v)$ ,  $s_{2,t}(\varphi_u, y_v)$  or both and in Fig. 12(b) we evaluate the performance between using a single and using three perpendicular cylinders aligned with the coordinate axes. The results clearly show that a significant increase in discriminative power is attained when capturing both the surface position and orientation as well as using three perpendicular cylinders instead of one.

## 5.3 Comparative Evaluation

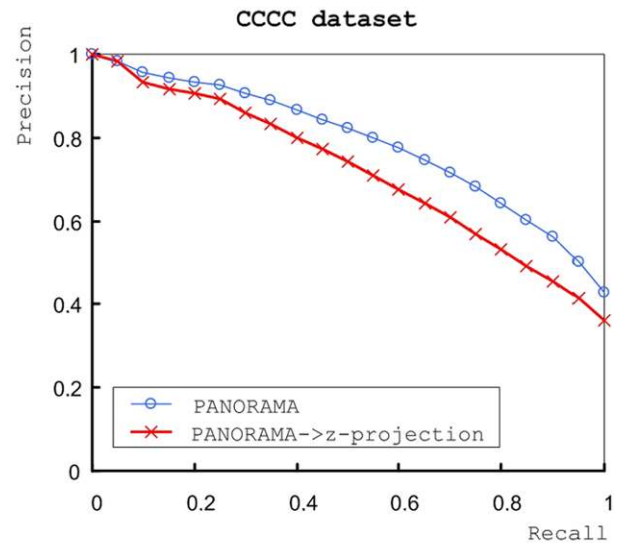
We next compare the PANORAMA descriptor against the following state-of-the-art methods:

- The **2D/3D Hybrid** descriptor developed by Papadakis et al. (2008a).
- The **DESIRE** descriptor developed by Vranic (2005).
- The Light Field descriptor (**LF**) developed by Chen et al. (2003).
- The spherical harmonic representation of the Gaussian Euclidean Distance Transform descriptor (**SH-GEDT**) developed by Kazhdan et al. (2003).

In Fig. 13, we give the precision-recall diagrams comparing the proposed PANORAMA descriptor against the other descriptors and show the increase in performance when LRF is employed. It is evident that the PANORAMA descriptor attains a better overall performance compared to the other methods. Interestingly, although the LF descriptor uses a plurality of 2D images (100 orthogonal projections) it is outperformed by the PANORAMA descriptor which uses a total of 12 cylindrical projections that are acquired from just three perpendicular cylinders. In addition, the PANORAMA descriptor is more discriminative than the 2D/3D Hybrid descriptor which uses a total of 12 orthogonal projections combined with 48 spherical functions. This strongly suggests that the cylindrical projection is a more effective representation for the purpose of 3D object retrieval compared to the conventional orthogonal projection and spherical function based representation. This is also confirmed by comparing



(a)

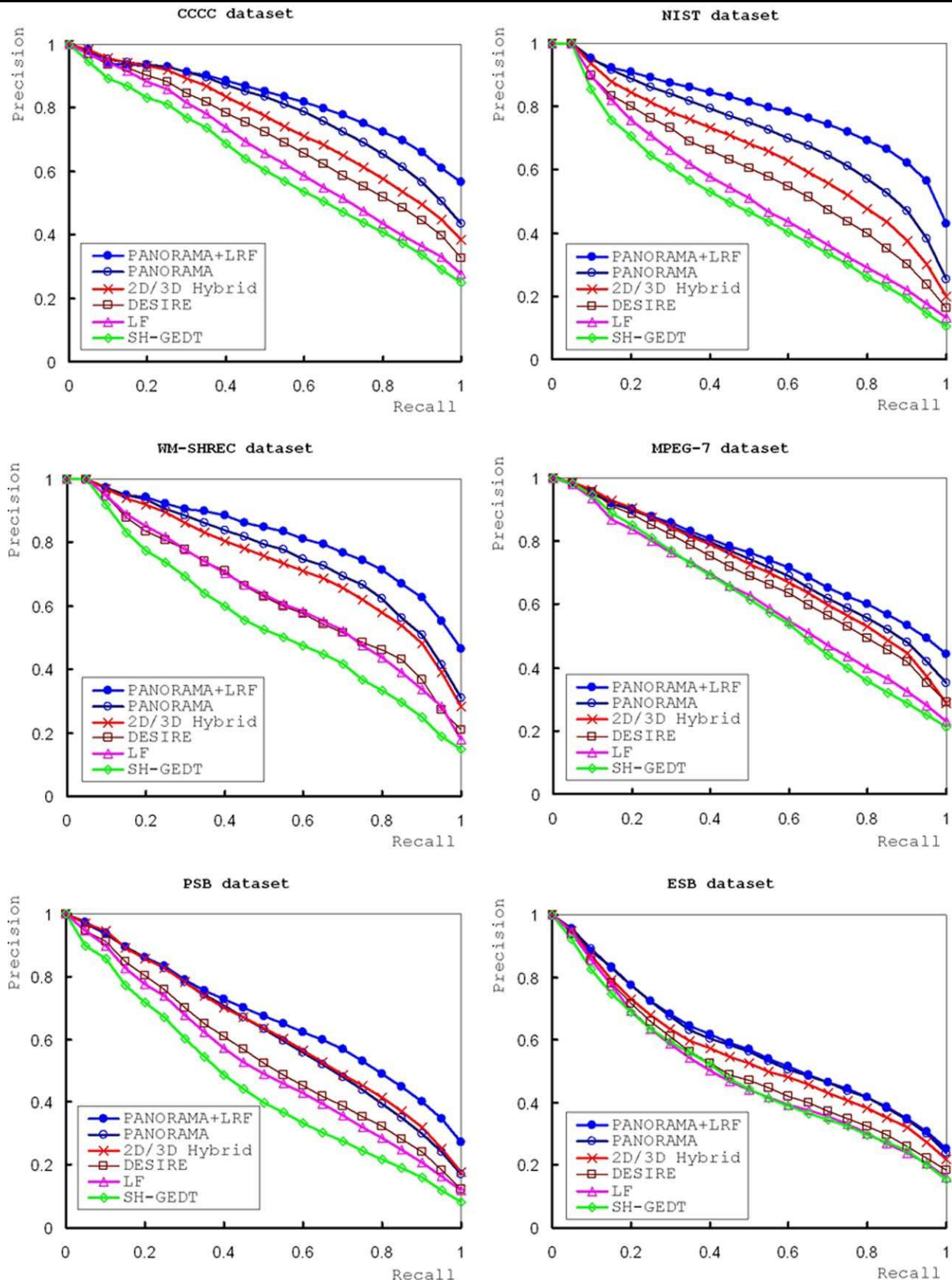


(b)

**Fig. 12** Performance evaluation of the PANORAMA descriptor when using: (a)  $s_{1,t}(\varphi_u, y_v)$  (PANORAMA  $\rightarrow$  Pos),  $s_{2,t}(\varphi_u, y_v)$  (PANORAMA  $\rightarrow$  Ori) or both (PANORAMA); (b) a single cylinder aligned with the  $z$  coordinate axis (PANORAMA  $\rightarrow$   $z$ -projection) and a set of three perpendicular cylinders aligned with the object's principal axes (PANORAMA)

PANORAMA with the DESIRE descriptor that also uses 12 orthogonal projections (6 depth buffers and 6 silhouettes) as well as features extracted using a spherical function.

In addition, it is evident that LRF adds a major gain in the overall performance, particularly on the CCCC, NIST, WM-SHREC and PSB datasets. This indicates that the *query drift* phenomenon that comes with LRF is compensated for the increased precision of the PANORAMA descriptor within the top retrieved results. These results are coherent to those that were obtained using the CRSP descriptor in



**Fig. 13** Precision-recall plots comparing the proposed PANORAMA descriptor against the 2D/3D Hybrid, DESIRE, LF and SH-GEDT descriptor in various 3D model datasets. The comparison includes the combination of the PANORAMA descriptor with local relevance feedback (LRF)

Papadakis et al. (2008b) and show that the employed LRF technique is general purpose and can be applied to any retrieval method as long as precision is high near the neighborhood of the query. It is also worth noticing that the LRF technique increases precision mainly at higher values of recall, as can be seen in Fig. 13. This implies that a user notices the increase in performance as he browses through the list of results until most or all models appear that belong to the query's class. It also implies that employing LRF more than once before showing the results to the user will not add any further gain in the retrieval performance. This is because the employed LRF technique uses the  $k$  nearest neighbors of a 3D model to update its descriptor but since precision is not increased at the early stages of recall, the  $k$  nearest neighbors of the model after employing LRF will mostly be the same as before employing LRF.

To further quantify the performance of the compared methods, we next employ the following measures:

- Nearest Neighbor (NN): The percentage of queries where the closest match belongs to the query's class.
- First Tier (FT): The recall for the  $(C - 1)$  closest matches were  $C$  is the cardinality of the query's class.
- Second Tier (ST): The recall for the  $2(C - 1)$  closest matches were  $C$  is the cardinality of the query's class.

These measures range from 0%–100% and higher values indicate better performance. In Table 3, we give the scores of each method for each measure for all datasets.

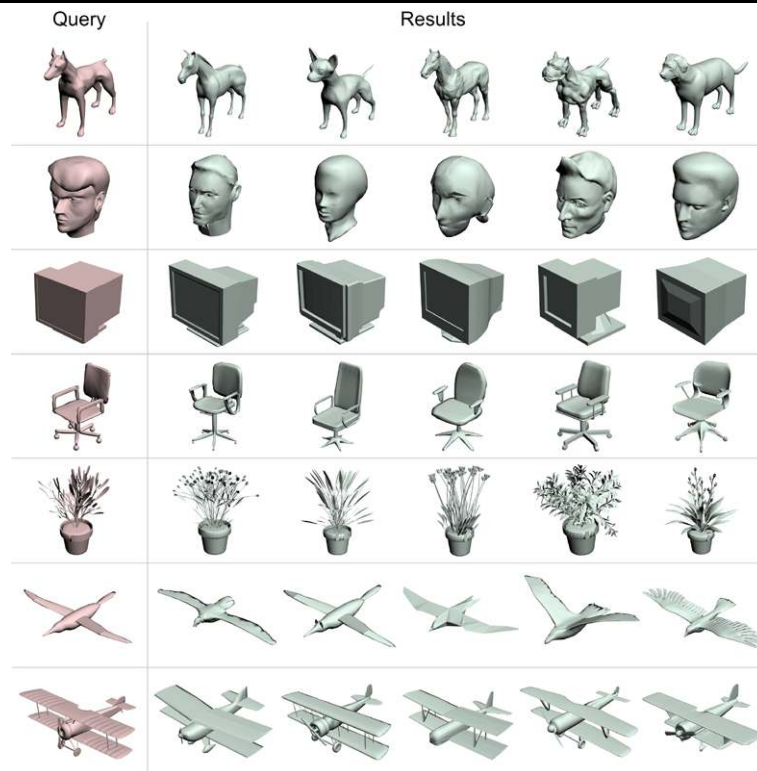
Apparently, the PANORAMA descriptor consistently outperforms all compared methods with respect to these measures as well. We can also observe that the nearest neighbor score is not particularly affected by the application of LRF, compared to the first and second tier measures whose values are significantly increased after employing LRF. This confirms our earlier conclusion that LRF does not change the precision at the early stages of recall and therefore it does not add any further gain in performance if it is applied multiple times.

As described in Sect. 3.3, the PANORAMA descriptor uses two kinds of features, namely those coming from the 2D DFT ( $s_F$ ) and the 2D DWT ( $s_W$ ). Comparing the retrieval performance between the two kinds, the  $s_F$  component was found to be more effective in most datasets than the  $s_W$  part. In fact, the  $s_W$  part had a slight advantage in retrieval performance within the WM-SHREC and ESB datasets. Overall however, the  $s_F$  part is about 3% better in terms of average precision, which can be justified by the fact that  $s_F$  has greater dimensionality than the  $s_W$  part. In particular, the dimensionality of  $s_F$  is  $2 * (B + 1) * (\frac{B}{2} + 1)$  while for the  $s_W$  component it is only  $2 * 3 * (3 * \log_2 B + 1)$ . Since we set the bandwidth to  $B = 64$ , this amounts to a dimension of 4160 and 114, respectively. This implies, that we can attain comparable retrieval performance using just the  $s_W$  part of the full descriptor.

**Table 3** Quantitative measures scores for the proposed PANORAMA, 2D/3D Hybrid, DESIRE, LF and SH-GEDT methods for the CCCC, NIST, WM-SHREC, MPEG-7, PSB and ESB datasets

Method	NN (%)	FT (%)	ST (%)
<b>CCCC</b>			
PANORAMA + LRF	87.4	<b>70.3</b>	<b>86.6</b>
PANORAMA	<b>87.9</b>	66.3	81.2
2D/3D Hybrid	87.4	60.2	75.8
DESIRE	82.8	55.6	70.0
LF	79.8	50.2	63.1
SH-GEDT	75.7	45.9	59.9
<b>NIST</b>			
PANORAMA + LRF	90.4	<b>71.5</b>	<b>84.1</b>
PANORAMA	<b>90.6</b>	63.4	77.5
2D/3D Hybrid	88.1	55.6	72.1
DESIRE	83.7	50.9	64.9
LF	84.1	43.9	56.0
SH-GEDT	76.5	40.5	53.7
<b>WM-SHREC</b>			
PANORAMA + LRF	<b>95.7</b>	<b>74.3</b>	<b>83.9</b>
PANORAMA	95.7	67.3	78.4
2D/3D Hybrid	95.5	64.2	77.3
DESIRE	91.7	53.5	67.3
LF	92.3	52.6	66.2
SH-GEDT	87.0	44.7	58.5
<b>MPEG-7</b>			
PANORAMA + LRF	<b>87.2</b>	<b>65.5</b>	<b>75.9</b>
PANORAMA	87.2	61.8	73.1
2D/3D Hybrid	86.1	59.6	70.7
DESIRE	86.4	57.7	67.7
LF	80.2	51.7	61.9
SH-GEDT	83.7	50.3	59.4
<b>PSB</b>			
PANORAMA + LRF	75.2	<b>53.1</b>	<b>65.9</b>
PANORAMA	<b>75.3</b>	47.9	60.3
2D/3D Hybrid	74.2	47.3	60.6
DESIRE	65.8	40.4	51.3
LF	65.7	38.0	48.7
SH-GEDT	55.3	31.0	41.4
<b>ESB</b>			
PANORAMA + LRF	<b>87.0</b>	<b>49.9</b>	<b>65.8</b>
PANORAMA	86.5	49.4	64.1
2D/3D Hybrid	82.9	46.5	60.5
DESIRE	82.3	41.7	55.0
LF	82.0	40.4	53.9
SH-GEDT	80.3	40.1	53.6

**Fig. 14** Examples of queries within the PSB dataset and the corresponding top 5 retrieved models using the PANORAMA descriptor. The retrieved objects are ranked from left to right in decreasing order of similarity



This is a significant advantage of the proposed method as it can be rendered much more efficient by reducing its storage requirements and time complexity, with negligible cost in discriminative power.

Altogether, the PANORAMA descriptor is very efficient in computation and comparison time. On a standard contemporary machine, it takes less than one second to extract and pairwise comparison time is approximately 0.23 ms which enables real-time retrieval from large repositories. These are the average values over the total set of models from the datasets used in our evaluation.

In Fig. 14, we provide a few examples of queries and the corresponding top 5 retrieved models from the PSB dataset using the proposed PANORAMA descriptor.

## 6 Conclusions

In this paper, we presented a novel 3D shape descriptor called PANORAMA, that allows effective content-based 3D model retrieval exhibiting superior performance, compared to other state-of-the-art approaches. PANORAMA is based on a novel 3D shape representation that uses a set of panoramic views of the 3D model which are obtained by projecting the model to the lateral surfaces of a set of cylinders that are aligned with the model's principal axes. A panoramic view is particularly descriptive of an object's shape as it captures a large portion of the object that would

otherwise require multiple orthogonal projections from different viewpoints. It is also beneficial compared to spherical function-based representations as the underlying sampling is uniform in the Euclidean space.

The 2-dimensional parameterization of a cylindrical projection allows the application of a variety of 2D features used for 2D shape-matching to be directly applied in the context of 3D shape matching. In this paper we have used the 2D Discrete Fourier Transform together with the 2D Discrete Wavelet Transform and their combination enables very effective and efficient 3D object retrieval. Using only the wavelet features part of the descriptor, we can greatly increase efficiency in terms of storage requirements and time complexity, with negligible cost in discriminative power.

Moreover, we have used the PANORAMA descriptor to examine the application of local relevance feedback in the context of content-based 3D object retrieval, using the method that was proposed in Papadakis et al. (2008b). The results of the evaluation showed that this method can add a considerable gain in the overall retrieval performance and can be readily applied to other retrieval methods that exhibit high precision near the neighborhood of the query.

**Acknowledgements** This paper is part of the 03 EΔ 520 research project "3D Graphics search and Retrieval", implemented within the framework of the "Reinforcement Programme of Human Research Manpower" (PENED) and co-financed by National and Community Funds (20% from the Greek Ministry of Development-General Secretariat of Research and Technology and 80% from E.U.-European Social Fund).

## References

- Akbar, S., Kung, J., Wagner, R., & Prihatmanto, A. S. (2006). Multi-feature integration with relevance feedback on 3D model similarity retrieval. In *Proceedings of the international conference on information integration and web-based applications services* (pp. 77–86) 2006.
- Ankerst, M., Kastenmuller, G., Kriegel, H. P., & Seidl, T. (1999). Nearest neighbor classification in 3D protein databases. In *ISMB* (pp. 34–43) 1999.
- Atmosukarto, I., Kheng, W., & Huang, Z. (2005). Feature combination and relevance feedback for 3D model retrieval. In *Proceedings of the 11th int. conf. on multimedia modeling*, 2005.
- Bang, H., & Chen, T. (2002). Feature space warping: an approach to relevance feedback. In *Proceedings of the int. conf. on image processing*, 2002.
- Bustos, B., Keim, D., Saupe, D., Schreck, T., & Vranic, D. (2004). Automatic selection and combination of descriptors for effective 3D similarity search. In *IEEE sixth int. symp. on multimedia software engineering* (pp. 514–521) 2004.
- Bustos, B., Keim, D., Saupe, D., Schreck, T., & Vranic, D. V. (2005). Feature-based similarity search in 3d object databases. *ACM Computing Surveys*, 37(4), 345–387.
- Chaouch, M., & Verroust-Blondet, A. (2007). 3D model retrieval based on depth line descriptor. In *Multimedia and expo, 2007 IEEE international conference on* (pp. 599–602) 2007.
- Chen, D. Y., Tian, X. P., Shen, Y. T., & Ohyoung, M. (2003). On visual similarity based 3D model retrieval. In *Eurographics, computer graphics forum* (pp. 223–232) 2003.
- Cornea, N., Demirci, M., Silver, D., Shokoufandeh, A., Dickinson, S., & Kantor, P. (2005). 3D object retrieval using many-to-many matching of curve skeletons. In *Proceedings of the international conference on shape modeling and applications* (pp. 368–373) 2005.
- Crucianu, M., Ferecatu, M., & Boujemaa, N. (2004). *Relevance feedback for image retrieval: a short survey* (Technical Report). INRIA.
- Daras, P., Zarpalas, D., Tzovaras, D., & Strintzis, M. G. (2006). Efficient 3-D model search and retrieval using generalized 3-D radon transforms. *IEEE Transactions on Multimedia*, 8(1), 101–114.
- Elad, M., Tal, A., & Ar, S. (2001). Content based retrieval of vrml objects: an iterative and interactive approach. In *Proceedings of the 6th Eurographics workshop on multimedia*, 2001.
- Fang, R., Godill, A., Li, X., & Wagan, A. (2008). A new shape benchmark for 3D object retrieval. In *International symposium on advances in visual computing* (pp. 381–392) 2008.
- Funkhouser, T., & Shilane, P. (2006). Partial matching of 3D shapes with priority-driven search. In *Fourth Eurographics symposium on geometry processing* (pp. 131–142) 2006.
- Google SketchUp (2009). <http://sketchup.google.com/>.
- Hilaga, M., Shinagawa, Y., Kohmura, T., & Kunii, T. L. (2001). Topology matching for fully automatic similarity estimation of 3D shapes. In *Proceedings of the 28th annual conference on computer graphics and interactive techniques* (pp. 203–212) 2001.
- Igarashi, T., Matsuoka, S., & Tanaka, H. (1999). Teddy: a sketching interface for 3D freeform design. In *SIGGRAPH '99: Proceedings of the 26th annual conference on computer graphics and interactive techniques* (pp. 409–416) 1999.
- Jayanti, S., Kalyanaraman, Y., Iyer, N., & Ramani, K. (2006). Developing an engineering shape benchmark for cad models. *Computer-Aided Design*, 38, 939–953.
- Kazhdan, M., Funkhouser, T., & Rusinkiewicz, S. (2003). Rotation invariant spherical harmonic representation of 3D shape descriptors. In *Eurographics/ACM SIGGRAPH symposium on geometry processing* (pp. 156–164) 2003.
- Kim, D. H., Park, I. K., Yun, I. D., & Lee, S. U. (2004). A new mpeg-7 standard: perceptual 3-d shape descriptor. In *Pacific conference in multimedia* (pp. 238–245) 2004.
- Kokare, M., Chatterji, B., & Biswas, P. (2003). Comparison of similarity metrics for texture image retrieval. In *TENCON 2003. Conference on convergent technologies for Asia-Pacific region* (Vol. 2, pp. 571–575) 2003.
- Leifman, G., Meir, R., & Tal, A. (2005). Semantic-oriented 3d shape retrieval using relevance feedback. *Visual Computer*, 21(8–10), 865–875.
- Lou, K., Jayanti, S., Iyer, N., Kalyanaraman, Y., & Ramani, K. (2003). A reconfigurable 3D engineering shape search system part II: database indexing, retrieval and clustering. In *Proceedings of the 23rd computers and information in engineering conference*, 2003.
- Novotni, M., & Klein, R. (2003). 3d Zernike descriptors for content based shape retrieval. In *8th ACM symposium on solid modeling and applications* (pp. 216–225) 2003.
- Novotni, M., Park, G. J., Wessel, R., & Klein, R. (2005). Evaluation of kernel based methods for relevance feedback in 3D shape retrieval. In *Proceedings of the 4th int. workshop on content-based multimedia indexing*, 2005.
- Yu, M., Atmosukarto, I., Leow, W. K., Huang, Z., & Xu, R. (2003). 3D model retrieval with morphing-based geometric and topological feature maps. In *5th ACM SIGMM international workshop on multimedia information retrieval* (pp. 656–661) 2003.
- Ohbuchi, R., Nakazawa, M., & Takei, T. (2003). Retrieving 3D shapes based on their appearance. In *5th. ACM SIGMM international workshop on multimedia information retrieval* (pp. 39–45) 2003.
- Ohbuchi, R., Minamitani, T., & Takei, T. (2005). Shape-similarity search of 3D models by using enhanced shape functions. *Computer Applications in Technology*, 23(2–4), 70–85.
- Olsen, L., Samavati, F., Sousa, M. C., & Jorge, J. (1999). A taxonomy of modeling techniques using sketch-based interfaces. In *Eurographics* (pp. 39–57) 1999.
- Osada, R., Funkhouser, T., Chazelle, B., & Dobkin, D. (2001). Matching 3D models with shape distributions. In *Int. conf. on shape modeling and applications* (pp. 154–166) 2001.
- Osada, R., Funkhouser, T., Chazelle, B., & Dobkin, D. (2002). Shape distributions. *ACM Transactions on Graphics*, 21(4), 807–832.
- Papadakis, P., Pratikakis, I., Perantonis, S., & Theoharis, T. (2007). Efficient 3D shape matching and retrieval using a concrete radialized spherical projection representation. *Pattern Recognition*, 40(9), 2437–2452.
- Papadakis, P., Pratikakis, I., Theoharis, T., Passalis, G., & Perantonis, S. (2008a). 3D object retrieval using an efficient and compact hybrid shape descriptor. In *Eurographics Workshop on 3D object retrieval*, 2008.
- Papadakis, P., Pratikakis, I., Trafalis, T., Theoharis, T., & Perantonis, S. (2008b). Relevance feedback in content-based 3d object retrieval: a comparative study. *Computer-Aided Design and Applications*, 5(5), 753–763.
- Passalis, G., Theoharis, T., & Kakadiaris, I. A. (2006). Ptk: a novel depth buffer-based shape descriptor for three-dimensional object retrieval. *Visual Computer*, 23(1), 5–14.
- Ricard, J., Coeurjolly, D., & Baskurt, A. (2005). Generalization of angular radial transform for 2D and 3D shape retrieval. *Pattern Recognition Letters*, 26, 2174–2186.
- Rochio, J. (1971). *Relevance feedback in information retrieval the SMART retrieval system—experiments in automatic document processing* (Vol. 14, pp. 313–323). Englewood Cliffs: Prentice Hall
- Ruthven, I., & Lalmas, M. (2003). A survey on the use of relevance feedback for information access systems. *Knowl. Eng. Rev.*, 18(2), 95–145.

- Schmidt, R., Wyvill, B., Sousa, M. C., & Jorge, J. A. (2005). Shapeshop: sketch-based solid modeling with blobtrees. In *Proceedings of the 2nd Eurographics workshop on sketchbased interfaces and modeling*, 2005.
- Shih, J. L., Lee, C. H., & Wang, J. (2007). A new 3D model retrieval approach based on the elevation descriptor. *Pattern Recognition*, 40(1), 283–295.
- Shilane, P., Min, P., Kazhdan, M., & Funkhouser, T. (2004). The Princeton shape benchmark. In *Shape modeling international* (pp. 167–178) 2004.
- Song, J., & Golshani, F. (2003). Shape-based 3D model retrieval. In *15th Int. conf. on tools with artificial intelligence* (pp. 636–640) 2003.
- Sundar, H., Silver, D., Gagvani, N., & Dickinson, S. (2003a). Skeleton based shape matching and retrieval. In *Proceedings of the shape modeling international* (pp. 130–139) 2003.
- Sundar, H., Silver, D., Gagvani, N., & Dickinson, S. (2003b). 3D shape matching with 3D shape contexts. In *Seventh central European seminar on computer graphics*, 2003.
- Theoharis, T., Papaioannou, G., Platis, N., & Patrikalakis, N. (2008). *Graphics and visualization: principles and algorithms*. Wellesley: AK Peters.
- Veltkamp, R., & ter Haar, F. (2007). *Shrec2007 3d shape retrieval contest* (Technical Report). Department of information and computing sciences, Utrecht university, 2007.
- Vranic, D. V. (2004). *3D model retrieval*. PhD thesis, University of Leipzig (2004).
- Vranic, D. (2005). Desire: a composite 3D-shape descriptor. In *IEEE international conference on multimedia and expo*, 2005.
- Zaharia, T., & Petreux, F. (2001). 3D shape-based retrieval within the mpeg-7 framework. In *SPIE conf. on nonlinear image processing and pattern analysis XII* (pp. 133–145) 2001.
- Zaharia, T., & Preteux, F. (2002). Shape-based retrieval of 3D mesh models (pp. 437–440).
- Zarpalas, D., Daras, P., Axenopoulos, A., Tzovaras, D., & Strintzis, M. G. (2007). 3D model search and retrieval using the spherical trace transform. *EURASIP Journal on Advances in Signal Processing* Article ID 23912, 14 pages.
- Zhang, J., Siddiqi, K., Macrini, D., Shokoufandeh, A., & Dickinson, S. (2005). Retrieving articulated 3-d models using medial surfaces and their graph spectra. In *International workshop on energy minimization methods in computer vision and pattern recognition*, 2005.
- Zhou, X. S., & Huang, T. S. (2001). Exploring the nature and variants of relevance feedback. In *Proceedings of the IEEE workshop on content-based access of image and video libraries* (pp. 94–101) 2001.