Int J Comput Vis (2010) 89: 229–247
DOI 10.1007/s11263-009-0277-2

A 3D Shape Retrieval Framework Supporting Multimodal Queries

Petros Daras · Apostolos Axenopoulos

Received: 30 September 2008 / Accepted: 21 July 2009 / Published online: 30 July 2009

Abstract This paper presents a unified framework for 3D shape retrieval. The method supports multimodal queries (2D images, sketches, 3D objects) by introducing a novel view-based approach able to handle the different types of multimedia data. More specifically, a set of 2D images (multi-views) are automatically generated from a 3D object, by taking views from uniformly distributed viewpoints. For each image, a set of 2D rotation-invariant shape descriptors is produced. The global shape similarity between two 3D models is achieved by applying a novel matching scheme, which effectively combines the information extracted from the multi-view representation. The experimental results prove that the proposed method demonstrates superior performance over other well-known state-of-the-art approaches.

Keywords 3D object retrieval · Multi-views · Multimodal queries · Image to 3D object · Sketch to 3D object

1 Introduction

It is widely known that human beings think with words. These words are harmonically combined creating sentences which are used to describe feelings, concepts, opinions, actions, desires, etc. In the special case where an object is to be described, the aforementioned sequence of actions which takes place in our brain is a little bit different due to the visual form of the object. Dr. Fodor in 1983 who introduced "the modularity of mind" (Fodor 1983), supports that our brain follows a hierarchical way of thinking starting from the simplest thought and ending at the most advanced one. With simple words this theory implies that if, for example, we want to describe a red rose in a bunch of flowers, the simplest "image" coming to our mind is ... image! To be more specific, it is a Three-Dimensional (3D) image, which clearly consists of all the information we want to describe and which is derived from the experience of seeing a red rose. In other words, a 3D object contains all the ground truth of a physical object, while its 2D views provide only a subset of the 3D object, an abstraction of the real world.

3D models have nowadays become ubiquitous for applications such as games (Bustos et al. 2005; Real-time 3D models), Computer-Aided Design (CAD) (Jayanti et al. 2006), molecular biology (Daras et al. 2006a; Tsatsaias et al. 2007), cultural heritage (Goodall et al. 2004), etc. The technology innovation in 3D scanners and computer-aided modeling software make it possible to easily construct complete 3D geometry models with relatively low cost and time, which in turn has triggered the rapid enlargement of 3D shape repositories. The latter, along with the explosion of the World Wide Web (WWW), has lead to research in the area of 3D content-based search and retrieval (Iyer et al. 2005; Tangelder and Veltkamp 2004; Bustos et al. 2007) using as query text, sketch and/or 3D object(s).

Those who are searching on the WWW are familiar with text-based search engines: by giving a few keywords it is possible to find related sites on the Internet. Keywords constitute a good descriptor in this particular case because the audiovisual information is organized for and around text. The situation is different for image and 3D object databases,

P. Daras (✉) · A. Axenopoulos
Informatics and Telematics Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece
e-mail: daras@iti.gr

A. Axenopoulos
e-mail: axenop@iti.gr

since a keyword description for an image/3D object is language, culture and context dependent, resulting in great lack of uniformity. On the other hand, a query-by-content (example) approach is much more simple and efficient. 2D images and 3D models can be added to a database as they are, without ordering them.

In this paper, we propose a unified framework for 3D shape retrieval. The framework supports multimodal queries (either sketches drawn by a user, or 2D images captured by a user, or 3D objects) by introducing a 2D multi-view-based approach able to handle the different types of multimedia data. More specifically, multiple 2D images are generated from a 3D object, by taking views from uniformly distributed viewpoints. For each image, a set of 2D rotation-invariant shape descriptors, based on the Polar-Fourier Transform, Zernike Moments and Krawtchouk Moments, is produced. The global shape similarity between two 3D models is achieved by applying a novel matching scheme, which effectively combines the information extracted from the multi-view representation. Furthermore, the method provides efficient search and retrieval capabilities using only a 2D image or a sketch as a query, when an input 3D model is not available. Finally, the proposed Compact Multi-View Descriptor (CMVD) is combined with a well-known transform-based method, the Spherical Trace Transform (Zarpalas et al. 2007), resulting in significantly increased performance.

1.1 Related Work

3D object retrieval is a relatively new and very challenging research field and a major effort of the research community has been devoted to the formulation of accurate and efficient 3D object search and retrieval algorithms. The main problems that pose obstacles in the efficiency of the existing approaches are the following: (i) the 3D object's degeneracies (e.g. holes, missing polygons, hidden polygons), (ii) the 3D object's pose normalization, (iii) invariance to shape representations, (iv) invariance to articulation or global deformation, (v) the trade-off between the time needed for the extraction (which heavily depends on a 3D object's Level-of-Detail) and matching of 3D objects' descriptors and the retrieval accuracy of a method.

The first problem is usually tackled successfully by applying a triangulation algorithm (e.g. Delaunnay triangulation) or a hole filling algorithm (Weyrich et al. 2004). Pose normalization implies invariance with respect to rotation, scaling and translation of a 3D object. Translation normalization is usually achieved by calculating the center of mass while scaling normalization uses the root of the average square radius. Another approach for scaling and translation normalization is the smallest enclosing sphere (Fischer and Gartner 2004). In order to achieve rotation normalization two widely acceptable solutions have been presented in the literature, namely the rotation normalization of the 3D object in a pre-processing step and the natively rotation invariant description of the 3D object. Both approaches present major advantages and serious drawbacks: Firstly, the vast majority of the utilized rotation normalization approaches are based on the PCA (e.g. Continuous PCA, Vranic 2003). Although algorithms that utilize pose normalization using PCA usually result in descriptors with higher discriminative power, some similar objects are not usually normalized in a similar manner (Vranic 2004). Several natively rotation invariant descriptors have been proposed so far, including the Spherical Harmonic Descriptor (SHD) (Kazhdan et al. 2003), the Light Field Descriptor (LFD) (Chen et al. 2003) as well as various histogram-based descriptors, such as those proposed by Kriegel et al. (2003), Ankerst et al. (1999), the D2 and other "Shape distributions" (Osada et al. 2002), the Surflet Pair Relation Histograms (SPRH) (Wahl et al. 2003) and the Enhanced Shape Functions by Ohbuchi et al. (2003a). Natively rotation invariant object description (Kazhdan et al. 2003) usually involve an integration-like technique which leads to inadequately discriminant descriptors (Vranic 2003). Concerning the invariance to shape representations, most of the existing methods work well either with polygonal meshes or polygon soups, while methods which rely on the topology of an object demand certain shape representation (e.g. watertight models) (Hilaga et al. 2001; Tung and Schmitt 2005; Mademlis et al. 2008a). The invariance to articulation is a hot research problem which has not been widely addressed so far and requires the extraction of local descriptors. A solution to the latter problem might also lead to more efficient partial matching algorithms.

The existing 3D object retrieval methods can be classified into four main categories: histogram-based, transform-based, graph-based, view-based and, finally, combinations of the above. In the first category the methods which have been proposed so far use histograms where the extracted local or global features of a 3D object are integrated. In this sense, Ohbuchi et al. (2002) employ shape histograms that are discretely parameterized along the principal axes of inertia of the model. Osada et al. (2001, 2002) introduce and compare shape distributions, which measure properties based on distance, angle, area, and volume measurements between random surface points. They evaluate the similarity between the objects using a metric that measures distances between distributions. Liu et al. (2006) propose the generalize shape descriptor (GSD) where a 3D histogram counts the number of specific local shape pairs at certain distances. Ankerst et al. (1999) introduce a 3D shape similarity model by defining two major ingredients: the shape histograms as

an intuitive and discrete representation of complex spatial objects and an adaptable similarity distance function for the shape histograms that may take into account small shifts and rotations by using quadratic forms. In Horn (1984), the extended Gaussian images (EGI) are introduced, where the surface normal orientation is mapped on a sphere, namely the Gaussian sphere. The EGI is obtained by having each triangle vote on the bin corresponding to its normal direction, with a weight equal to the area of the triangle. In Kang and Ikeuchi (1993), EGI has been generalized to the complex extended Gaussian image (CEGI), which stores for each bin also the normal distance of the surface points to the origin. The aforementioned methods are, in general, ease to implement but usually they are not discriminating enough to make subtle distinctions between classes of shapes.

Transform-based methods are employed either on the surface or on the volume of a 3D model. In Vranic and Saupe (2002, 2001), a method where the descriptor vector is obtained by forming a complex function on the sphere, is presented. Then, spherical harmonics analysis is used to form the rotation invariant descriptor vector. In Kazhdan et al. (2003), the Spherical Harmonic Representation is proposed which transforms rotation dependent descriptors into rotation invariant ones. In Novotni and Klein (2003), the theoretical framework for the 3D Zernike moments (Canterakis 1999) is extended and applied for 3D content-based search and retrieval. These are computed as a projection of the function which defines the object, onto a set of orthonormal functions within the unit ball. The 3D Zernike descriptors are natively invariant under rotation. In Papadakis et al. (2007), the authors apply PCA on the face normals of a model. Then, the 3D model is decomposed into a set of spherical functions which represents not only the intersections of the corresponding surface with rays emanating from the origin but also points in the direction of each ray which are closer to the origin than the furthest intersection point. All the presented approaches are extracting sophisticated descriptors that exploit some significant properties (e.g. rotation invariance transformation Kazhdan et al. 2003; Novotni and Klein 2003 and highly discriminative descriptor sets Vranic 2004). In most of the aforementioned approaches, which are directly applied to the 3D geometry of the objects, the retrieval efficiency may be decreased when they deal with non-perfect polygon meshes. In order to avoid the latter, voxel-based representation is preferred, which is more robust to mesh degeneracies and levels of detail. In Kazhdan et al. (2003), for example, the method scan-converts faces into voxels before applying the spherical harmonic decomposition. Alternatively, the following methods rely on a voxel-based representation of the 3D object's volume: in Daras et al. (2006b), a voxel-based 3D search and retrieval method based on the Generalized Radon Transform (GRT) is proposed, while in Zarpalas et al. (2007), the Spherical Trace Transform (STT) is presented. The STT is among the best algorithms which have ever been presented (in terms of retrieval accuracy). In general, the transform-based methods have high retrieval accuracy but usually, pose invariance is achieved by discarding the "phase" of the transform coefficients at the expense of some shape information.

One possible solution to the aforementioned problem is the graph-based methods, which produce descriptors fundamentally different from other vector-based descriptors. They are more elaborated and complex, in general harder to obtain; but they have the potential of encoding geometrical and topological shape properties in a more faithful and intuitive manner. In Hilaga et al. (2001) a technique, called Topology Matching, is introduced, which calculates the similarity between polyhedral models by comparing multiresolutional Reeb graphs (MRGs). Based on the idea of MRG matching (Hilaga et al. 2001), Chen and Ouhyoung (2002) propose a 3D model retrieval system, where a preprocessing step has been added before the Reeb Graph extraction in order to accelerate the graph-matching and retrieval processes. The work in Hilaga et al. (2001) has been further extended in Tung and Schmitt (2004, 2005), where the Reeb graph is augmented with geometrical attributes leading to the creation of a flexible multiresolutional representation, called an augmented Reeb graph. In Katz and Tal (2003), a hierarchical mesh decomposition algorithm is proposed. The algorithm computes a 3D object mesh decomposition, which generally refers to segmentation at regions of deep concavities. A mesh decomposition method, similar to Katz and Tal (2003), is presented in Tal and Zuckerberger (2006). This decomposition is represented as an attributed graph, which is considered the signature of the object. The method is less computationally expensive than the one in Katz and Tal (2003), which makes it appropriate for search and retrieval in large databases. In Mademlis et al. (2008a), a method which combines topological and geometrical information is proposed, which is invariant to geometric transformations of a 3D object, as well as to the different poses of articulated objects. The drawbacks of the graph-based methods are that it is difficult to implement them, they do not generalize easily to all 3D shape representation formats and they require dedicated matching schemes.

2D view-based methods consider the 3D shape as a collection of 2D projections taken either from canonical or non-canonical viewpoints. Each projection is then described by standard 2D image descriptors like Fourier descriptors or Zernike moments. They rely on the assumption that as the 3D models are completely given, projections can be produced in a controlled manner so that nuisance effects of occlusion (except self-occlusions), clutter or affine deformations are avoided. In 2002, Mahmoudi and Daoudi (2002) introduce a method for indexing 3D models by using a set of

7 characteristic views (CVs), three principals and four secondaries. An object contour-based shape descriptor based on the Curvature Scale Space (CSS) representation of the contour is, then, generated for each CV. In Vranic (2004), two view-based descriptors, the Depth-Buffer and the Silhouette Descriptor, are presented. After a pose normalization using Continuous PCA (CPCA), depth images and silhouettes are extracted from the faces of a bounding cube. For each 2D image, a set of 2D Fourier coefficients are extracted, producing the descriptor vectors. Similar to the Depth-Buffer descriptor (Vranic 2004), the Elevation descriptor, presented in Shih et al. (2007), takes depth images from the six faces of a bounding cube. Since the method does not utilize any pose normalization technique, it takes several combinations of pairs of views during the matching procedure. The method presented in Chen et al. (2003) avoids the pose normalization problem by taking multiple sets of silhouette images from 60 different rotations of the 3D object. As 2D descriptors per view, Zernike moments and Fourier descriptors are used. Chaouch and Verroust-Blondet (2006), propose a novel framework to provide more accurate descriptions of 3D models by associating a relevance index to 2D images (either depth-buffer or silhouette images). It is based on the notion that some 2D views of the 3D object are more significant than others, therefore they should be assigned different weight during dissimilarity computation. Ohbuchi et al. (2008) recently proposed a view-based 3D model retrieval method based on multi-scale local visual features. The features are extracted from 2D range images of the model. For each range image, a set of 2D multi-scale local visual features is computed by using the Scale Invariant Feature Transform (SIFT) (Lowe 2004) algorithm. Another novel approach is proposed by Napoleon et al. (2008). The method is based on a set of 2D multi-views and utilizes CPCA for pose normalization. The method introduces a Multi-scale Contour Representation for each multi-view, which describes the convexities and concavities of the silhouettes at different scale levels. The method presents noticeable results in SHREC'09 Generic and Structural Shape Retrieval datasets.

The 2D view-based methods have the advantages of being highly discriminative, can work for articulated objects, can be effective for partial matching and can also be beneficial for 2D sketch-based and 2D image-based queries. Their only drawback is that they discard valuable 3D information (due to the self-occlusion). In order to avoid the latter, a combination of a 2D view-based method with a 3D-based method (such as transform-based) is expected to achieve much higher performance.

In all categories of 3D object retrieval methods described above, search is performed by using as query a 3D object. However, in a real-world 3D search and retrieval scenario, an input 3D object may not always be available. On the other hand, a 2D image or a hand-drawn sketch is easier to be provided as query by the user. Consequently, a 3D object retrieval mechanism able to support multimodal queries is clearly much more competitive than others which support only 3D queries. Although 3D search and retrieval with multimodal support is a challenging research issue, only few approaches have been proposed so far. In Funkhouser et al. (2003), the proposed "Princeton 3D search engine" supports 2D sketch, 3D sketch, 3D model and text as queries. In Chen et al. (2003), retrieval of 3D objects from 2D sketches is achieved by matching the 2D descriptors of the sketch with the corresponding descriptors of the 100 2D silhouettes of the 3D object. Recently, Ansary et al. (2007) proposed a framework for 3D model retrieval using 2D (still) photographic images, sketches, as well as 3D models. The search engine is based on the Adaptive Views Clustering (AVC) algorithm (Filali Ansary et al. 2007), which uses statistical model distribution scores to select the optimal number of views to characterize a 3D model.

1.2 Proposed Work

The proposed method can be summarized in the block diagram presented in Fig. 1. The input 3D object is a triangulated mesh, in one of the common 3D file formats (VRML, OFF, 3DS, etc.). As a first step, a pose estimation takes place, which includes translation, scaling and rotation of the object. After the pre-processing step, a set of 18 2-dimensional views, taken from the vertices of a bounding 32-hedron is extracted. Both binary (black/white) and depth images are generated. In each of the extracted 2D images, a set of 2D functionals is applied, resulting in a descriptor vector for each view.

The proposed Compact Multi-View Descriptor (CMVD) belongs to the category of the 2D view-based approaches and, thus, holds the following advantages:

Robustness to Object Degeneracies, Holes, Missing Polygons. Several feature extraction methods, proposed so far, are applied to the 3D objects' surface, which in turn requires perfect triangulation of the 3D meshes (e.g. watertight objects). These techniques can work only with a limited number of objects, since the majority of models publicly available contain holes, missing polygons and other design defects. On the other hand, the view-based methods can be applied irrespectively to the object's triangulation. The only potential weakness of the view-based methods is that the use of a bounding polyhedron for scale normalization may cause undesirable scaling of the 3D object, in cases when the existence of outliers changes significantly the size of the original object.

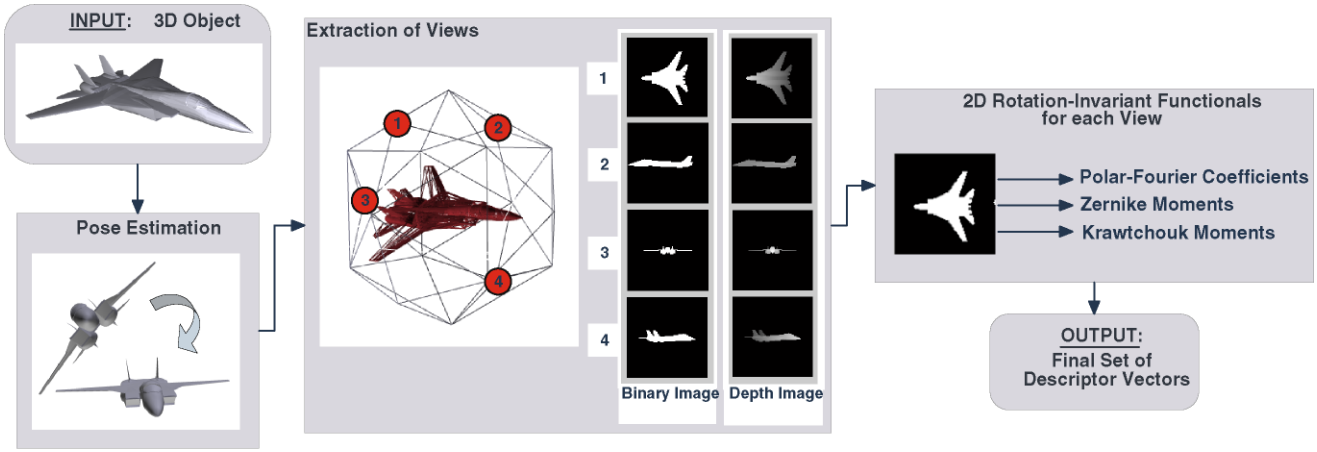

Fig. 1 Block diagram of proposed descriptor extraction method

Robustness with Respect to the to Objects' Level of Detail (LoD). The proposed descriptor extraction method is applied to a set of 2D images with the same pixel resolution. Therefore, the extracted descriptors are independent of the 3D object's complexity.

Unified Framework. Similarly to partial matching, the proposed method can provide 3D model retrieval capabilities even when the input 3D model is not available. A variety of queries, such as 2D images, hand-drawn sketches, 3D objects, are supported and can alternatively be used, which results in a unified framework for 3D object retrieval, beyond the traditional 3D search engines. Additionally, the image-based search capabilities can broaden the application areas and include real-life mobile application scenarios.

High Discriminative Power. In general, methods in which 3D matching is based on multi-views comparison instead of global shape geometry comparison, have proven to be more efficient in terms of retrieval accuracy. The performance of the proposed method confirms the above conclusion, as it will be presented in the experimental results section.

Suitable for Partial Matching and Articulated Objects. In general, a view-based method does not directly extract global 3D shape descriptors; it combines descriptors taken from specific parts (multiple views) of a 3D object in order to describe its global shape. If, instead of a 3D object, only a part of it is available as query, a view-based method could provide a solution for partial 3D object retrieval by efficiently matching only a subset of views of a 3D object. However, designing a matching framework suitable for partial 3D shape retrieval is not so straightforward and it is still an open research issue. Similarly, a view-based method could be appropriately modified in order to deal with articulated objects.

Despite the numerous common advantages, the proposed approach introduces the following novel features:

A Set of Three Complementary 2D Functionals: The proposed method combines the characteristics of three different 2D functionals to describe the shape of a single view: Polar-Fourier Coefficients, Zernike Moments and Krawtchouk Moments. The first two have been already used in similar approaches, while Krawtchouk Moments are used for the first time in this paper to deal with view-based 3D object retrieval. The reason we choose to use these functionals is threefold: (a) they are able to store image information with minimal information redundancy, (b) they have the important property of being rotation invariant and (c) they seem to be among the most powerful methods for 2D shape recognition as it is evident by Belkasim et al. (1991), Khotanzad and Hong (1990), Zhang and Lu (2002), Yap et al. (2003) and was also proved by our experimental results. However, it should be stated that the proposed framework is modular enough to allow for integration of other high discriminant transforms and moment invariants that hold the abovementioned desired properties. Having such a highly discriminative 2D descriptor makes the proposed method appropriate for 3D shape retrieval tasks where only a 2D image or sketch is given as query.

Efficient Selection of the Multi-Views: Several approaches have been already proposed towards the selection of the optimal set of the 2D views, such as the 6 vertices of an 8-hedron (Vranic 2004), the 20 vertices of a 12-hedron (Chen et al. 2003) or even the 42 vertices of an 80-hedron (Ohbuchi et al. 2008). The 18-views representation proposed in this paper is used for the first time, it provides a non-redundant set of views and it has the following advantage over the 20-view and 42-view representations: the 18 views extracted from a 32-hedron are symmetric with respect to 90 degrees rotations about the three orthogonal axes (x, y, z),

which makes this representation appropriate for the matching scheme described below.

Efficient Multi-View Matching Scheme: In most of the view-based methods proposed so far, similarity matching is based on pairwise comparison of 2D views. In Ohbuchi et al. (2003b), an all-to-all comparison of 42-to-42 views (1764 comparisons) is performed. The drawback of this matching method is that it does not take into account the correspondence between neighboring views, thus, it discards valuable 3D information. This problem is overcome in the Light Field Descriptor (LFD) (Chen et al. 2003), where, instead of matching single pairs of views, a matching of a sequence of corresponding views is performed. In order to be invariant to rotations of the 3D objects, the method applies this matching scheme for all 5460 possible rotations of the model. This increases significantly the similarity matching time and makes it inappropriate for on-line retrieval applications. In Vranic (2004), the author applies Continuous PCA (CPCA) for pose normalization and matches the sequence of 6 views taken from the vertices of a regular 8-hedron only once, without any further rotation of the model. However, as it will be explained in the sequel, rotation normalization based on PCA and its variations (VCA, CPCA) is not always accurate. The matching method proposed in this paper involves rotation of the 3D object 24 times in 90-degrees intervals around the three orthogonal axes. Thus, exhaustive matching of all possible rotations as in the case of LFD (Chen et al. 2003) is avoided, while, at the same time, the method compensates for the inherent limitations of PCA. This might be more time consuming than the method in Vranic (2004) (not significantly), but it provides higher retrieval accuracy.

Consequently, the method proposed in this paper demonstrates higher retrieval accuracy than the view-based methods presented in Chen et al. (2003), Ohbuchi et al. (2008) and outperforms or is competitive with the best 3D shape retrieval methods presented so far. The performance can be further improved if the advantages of the proposed method are combined with an efficient transform-based method, since the latter extracts different type of information from 3D content. Among the state-of-the-art transform-based methods, the Spherical Trace Transform (STT), presented in Zarpalas et al. (2007), is combined with the proposed method leading to significantly improved results.

The rest of the paper is organized as follows: Sect. 2 analyzes the descriptor extraction procedure, which consists of a pose estimation step, a views generation step and the computation of 2D functionals. In Sect. 3, the shape matching framework for both 2D/3D and 3D/3D matching is described. Experimental results evaluating the proposed method and comparing it with other methods are presented in Sect. 4. Finally, conclusions are drawn in Sect. 5.

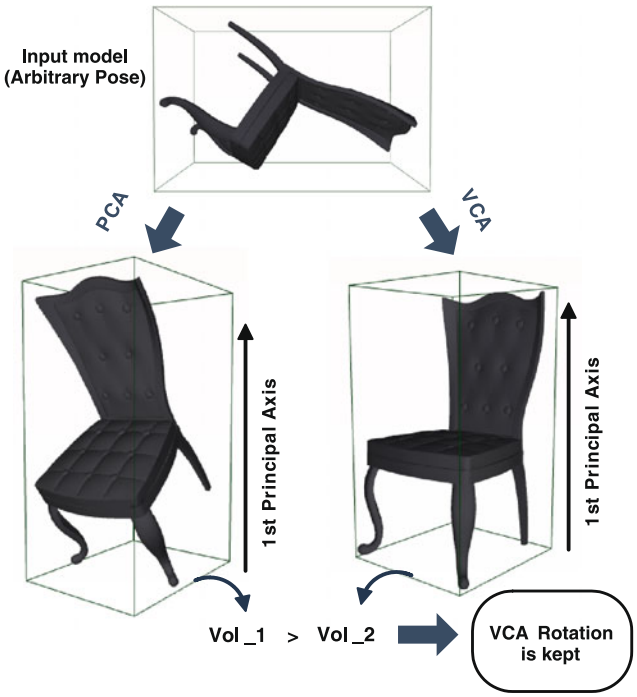

Fig. 2 Rotation normalization using both PCA and VCA. The rotation that produces the model with the smallest bounding volume is selected

2 Descriptor Extraction Method

2.1 Pose Estimation

The Pose Estimation procedure initially involves the translation and scaling of the 3D object. The model is translated so that the center of mass coincides with the center of the coordinate system and scaled in order to lie within a bounding sphere of radius 1.

After translation and scaling, a rotation estimation step is required, since the 3D object may have an arbitrary orientation. In order to achieve the best possible result, a combination of the two dominant rotation estimation methods, the Principal Component Analysis (PCA) (Vranic et al. 2001) and the Visual Contact Area (VCA) (Pu and Ramani 2005), which have been proposed so far in the literature, is utilized. The VCA method achieves more accurate rotation estimation results than PCA when the 3D objects are composed of large flat areas. Otherwise, PCA produces better results than VCA. In this paper for every model, rotation normalization is estimated using both PCA and VCA. Then, the volumes of the bounding boxes parallel to principal axes are computed and the rotated object with the minimum bounded volume is chosen. In Fig. 2, an example of the combined use of PCA and VCA is illustrated. Both PCA and VCA rotation normalization methods are applied to the arbitrarily rotated input model. It is obvious from the rotated models that, in this case, VCA achieves better rotation estimation. Since the

bounding volume of the VCA-rotated model is smaller than that of the PCA-rotated model, the former is chosen.

The proposed rotation estimation framework leads to the automatic detection of the model's three principal axes with a quite satisfying level of success. However, it does not provide information about the orientation of the principal axis. Therefore, for each of the three principal axes, there are two possible orientations, which results in 8 different alignments of the model. Taking also into account the fact that the first principal axis may not always be successfully selected among the three principal axes, this leads to a set of $3 \times 8 = 24$ different alignments.

It must be noted that the proposed rotation estimation method is introduced to deal with the arbitrary rotations of the model and not to overcome the inherent limitations of PCA and VCA in identifying the first principal axis and estimating the model's orientation. The problem of having 24 possible alignments is overcome by appropriately selecting the set of 2D views as well as by introducing an efficient matching method, which will be elaborated in the following sections.

2.2 A Set of Uniformly Distributed Views

The proposed method is based on the matching of multiple 2D views, which can be extracted from a 3D object by selecting a set of different viewpoints. In order to be uniformly distributed, the viewpoints are chosen to lie at the vertices of a regular polyhedron. The type of the polyhedron and the level of tessellation need to be carefully considered in order to provide the optimal solution. As mentioned in the previous Sect. 2.1, the rotation estimation method may not effectively identify the first principal axis among the three principal axes. Thus, in order to compensate for this inherent limitation of pose normalization, during similarity matching between two models, the second model has to be rotated 24 times in 90 degrees intervals around the three principal axes. This requires also 24 different sets of multi-views to be extracted. This computational burden can be significantly decreased by selecting as a base polyhedron a regular 8-hedron or one of its derivatives (32-hedron, 128-hedron, etc.). The advantage of this family of polyhedra is that they are symmetric with respect to the three Cartesian axes (x, y and z). By taking views from the vertices of an 8-hedron (or a 32-hedron, 128-hedron, etc.), a 90-degree rotation about one of the three Cartesian axes will not generate new views, it will just change the order of the existing ones. Consequently, in the pre-processing step, the set of multi-views of the object needs to be extracted only once, which results in a compact representation. On the other hand, if another type of polyhedron is chosen, such as 12-hedron (Chen et al. 2003), 20-hedron or 80-hedron (Ohbuchi et al. 2008), several sets of multi-views need to be extracted.

Another important aspect regarding the optimal selection of the base polyhedron is the number of views. In Depth-Buffer and Silhouette descriptors method (Vranic 2004), views are taken from the 6 vertices of a regular 8-hedron. It has been already proven in the literature (Lindstrom and Turk 2000; Huber and Hebert 2001) that 15 to 20 views can roughly represent the shape of a 3D model. More specifically, the method presented in Lindstrom and Turk (2000) uses 15 to 20 views to reconstruct real-world objects. In Sect. 4, the performance of the CMVD descriptor using both an 8-hedron and a 32-hedron was evaluated in Princeton Shape Benchmark Database (Shilane et al. 2004). It was found that the 18 views of the 32-hedron achieve better retrieval accuracy comparing with the 6 views of the 8-hedron, while the compactness with respect to the number of views is kept. Based on the above, the 32-hedron was eventually selected.

In order to render the multi-view images, the camera viewpoints are placed at the 18 vertices of the 32-hedron. The image rendering uses orthographic projection, i.e. views are taken from planes tangential to the 32-hedron at each viewpoint, as opposed to perspective projection, where views are taken directly from the viewpoint. Two 2D image types are available:

Binary Images: the rendered images are only silhouettes, where the pixel values are 1 if the pixel lies inside the model's 2D view and 0 otherwise.

Depth Images: the pixel intensities are proportional to the distance of the 3D object from each sample point of the corresponding tangential plane.

Although binary images provide an efficient and robust representation of a 2D view, depth images contain more information and produce better retrieval results, if appropriately exploited.

2.3 Computing 2D Functionals on each View

The set of uniformly distributed views, described above, consists of 2D binary images and depth images. In each image, three rotation-invariant functionals are applied in order to produce the final set of descriptors per view: (a) the 2D Polar-Fourier transform, (b) 2D Zernike Moments and (c) 2D Krawtchouk Moments. Each functional describes different 2D geometric properties and each one's role is complementary for the description of the 2D view.

Let $f_t(i, j)$ be the 2D image, where $i, j = 0, \ldots, N-1$, $N \times N$ the size of the image, $t = 1, \ldots, N_V$ and N_V the total number of views. The values of $f_t(i, j)$ are either 0 or 1, for the binary images, while in the case of depth images, the values can be any real number between 0 and 1.

2.3.1 2D Polar-Fourier Transform

The Discrete Fourier Transform (DFT) is computed for each $f_t(i, j)$:

$$FT(k,m) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} f_t(i,j)\exp\left(-\hat{j}\left(\frac{2\pi ik}{N}+\frac{2\pi jm}{N}\right)\right),$$
$$k, m = 0, \ldots, N-1. \quad (1)$$

In the DFT, shifts in the spatial domain cause corresponding linear shifts in the phase component:

$$FT_t(k,m)\exp[-\hat{j}(ak+bm)] \leftrightarrow f_t(i+a, j+b). \quad (2)$$

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

$$r_{ij} = \sqrt{(c_1 i + c_2)^2 + (c_1 j + c_2)^2},$$
$$\xi_{ij} = \tan^{-1}\left(\frac{c_1 j + c2}{c_1 i + c_2}\right), \quad (3)$$
$$c_1 = \frac{\sqrt{2}}{N-1}\cdot r_{max}, \qquad c_2 = -\frac{1}{\sqrt{2}}\cdot r_{max},$$
$$i, j = 0, \ldots, N-1.$$

(1) becomes:

$$FT_t(k,m) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} f_t(r_{ij}, \xi_{ij})\exp(-\hat{j}(kr_{ij} + m\xi_{ij})) \quad (4)$$

and rotation is converted to a circular translation of ξ. Then, the first $K \times M$ harmonic amplitudes $|FT_t(k,m)|$ are considered for each $f_t(i, j)$.

For faster extraction of the Fourier descriptors, the FFT was used instead of DFT. In order for the FFT to be applicable, the following image resolutions were used for the experiments: 64×64 pixels, 128×128 pixels and 256×256 pixels.

2.3.2 2D Zernike Moments

Zernike moments (Vinanco et al. 2003) are defined over a set of complex polynomials which forms a complete orthogonal set over the unit disk and are rotation invariant. The Zernike moments are calculated for each $f_t(i, j)$ with spatial dimension $N \times N$, as:

$$Z_{km} = \frac{2(k+1)}{\pi(N-1)^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} R_{km}(r_{ij})e^{-\hat{j}m\xi_{ij}} f_t(i,j),$$
$$0 \le r_{ij} \le 1, \quad (5)$$

where $k \in N^+$, $|m| \le k$ and $k - |m|$ is even and R the radial polynomials (Vinanco et al. 2003):

$$R_{km}(r) = \sum_{s=0}^{\frac{k-|m|}{2}} (-1)^s \frac{(k-s)!}{s!(\frac{k+|m|}{2}-s)!(\frac{k-|m|}{2}-s)!} r^{k-2s}. \quad (6)$$

The discrete polar coordinates are defined as in (4) with $r_{max} = 1$. The definition of the radial polynomial leads to $R_{km}(r) = R_{k,-m}(r)$. It can then easily be shown that $Z_{km} = Z_{k,-m}$. The number of Zernike moments for any order, k, is given by $k + 1$, while the number of moments up to order k is $(k/2+1)(k+1)$ (although because of the relationship between Z_{km} and $Z_{k,-m}$ given above, only the moments with $m \ge 0$ need to be known).

2.3.3 2D Krawtchouk Moments

Krawtchouk moments (Yap et al. 2003) are a set of moments formed by using Krawtchouk polynomials as the basis function set. The nth order classical Krawtchouk polynomials are defined as:

$$K_n(x; p, N) = \sum_{\kappa=0}^{N} a_{\kappa,n,p}x^{\kappa} = {}_2F_1\left(-n, -x; -N; \frac{1}{p}\right), \quad (7)$$

where $x, n - 0, 1, 2, \ldots, N$, $N > 0$, $p \in (0, 1)$, ${}_2F_1$ is the hypergeometric function defined as:

$${}_2F_1(a, b; c; z) = \sum_{\kappa=0}^{\infty} \frac{(a)_\kappa (b)_\kappa}{(c)_k}\frac{z^\kappa}{\kappa!} \quad (8)$$

and $(a)_\kappa$ is the Pochhammer symbol given by:

$$(a)_\kappa = a(a+1)\cdots(a+\kappa-1) = \frac{\Gamma(a+\kappa)}{\Gamma(a)}, \quad (9)$$

where $\Gamma(.)$ is the gamma function.

For each $f_t(i, j)$ with spatial dimension $N \times N$, the Krawtchouk moment invariants can be defined using the classical geometric moments:

$$M_{km} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} i^k j^m f_t(i, j). \quad (10)$$

The standard set of geometric moment invariants, which are independent to rotation (Hu 1962) can be written as:

$$\nu_{km} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}[i\cos\xi + j\sin\xi]^k \times [j\cos\xi - i\sin\xi)]^m f_t(i, j), \quad (11)$$

where

$$\xi = (1/2)\tan^{-1}\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \quad (12)$$

and μ are the central moments:

$$\mu_{pq} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i-\bar{x})^p(j-\bar{y})^q f_t(i,j),$$
$$p,q = 0,1,2,\ldots. \tag{13}$$

The value of ξ is limited to $-45° \leq \xi \leq 45°$. In order to obtain the exact angle ξ in the range of 0° to 360° modifications described in detail in Teague (1979) are required.

Following the analysis described in Yap et al. (2003), the rotation invariant Krawtchouk moments are computed by:

$$\tilde{Q}_{km} = [\rho(k)\rho(m)]^{-(1/2)} \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} a_{i,k,p_1} a_{j,m,p_2} v_{ij}, \tag{14}$$

where the coefficients $a_{\kappa,n,p}$ can be determined by (7), and $\rho(k)$, $\rho(m)$ can be calculated from the orthogonality condition (Yap et al. 2003). It should be noted that in our experiments the parameters p_1, p_2 were set to 0.5 (Yap et al. 2003).

A compact representation of the multi-view descriptor implies also a small number of descriptors per view, otherwise the shape matching time would be prohibitive. The numbers of descriptors for each of the above functionals are determined as follows:

Concerning the amplitudes of the Polar-Fourier coefficients, the values of up to order k_{FT} are selected, resulting in a total of $N_{FT} = (k_{FT} \times (k_{FT}+1))/2$ descriptors.

Similarly, selecting Krawtchouk moments of order k_{Kraw} results in $N_{Kraw} = (k_{Kraw} \times (k_{Kraw}+1))/2$ descriptors.

Finally, selecting Zernike moments of order k_{Zern} would give $(k_{Zern}/2+1)(k_{Zern}+1)$ descriptors, which is reduced to $N_{Zern} = (k_{Zern}/2+1)((k_{Zern}+1)/2+1)$ descriptors due to the relationship between Z_{km} and $Z_{k,-m}$ described in Sect. 2.3.2.

The total number of descriptors N_D for each view is given below:

$$N_D = N_{FT} + N_{Zern} + N_{Kraw}. \tag{15}$$

In order to produce a compact descriptor vector, the number N_D should be relevantly small, which implies small values of orders k_{FT}, k_{Zern} and k_{Kraw}, but not significantly small so as not to discard valuable information. It was found experimentally, that the optimal values are: $k_{FT} = 12$, $k_{Kraw} = 12$ and $k_{Zern} = 13$.

3 Matching Method

Similar to existing view-based approaches, the proposed framework measures the similarity between two 3D objects by summing up the similarity from all the corresponding images.

Let $\mathbf{D}_t$ be the descriptor vector of the tth view, which is extracted according to the procedure described in Sect. 2.3. The dissimilarity metric between a corresponding pair of views of two models A and B is given by the L1-distance:

$$d_t = \sum_{k=1}^{N_D} |D_t^A(k) - D_t^B(k)| \tag{16}$$

where N_D is the number of descriptors per view.

3.1 3D/3D Matching

Let now A and B be two 3D models, with descriptor vectors $\mathbf{D}_t^A$ and $\mathbf{D}_t^B$, respectively, where $t = 0,\ldots,N_V$ and N_V the total number of views. It must be noted that in the case of depth images, $N_V = 18$, while in the case of silhouettes only half of the views ($N_V = 9$) are kept. This is due to the fact that the symmetrical silhouette images are identical, thus, produce the same descriptors. The total dissimilarity d between the models A and B is given by the following equation:

$$d = \sum_{t=1}^{N_V} d_t, \tag{17}$$

where d_t is the dissimilarity of the tth view described in (16). Note that the dissimilarity metric does not include matching of all views of model A with all views of model B ("all-to-all" matching), it includes matching of only the corresponding views (i.e. matching of $View_1^A$ with $View_1^B$, $View_2^A$ with $View_2^B$ and so on). The 3D/3D matching procedure is depicted in Fig. 3. The numbering of views has been arbitrarily chosen but it is consistent for every 3D model. This results in a significantly fast matching procedure, however, it requires that rotation normalization provide 100% success, not only in terms of identification of the three principal axes but also in terms of orientation of each axis.

In Fig. 4, the results of rotation normalization for three 3D models, by using the method described in Sect. 2.1, are presented. The method succeeds in detecting the first principal axis (x) in all three cases, while, in the third case, it confuses the second with the third principal axis (y, z). Although all three models look similar, a comparison of the first two with the third, using (17), would produce an unexpectedly large dissimilarity. This is due to the fact that the views numbers in the third model correspond to different model parts compared to those in the first two models (e.g. in the third case, view #2 depicts the back of the animal, while in the first two cases, view #2 depicts the left side view, etc.).

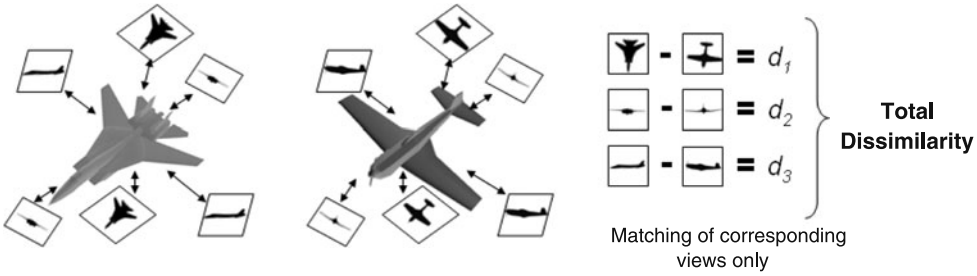

Fig. 3 The proposed Similarity Matching Framework. The total dissimilarity between two 3D objects is the sum of the dissimilarities of the corresponding views

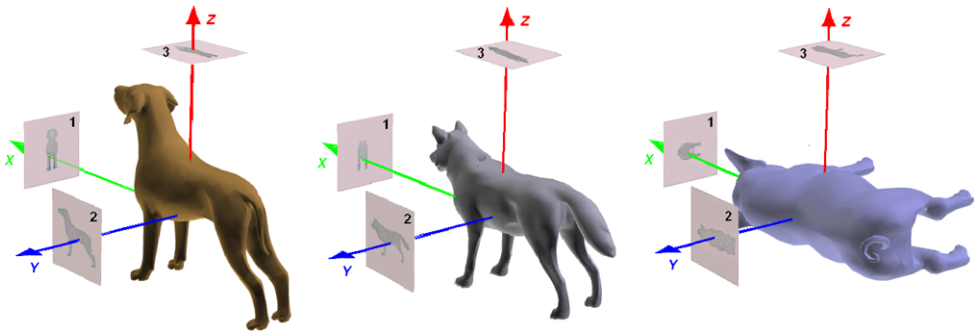

Fig. 4 The results of unsuccessful rotation normalization in 3D/3D matching. In the third 3D object, the method confuses the second with the third principal axis, which leads in false detection of the corresponding views

The above problems cannot be avoided due to the inherent limitations of PCA and VCA. However, for a model rotated by an arbitrary angle, the proposed rotation estimation method is accurate in detecting the three principal axes, disregarding of the axis order and orientation. Taking the above into account, the similarity matching of two 3D models can be improved, if, instead of a single set of views, 24 different sets of views are used for the second model. This number is equal to the different possible alignments of the 3D model, after the rotation normalization step is applied. In order to produce these sets of views, the 3D model should be rotated 24 times at intervals of 90 degrees. In the case that the multiple views are taken from the six vertices of a regular 8-hedron, then, in each of the 24 rotations, the viewpoints will always lie at these six vertices. Thus, the views (and consequently the 2D descriptors) need to be extracted only once. In order to have an adequate set of views, a 32-hedron (produced by the 8-hedron at the first level of tesselation) is used instead. The total dissimilarity d between A and B is now modified as:

$$d = \min\{d^r\} = \min\left\{\sum_{t=1}^{N_V} d_t^r\right\}, \tag{18}$$

where $r = 1, \ldots, 24$ is the total number of rotations of the second model, d^r is the dissimilarity of the rth rotation and $N_V = 18$ is the number of views of the 32-hedron.

3.2 2D/3D Matching

Retrieval of 3D models can also be achieved if, instead of a 3D model, a single 2D image is used as query. In order to measure the dissimilarity, the query 2D image is compared to the N_V views of the 3D model and the most similar (to the image) view is selected:

$$d = \min\{d_t\} = \min\left\{\sum_{k=1}^{N_D} |D^Q(k) - D_t^B(k)|\right\} \tag{19}$$

where $t = 1, \ldots, N_V$, $N_V = 18$ is the total number of views of model B, $D^Q(k)$ are the descriptors of the query image Q and $D_t^B(k)$ are the descriptors of the tth view of model B.

It is obvious that 2D/3D matching cannot be as efficient as 3D-3D matching, since a 2D image is unable to capture the global visual information of an object. However, it is much easier to provide a 2D image as query than a 3D model (e.g. take a photo of an object or draw a sketch). In order to produce a valid dissimilarity metric, the query 2D image should be of the same type as the multiple views generated from a 3D model, i.e. either binary (black/white) images or depth images. Since depth images are usually difficult to create, binary images are preferred.

3.3 Computational Aspects

A method for 3D search and retrieval should not only be effective in retrieving similar objects but also be adequately fast, which makes it appropriate for online applications. Therefore, a major aspect in the proposed method is the computation time.

The main time-consuming parts throughout the descriptor extraction procedure are the multiple views generation and the 2D functionals computation. The views generation

Table 1 Average computation times for descriptor extraction and matching procedures

Action	Time (msec)
Views generation	2587
Polar-Fourier descriptors extraction	63
Krawtchouk descriptors extraction	398
Zernike descriptors extraction	811
Matching between 2 models	10

time per object varies from a few milliseconds to a few seconds and depends on the number of triangles that constitute the object's 3D mesh. On the other hand, the 2D functionals computation time depends on calculations performed on a fixed number of equally-sized 2D images, thus, has only slight variations.

Similarly, the matching time between two 3D models is constant, since it involves distance computation of equally-sized descriptor vectors.

However, this time may become prohibitive when the number of database models increases significantly and each 3D model should be rotated 24 times according to the proposed matching framework. In this case, a method based on look-up-table, similar to the one presented in Chen et al. (2003) is utilized to speed-up the matching process. More specifically, the descriptor values for each 2D view are quantized to 8 bits. Then, a 256×256 matrix is created to store the pairwise descriptor distances. During the matching step, calculation of distance between 2 descriptors is reduced to a simple look-up to the 256×256 matrix.

In Table 1, the average computation times for descriptor extraction and matching procedures are summarized. The times were obtained using a PC with a 2.4 GHz processor and 3GB RAM, running operating system Windows XP.

The matching time presented in Table 1 corresponds to the depth-buffer CMVD descriptor, which uses all 18 views. In the case of the binary CMVD descriptor, where only half views are used, the time is reduced to 5 msec.

The proposed method uses less number of views (only 18) than the competing view-based methods presented in Chen et al. (2003) and Ohbuchi et al. (2008). In terms of descriptor extraction time, CMVD is much faster than both of the above. Moreover, in the similarity matching between two models, CMVD is faster than the LFD presented by Chen et al. (2003). The method presented by Ohbuchi et al. (2008), however, extracts only one descriptor vector (of dimension up to 1500). This results in a faster similarity matching.

The matching time of the proposed method can be significantly decreased, by taking a closer look at the rotation normalization results. More specifically, in most of the cases, the rotation normalization method confuses the first with the second principal axes, while the third principal axis is correctly predicted. Assuming correct prediction of the 3rd principal axis, the 3D object has to be rotated only 8 times instead of 24. This makes the matching process 3 times faster, without significant degradation of performance.

4 Experiments

4.1 Evaluation of the 3D/3D Matching Method

The proposed method was experimentally evaluated using three different databases. The first one was compiled from the Internet by us and it is called "the ITI database" (The VICTORY 3D Search Engine). It consists of 544 3D models classified in 13 different categories: 27 animals, 17 spheroid objects, 64 conventional airplanes, 55 delta airplanes, 54 helicopters, 48 cars, 12 motorcycles, 10 tubes, 14 couches, 42 chairs, 45 fish, 53 humans, and 103 other models. This choice reflects primarily the shape of each object and secondarily its function. The average numbers of vertices and triangles of the models in the new database are 5080 and 7061, respectively. The second dataset, the "Princeton Shape Benchmark (PSB)", was formed in Princeton University (Shilane et al. 2004) and it consists of 907 3D models classified into 35 main categories. This classification reflects primarily the function of each object and secondarily its form. Finally, the third dataset, the "Engineering Shape Benchmark (ESB)", contains a total of 867 3D CAD models from the mechanical engineering domain (Jayanti et al. 2006). They are classified into 3 main classes: 107 flat-thin wall components, 281 rectangular-cubic prism and 479 solids of revolution.

To evaluate the proposed method, each 3D model was used as a query object. The retrieval performance was evaluated in terms of the well-known "precision-recall", where precision is the proportion of the retrieved models that are relevant to the query and recall is the proportion of relevant models in the entire database that are retrieved in the query.

Two variations of the proposed method are used: the CMVD-Binary that uses binary images and the CMVD-Depth that uses depth images. First of all, experiments were performed for various 2D image resolutions and numbers of views, as well as for each 2D functional separately, in order to select the optimal values of the above parameters. In Fig. 5, the precision-recall diagram of the CMVD-Depth method in PSB database is presented for three different 2D image resolutions: 64×64, 128×128 and 256×256 pixels. By increasing the views resolution from 64×64 to 128×128 pixels, the performance is significantly improved. Further increase of resolution to 256×256 pixels produces no further improvement, which leads in the selection of 128×128 pixels resolution. In Fig. 6, the results of the

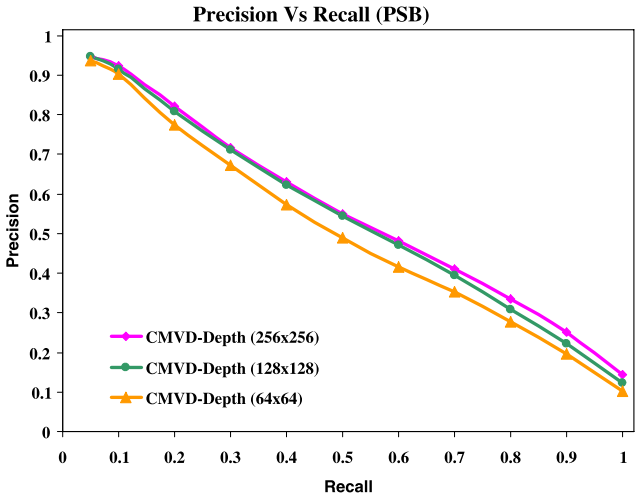

Fig. 5 Precision-recall diagram of the proposed method, in PSB database, using different 2D image resolutions: 64 × 64, 128 × 128 and 256 × 256 pixels

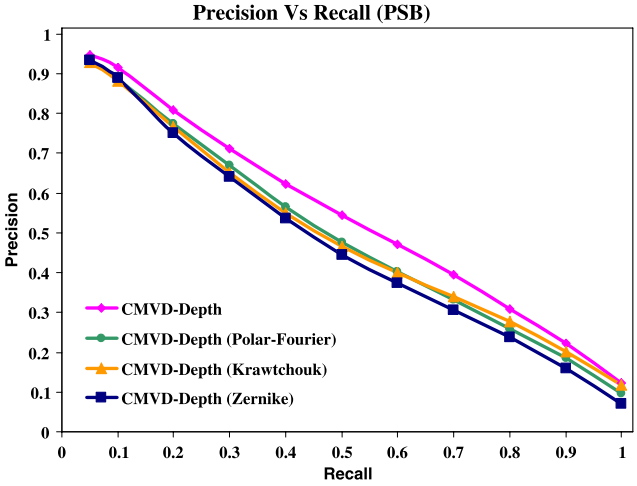

Fig. 7 Precision-recall diagram of the proposed method, in PSB database, using each 2D functional separately and their combination

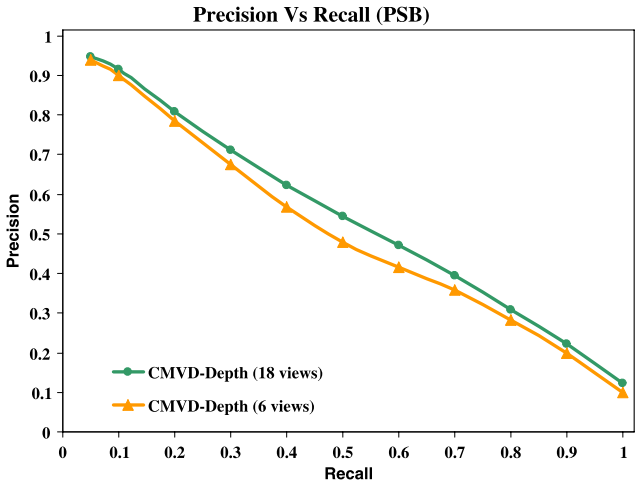

Fig. 6 Precision-recall diagram of the proposed method, in PSB database, using different bounding polyhedra: 8-hedron (6 views) and 32-hedron (18 views)

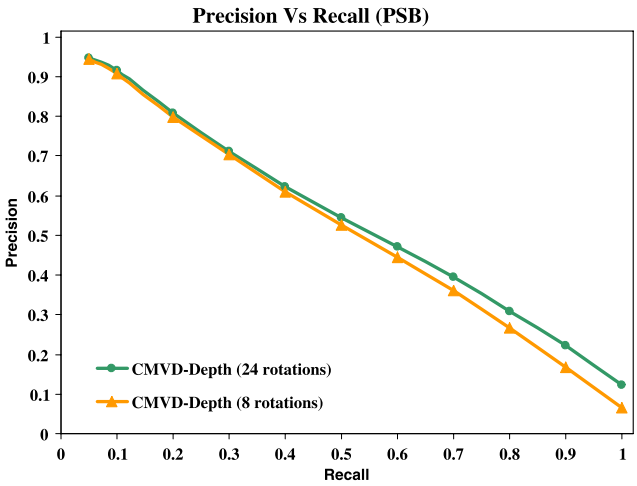

Fig. 8 Precision-recall diagram of the proposed method, in PSB database, using either the 24-rotation matching scheme or the much faster 8-rotation matching. In the 8-rotation matching, we assume that the third principal axis is correctly predicted

CMVD-Depth method, using two different bounding polyhedra, are depicted. It is obvious that the 18-view representation of a 32-hedron produces better retrieval results than the 6-view representation of the 8-hedron, which maintains the assumption that 18 views provide a better representation of the overall 3D shape than the 6 views.

Figure 7 presents the results of CMVD-Depth method in PSB database, using as 2D descriptor either each 2D functional separately or their combination. It is worth to mention that all three 2D rotation-invariant functionals demonstrated satisfactory performance. Their combination, however, achieved even higher retrieval accuracy. This relies on the fact that each functional describes different 2d geometric properties and each one's role is complementary for the description of the 2D view.

The proposed similarity matching framework requires rotation of one of the two 3D objects 24 times, in order to compensate for the inherent limitations of rotation normalization. In order to reduce the matching time, the proposed matching method can be slightly modified as follows: assuming that the 3rd principal axis is correctly predicted (which is true for most of the cases), the 3D object has to be rotated only 8 times, instead of 24. This makes the matching process 3 times faster, without significant degradation of performance. In Fig. 8, the precision-recall diagram of the CMVD-Depth method in PSB database, using 8 and 24 rotations in similarity matching, is presented. For recall values

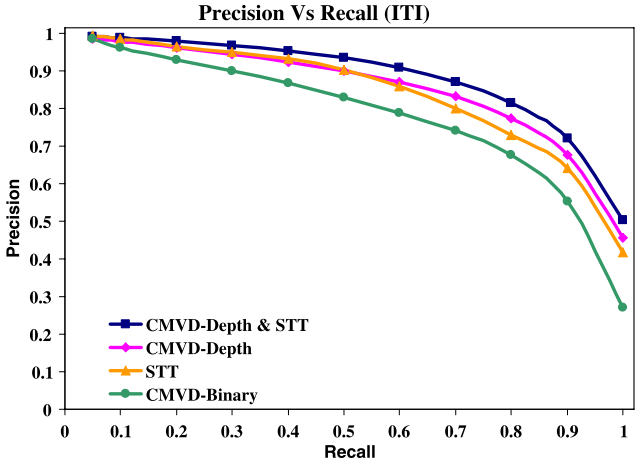

Fig. 9 Precision-recall curves diagram of the proposed method using the ITI database

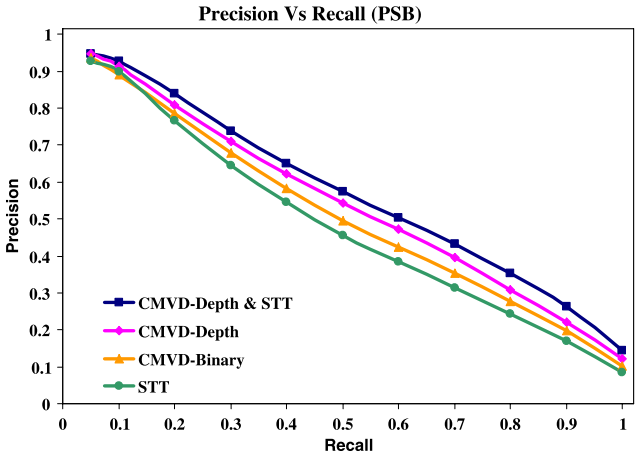

Fig. 10 Precision-recall curves diagram of the proposed method using the PSB database

up to 0.5, the performance of the above matching variations is almost the same. The decrease in retrieval accuracy of the 8-rotation matching, for recall values higher than 0.5, is of little importance, taking into account a real-life 3D shape retrieval scenario. In this case, the retrieval performance in only the first ranked positions matters (the user would not browse the entire ranked list, the first 10–20 retrieved objects suffice), while retrieval time should be considerably low.

In Fig. 9, the precision-recall diagrams of the proposed method using the ITI database are depicted. It is obvious that the use of depth images instead of binary images improves the performance of the method, since the depth image can capture more details in a 3D object. The proposed method is also compared with the Spherical Trace Transform (STT), which was presented in Zarpalas et al. (2007). The Spherical Trace Transform is a transform-based method and produces rotation-invariant descriptor vectors with high discriminative power. Although STT has demonstrated very high performance in ITI database, the CMVD-Depth method is slightly better.

An attempt to combine the advantages of the proposed view-based method and the STT (transform-based) method was also made. The retrieval results of the combined "CMVD-Depth & STT" method are depicted in Fig. 9. As expected, the combination of different types of descriptors has achieved the highest retrieval performance.

Similar results are obtained using the PSB and the ESB databases. In Figs. 10 and 11, the precision-recall diagrams of the proposed method (CMVD-Binary and CMVD-Depth) and the combination of CMVD-Depth with STT (CMVD-Depth & STT) are shown for the PSB and the ESB database, respectively. The diagrams demonstrate the retrieval efficiency of CMVD-Depth method, as well as of the combination of CMVD-Depth with the Spherical Trace Transform.

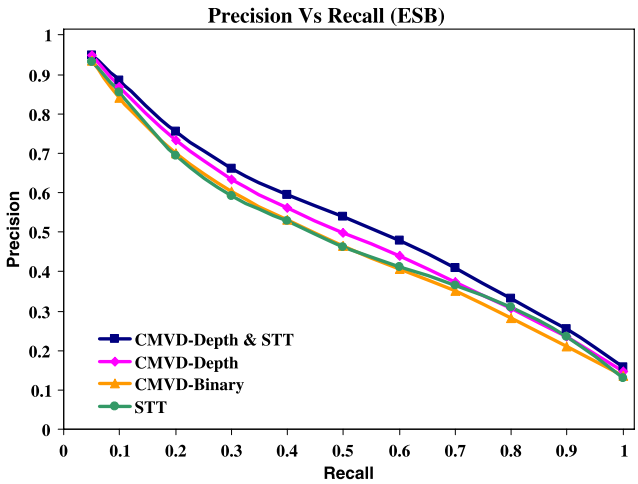

Fig. 11 Precision-recall curves diagram of the proposed method using the ESB database

Our results were also compared to those of the following three methods. The first two belong to the category of the "view-based" methods, while the third combines 2D view-based methods with a transform-based method.

- *The light field descriptor (LFD)*: Similar to the proposed method, LFD (Chen et al. 2003) uses a representation of a model as a collection of images rendered from uniformly sampled positions on a view sphere. The distance between two descriptors is defined as the minimum L1-difference, taken over all rotations and all pairings of vertices on two 12-hedra. The advantages of CMVD over LFD are: (a) significantly less number of rendered 2D views, since only 18 views are extracted, (b) less number of different rotations, since a pose normalization step has been added, and (c) the use of depth images instead of only

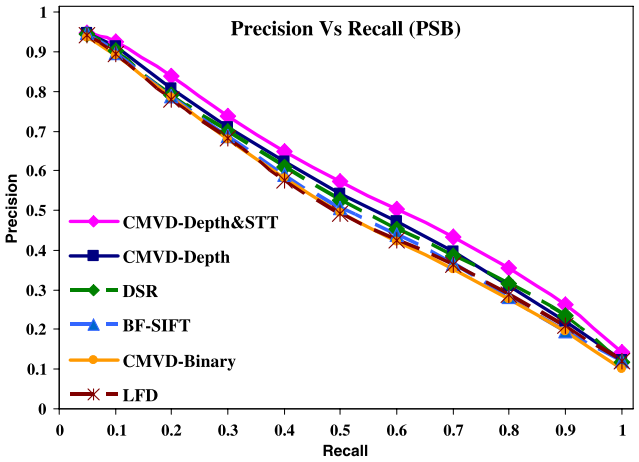

Fig. 12 Comparison of the proposed method with LFD, BF-SIFT and DSR in terms of precision-recall, using the PSB database

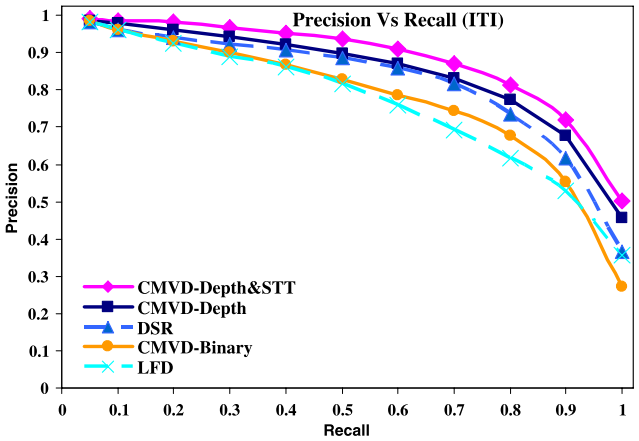

Fig. 13 Comparison of the proposed method with LFD and DSR in terms of precision-recall, using the ITI database

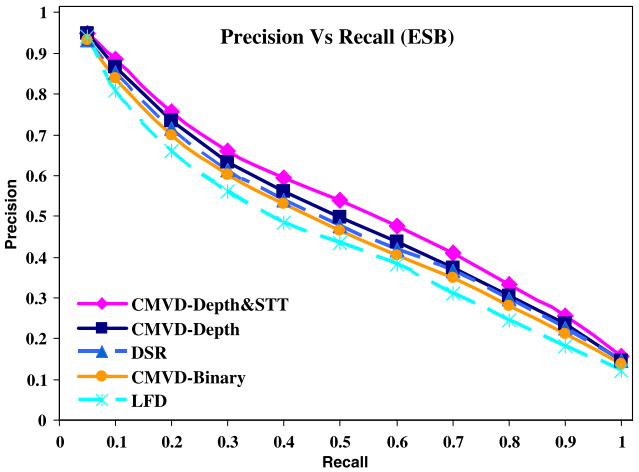

Fig. 14 Comparison of the proposed method with LFD and DSR in terms of precision-recall, using the ESB database

black/white silhouettes, which captures more details per view.

- *The bag-of-features SIFT algorithm (BF-SIFT)*: BF-SIFT (Ohbuchi et al. 2008) is also based on the rendering of a set of range images from multiple viewpoints. For each image, the method uses the well-known Scale Invariant Feature Transform (SIFT) to extract local features. Finally, the method integrates all the local features of the model into a single feature vector by using the Bag-of-Features approach.
- *DSR*: This method (Vranic 2004) is denoted as hybrid DSR and it stands for Depth-Buffer, Silhouette and REXT (Radialized Spherical Extent Function). It is a combination of two view-based methods (Depth-Buffer descriptor and Silhouette descriptor) and the well-known transform-based method REXT (Vranic 2003). DSR is one of the best-known shape matching methods that produce very accurate retrieval results, which reinforces the notion that the combination of view-based with transform-based methods achieve the highest efficiency of descriptors.

It should be noted that we did not implement the above methods. The performance of the first and the third method was computed by using the executables taken from the home pages of the authors, while the results of the second method were directly extracted from the ones presented in Ohbuchi et al. (2008) and are available only for the PSB database.

Figure 12 contains a numerical precision versus recall comparison of CMVD-Depth with the aforementioned methods using the PSB database. It is clear that the proposed method outperforms all others. The difference is even more noticeable when the proposed method is combined with the Spherical Trace Transform (CMVD-Depth & STT).

Similar results are obtained using the ITI and the ESB databases. Figures 13 and 14 illustrate the precision-recall diagrams, using the ITI and the ESB database respectively, where the proposed method (CMVD-Depth) and the combination of CMVD-Depth with STT (CMVD-Depth & STT) are compared to the LFD and DSR methods. The results are impressive, since both of the proposed approaches outperform the other existing 3D shape retrieval methods. It is worth to mention that CMVD-Depth alone is slightly better than DSR, which combines view-based and transform-based information.

4.2 Evaluation of the 2D/3D Matching Method

The experimental results presented above have proven that the proposed method is very effective in retrieving similar 3D objects from a database, using a 3D model as query. However, an input 3D model is not always available and it cannot be created from scratch by a non-expert user. On the other hand, using as query a 2D image or a hand-drawn sketch is more convenient for inexperienced users.

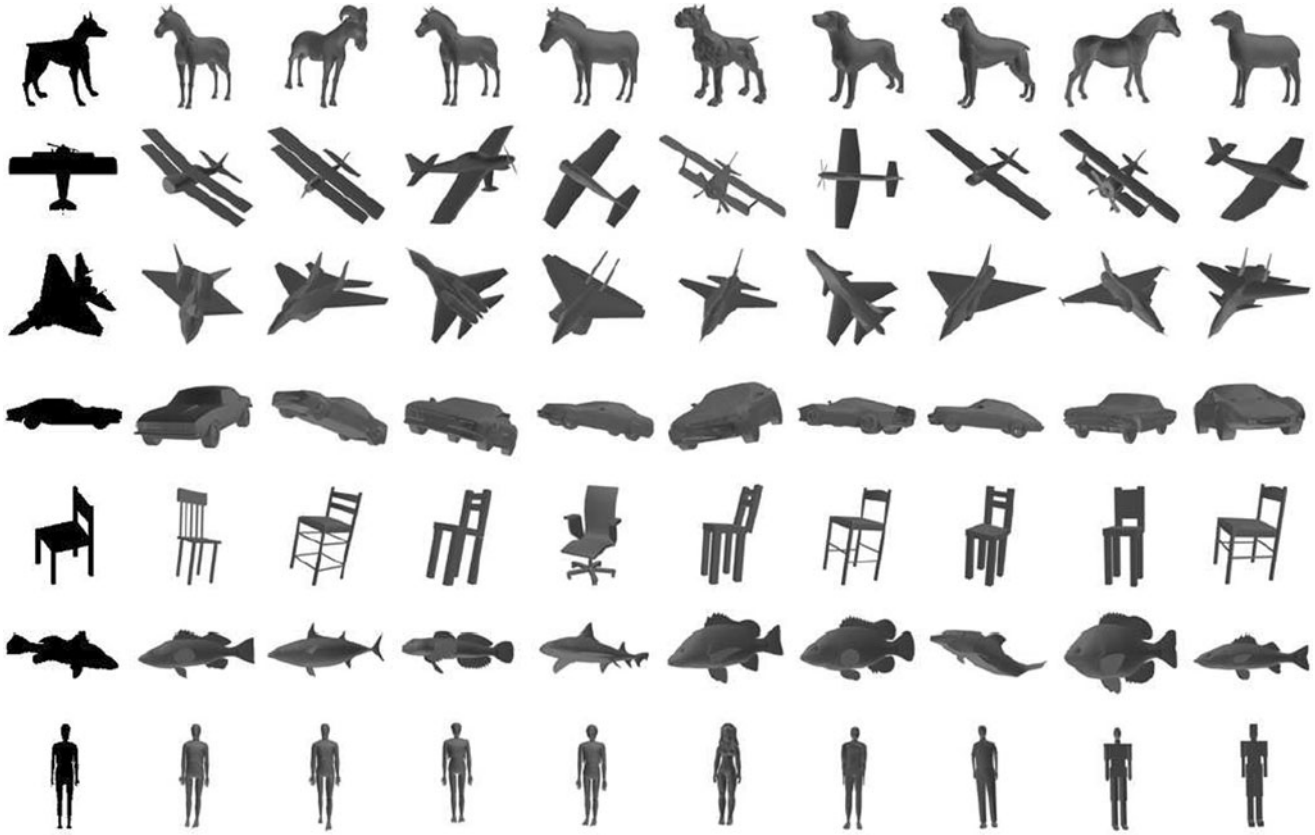

Fig. 15 Retrieved 3D models using only a single view as query. The first column depicts the query 2D image, while the rest depict the retrieved 3D objects

As explained in the previous sections, the proposed framework provides efficient search and retrieval capabilities using only a 2D image or a sketch as a query, when an input 3D model is not available. In this case, the matching is performed by computing the dissimilarity between the query image and the multiple 2D views of the 3D object and finally selecting the view with the lowest dissimilarity. This requires that the input 2D images (or sketches) are in the binary (black/white) form, since depth images cannot be easily sketched or retrieved.

It is clear that 2D/3D matching cannot be compared with 3D/3D matching, in terms of retrieval performance, since the amount of information enclosed in a 2D query is significantly less than in a 3D object. In order to measure the performance of the 2D/3D matching method, an evaluation scheme more qualitative than quantitative, other than precision-recall, is needed. The evaluation procedure, followed in this paper for 2D/3D matching, is described below.

For each 3D object of the ITI database, a single 2D view, preferably the most significant, was selected from the set of the 18 binary images. Each of the selected binary images was used as query in order to retrieve 3D objects from the ITI database. Note that the term "significant" has been used to characterize the view that an ordinary user would choose, if asked to represent a 3D object in two dimensions. In order to ensure validity of results, an experiment has been conducted, where 50 users were asked to choose among all views of several 3D objects the most representative ones. Although the identification of the most significant view for a 3D object is highly subjective, the one that reflected the majority of users was eventually selected. Finally, for each 2D query, a rank list of retrieved results is generated and a qualitative evaluation of the k-first results, in terms of similarity to the query, takes place.

The results are impressive, since, for the majority of the queries, the 3D objects retrieved at the first positions of the rank list belong to the same category as the query. Figure 15 depicts the retrieved results for seven example 2D queries. The first model is the query binary image, while the rest are the first nine retrieved 3D objects. Note that the 3D object that corresponds to the query image has been removed from the rank list.

A query 2D image can be taken either by drawing a sketch or from the user's digital camera. In order to support these types of queries, a user-friendly interface has been appropriately designed within the VICTORY project (The VICTORY 3D Search Engine). In Fig. 16, screenshots of the

Fig. 16 Retrieved 3D models using as query a hand-drawn sketch

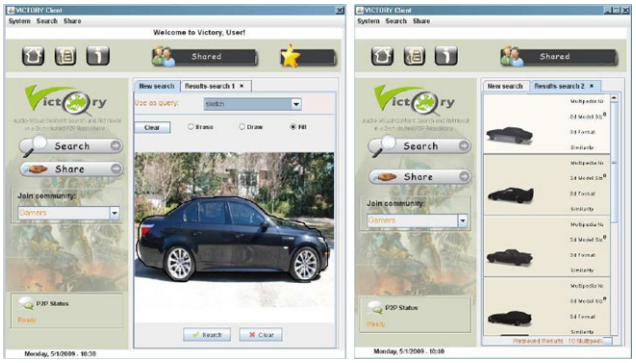

Fig. 17 Retrieved 3D models using as query a 2D image. The image is separated from the background by using the proposed interface for user-assisted segmentation

VICTORY search and retrieval tool are given. The interface provides a typical drawing tool, allowing the user to easily draw a sketch of the query., as well as a panel to visualize the retrieved results.

Despite the simplicity of the proposed interface, the quality of the input hand-drawn sketches (and consequently of the retrieved results) depends on the users' drawing skills. However, in such a search and retrieval scenario, even few relevant hits among the retrieved results are enough. Then, as a second step, the user can select one of the relevant retrieved 3D objects as new query in order to retrieve more relevant results.

Alternatively, the tool provides an extra functionality to load a 2D image and draw a contour of the desired object (Fig. 17). This manual segmentation, which separates the query image from the background, is very useful and produces more accurate results. The retrieved 3D objects in the first positions of the rank lists are all similar to the queries, which demonstrates the efficiency of the proposed method to support multiple types of queries.

4.3 Evaluation of CMVD in SHREC'09

The performance of the proposed method was also compared with the best shape retrieval methods worldwide, in

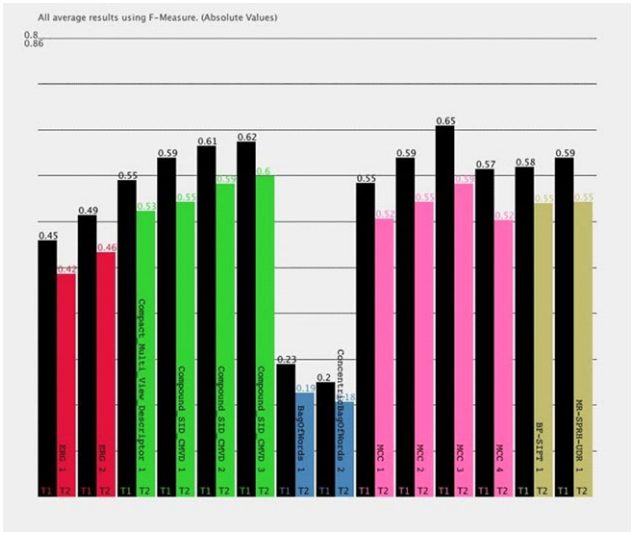

Fig. 18 Average results in SHREC'09 structural shape retrieval track using F-measure. The *black columns* depict the 1st Tier results, while the *colored ones* depict the 2nd Tier results

Shape Retrieval Contest 2009 (SHREC'09). More specifically, CMVD participated in the following tracks:

- *Structural shape retrieval*,
- *Shape retrieval contest on a new generic shape benchmark* and
- *Shape retrieval contest of partial 3D models*.

4.3.1 Structural Shape Retrieval

In this track (Hartveldt et al. 2009), the shape repository contained 200 models. The models were classified in 10 main classes. Each main class contained a pair of (two) subclasses. Matching models from inside a class are highly relevant and matching results between models the two pair class are marginally relevant. The performance of the shape retrieval was measured using first and second tier precision and recall, finally presented as the F-measure. This measure gives a nice overview of the complete retrieval performance and the method especially.

Five groups have participated in the Structural Shape Retrieval track and they have submitted 14 sets of rank lists. In this track, a combination of the proposed CMVD descriptor with a transform-based method, presented in Mademlis et al. (2008b), was used. The results are presented in Fig. 18.

Regarding the First Tier, it is clear from the results that for all different measures (Precision, Recall and F-measure), the proposed approach is ranked second among 14 different submitted runs. The retrieval performance of the method is improved in the Second Tier, where the third run of the compound CMVD approach is always ranked first for all different measures. This is very important, if we also consider that

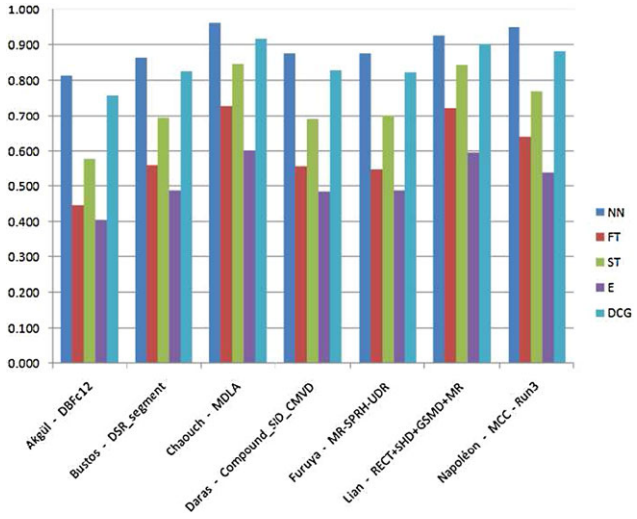

Fig. 19 Bar plot of the Nearest Neighbor (NN), First Tier (FT), Second Tier (ST), E-measure (E) and Discounted Cumulative Gain (DCG) for the best runs of each participant, in SHREC'09 generic shape retrieval track

the second tier provides a better overview of the complete retrieval performance than the first tier, since it relies on the first 20 retrieved objects, instead of 10 used in the first tier. Similar performance is also achieved by the MCC method proposed by Napoleon (Napoleon et al. 2008). The third run of the MCC method is ranked first in the first tier and second in the second tier, just after the proposed compound CMVD method.

4.3.2 Shape Retrieval Contest on a New Generic Shape Benchmark

This dataset (Akgul et al. 2009) consists of 720 3D objects classified in 40 categories. The query set is composed of 80 3D objects, two from each category. The performance of the shape retrieval was measured using the following evaluation measures: Precision-Recall curve; Average Precision (AP) and Mean Average Precision (MAP); E-Measure; Discounted Cumulative Gain; Nearest Neighbor, First-Tier (Tier1) and Second-Tier (Tier2).

Seven groups have participated in the Generic Shape Retrieval track and they have submitted 22 sets of rank lists. In this track, the same compound CMVD approach with the one presented in Structural Shape Retrieval Track was used.

In Fig. 19, a bar plot of the Nearest Neighbor (NN), First Tier (FT), Second Tier (ST), E-measure (E) and Discounted Cumulative Gain (DCG) for each method is presented. It is obvious that the MDLA method of Chaouch achieved the best performance among the seven participants, while the performance of the proposed compound CMVD method is acceptable with respect to the overall view. The MCC method by Napoleon, which was competitive with the proposed method in Structural Shape Retrieval Track, seems to outperform the compound CMVD method in this track and is ranked second after the MDLA method of Chaouch.

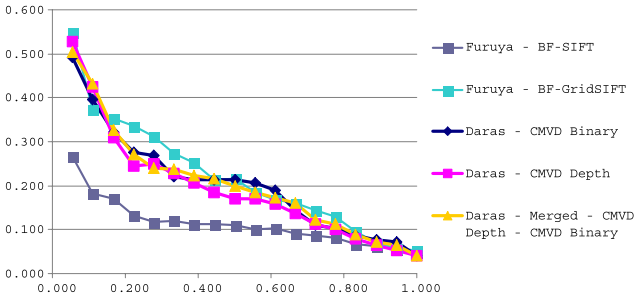

Fig. 20 Precision-recall curves in SHREC 2009 partial shape retrieval track

4.3.3 Shape Retrieval Contest of Partial 3D Models

This dataset (Axenopoulos et al. 2009) consists of 720 3D objects classified in 40 categories. This query set is composed of 20 range images, which are acquired by capturing range data of 20 models from arbitrary view directions.

Three variations of the CMVD descriptor have been used for evaluation, resulting in three different runs: (a) CMVD-Binary that uses binary images, (b) CMVD-Depth that uses depth images and (c) merged CMVD-Depth and CMVDBinary, which is a weighted sum of the dissimilarities computed by each method separately. The performance has been evaluated using 6 different evaluation metrics: Nearest Neighbor (NN), First Tier (FT), Second Tier (ST), E-measure, Discounted Cumulative Gain (DCG) and Precision Recall Diagram.

In Fig. 20, the precision-recall curves of all methods are depicted. It is clear from the results that all three runs of our CMVD approach outperform the BF-SIFT method, while they are competitive with the BF-GridSIFT method. More specifically, our method outperforms BF-GridSIFT for recall values close to 0.1, from 0.5 to 0.6 and greater than 0.9. This means that we have a better retrieval accuracy for the first 10% of relevant retrieved results, i.e. in the first places of the rank list, while we have comparative or even better retrieval accuracy for the last 50% of relevant retrieved results. Furthermore, CMVD-Depth has the same NN (0.45) with BF-GridSIFT. Nearest Neighbor indicates the relevance to the query of the first retrieved result. Regarding the Emeasure and DCG, the CMVD-Binary and the merged-CMVD achieved better performance respectively. Their values are very close to the ones of BF-GridSIFT.

The overall results demonstrate that the CMVD descriptor is among the best 3D object retrieval methods that use as query only a single 2D view, such as a range scanned image. This maintains the assumption that the combination of the three 2D rotation-invariant functionals, which is introduced in this paper, provides a very powerful descriptor for a single view.

5 Conclusions

In this paper, a unified framework for 3D object retrieval was presented. The method provides search and retrieval capabilities by supporting multimodal queries (3D objects, 2D images or sketches). The proposed view-based approach creates a compact representation of a 3D object as a set of multiple 2D views (both binary and depth images) taken from uniformly distributed viewpoints. For each view, a set of 2D rotation-invariant shape descriptors, based on the Polar-Fourier Transform, Zernike Moments and Krawtchouk Moments, is produced. The paper also introduced a novel matching scheme, which calculates the global shape similarity between two 3D models by effectively combining the information extracted from the multi-view representation.

The proposed Compact Multi-View Descriptor (CMVD) was evaluated in terms of retrieval performance using three different databases. The results were compared to those of the best-known retrieval methods in the literature and clearly demonstrate that the proposed method outperforms all others in terms of precision-recall. Another interesting conclusion is that the combination of a view-based method, such as CMVD, with a transform-based method, such as the Spherical Trace Transform, achieved the highest retrieval performance.

Finally, the effectiveness of the proposed 3D shape retrieval framework, using a single 2D image as query, was tested. The results were impressive, since the method was able to retrieve relevant objects by exploiting a very limited amount of information enclosed in a 2D image.

Acknowledgements This work was supported by the EC project VICTORY (http://www.victory-eu.org).

References

Akgul, C., Axenopoulos, A., Bustos, B., Chaouch, M., Daras, P., Dutagaci, H., Furuya, T., Godil, A., Kreft, S., Lian, Z., Napoleon, T., Mademlis, A., Ohbuchi, R., Rosin, P. L., Sankur, B., Schreck, T., Sun, X., Tezuka, M., Yemez, Y., Verroust-Blondet, A., & Walter, M. (2009). SHREC 2009—generic shape retrieval contest. In *30th international conference on EUROGRAPHICS 2009, workshop on 3D object retrieval*. Munich, Germany, March 2009.

Ankerst, M., Kastenmuller, G., Kriegel, H. P., & Seidl, T. (1999). 3D shape histograms for similarity search and classification in spatial databases. In *Proc. of the 6th int. symp. spatial databases (SSD1999)*. Hong Kong, 1999.

Ansary, T. F., Vandeborre, J.-P., & Daoudi, M. (2007). 3D-model search engine from photos. In *Proc. ACM CIVR 2008* (pp. 89–92).

Axenopoulos, A., Daras, P., Dutagaci, H., Furuya, T., Godil, A., & Ohbuchi, R. (2009). SHREC 2009—shape retrieval contest of partial 3D models. In *30th international conference on EUROGRAPHICS 2009, workshop on 3D object retrieval*. Munich, Germany, March 2009.

Belkasim, S. O., Shridhar, M., & Ahmadi, M. (1991). Pattern recognition with moment invariants: A comparative study and new results. *Pattern Recognition*, *24*(12), 1117–1138.

Bustos, B., Keim, D. A., Saupe, D., Schreck, T., & Vranic, D. V. (2005). Feature-based similarity search in 3D object databases. *ACM Computing Surveys*, *37*(4), 345–387.

Bustos, B., Keim, D., Saupe, D., & Schreck, T. (2007). Content-based 3d object retrieval. *IEEE Computer Graphics and Applications*, *27*(4), 22–27.

Canterakis, N. (1999). 3D Zernike moments and Zernike affine invariants for 3D image analysis and recognition. In *Proc. of the Scandinavian conference on image analysis*, 1999.

Chaouch, M., & Verroust-Blondet, A. (2006). Enhanced 2D/3D approaches based on relevance index for 3D-shape retrieval. In *SMI'06*. Matsushima, Japan, June 2006.

Chen, D.-Y., & Ouhyoung, M. (2002). A 3D object retrieval system based on multi-resolution reeb graph. In *Proc. of Computer Graphics Workshop*. Tainan, ROC.

Chen, D.-Y., Ouhyoung, M., Tian, X.-P., Shen, Y.-T., & Ouhyoung, M. (2003). On visual similarity based 3d model retrieval. In *Proc. of Eurographics* (pp. 223–232). Granada, Spain.

Daras, P., Zarpalas, D., Axenopoulos, A., Tzovaras, D., & Strintzis, M. G. (2006a). Three-dimensional shape-structure comparison method for protein classification. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *3*(3), 193–207.

Daras, P., Zarpalas, D., Tzovaras, D., & Strintzis, M. G. (2006b). Efficient 3-d model search and retrieval using generalized 3-d radon transforms. *IEEE Transactions on Multimedia*, *8*(1), 101–114.

Filali Ansary, T., Daoudi, M., & Vandeborre, J.-P. (2007). A Bayesian 3D search engine using adaptive views clustering. *IEEE Transactions on Multimedia*, *9*(1), 78–88.

Fischer, K., & Gartner, B. (2004). The smallest enclosing ball of balls: Combinatorial structure and algorithms. *International Journal of Computational Geometry and Applications (IJCGA)*, *14*, 341–387.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge: Cambridge Bradford Books, MIT Press.

Funkhouser, T., Min, P., Kazhdan, M., Chen, J., Halderman, A., Dobkin, D., & Jacobs, D. (2003). A search engine for 3D models. *ACM Transactions on Graphics*, *22*(1), 83–105.

Goodall, S., Lewis, P. H., Martinez, K., Sinclair, P. A. S., Giorgini, F., Addis, M., Boniface, M. J., Lahanier, C., & Stevenson, J. (2004). SCULPTEUR: Multimedia retrieval for museums. In *Proc. of the image and video retrieval: Third international conference (CIVR'04)* (pp. 638–646).

Hartveldt, J., Spagnuolo, M., Axenopoulos, A., Biasotti, S., Daras, P., Dutagaci, H., Furuya, T., Godil, A., Li, X., Mademlis, A., Marini, S., Napoleon, T., Ohbuchi, R., & Tezuka, M. (2009). SHREC 2009 track: Structural shape retrieval on watertight models. In *30th international conference on EUROGRAPHICS 2009, workshop on 3D object retrieval*. Munich, Germany, March 2009.

Hilaga, M., Shinagawa, Y., Kohmura, T., & Kunii, T. L. (2001). Topology matching for fully automatic similarity estimation of 3D shapes. In *Proc. of ACM SIGGRAPH 2001* (pp. 203–212).

Horn, B. (1984). Extended Gaussian images. *Proceedings of the IEEE*, *72*(12), 1671–1686.

Hu, M. K. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, *8*, 179–197.

Huber, F., & Hebert, M. (2001). Fully automatic registration of multiple 3D data sets. In *Proc. of IEEE workshop on computer vision beyond the visible spectrum: Methods and applications (CVBVS)*. Kauai, Hawaii, USA, Dec. 2001.

Iyer, M., Jayanti, S., Lou, K., Kalyanaraman, Y., & Ramani, K. (2005). Three dimensional shape searching: State-of-the-art review and future trends. *Computer Aided Design*, *5*(15), 509–530.

Jayanti, S., Kalyanaraman, K., Iyer, N., & Ramani, K. (2006). Developing an engineering shape benchmark for CAD models. *Computer-Aided Design*, *38*(9), 939–953.

Kang, S. B., & Ikeuchi, K. (1993). The complex EGI: A new representation for 3D pose determination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *15*(7), 707–721.

Katz, S., & Tal, A. (2003). Hierarchical mesh decomposition using fuzzy clustering and cuts. *ACM Transactions on Graphics* 954–961.

Kazhdan, M., Funkhouser, T., & Rusinkiewicz, S. (2003). Rotation invariant spherical harmonic representation of 3D shape descriptors. In *Proc. of symposium on geometry processing*, Jun. 2003.

Khotanzad, A., & Hong, Y. H. (1990). Invariant image recognition by Zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*(5), 489–497.

Kriegel, H.-P., Kroeger, P., Mashael, Z., Pfeifle, M., Poetke, M., & Seidl, T. (2003). Effective similarity search on voxelized cad objects. In *Proc. of IEEE eighth international conference on database systems for advanced applications*. Washington, DC, USA.

Lindstrom, P., & Turk, G. (2000). Image-driven simplification. *ACM Transactions on Graphics*, *19*(3), 204–241.

Liu, Y., Zha, H., & Qin, H. (2006). The generalized shape distributions for shape matching and analysis. In *Proc. of the IEEE international conference on shape modeling and applications (SMI2006)*. Matsushima, Japan.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2).

Mademlis, A., Daras, P., Axenopoulos, A., Tzovaras, D., & Strintzis, M. G. (2008a). Combining topological and geometrical features for global and partial 3D shape retrieval. *IEEE Transactions on Multimedia*, *10*(5), 819–831.

Mademlis, A., Daras, P., Tzovaras, D., & Strintzis, M. G. (2008b). 3D object retrieval based on resulting fields. In *29th international conference on EUROGRAPHICS 2008, workshop on 3D object retrieval*. Crete, Greece, April 2008.

Mahmoudi, S., & Daoudi, M. (2002). 3D models retrieval by using characteristic views. In *ICPR'02* (pp. 11–15). Quebec, Canada, Aug. 2002.

Napoleon, T., Adamek, T., Shmitt, F., & O'Connor, N. E. (2008). SHREC'08 entry: Multi-view 3D retrieval using multi-scale contour representation. In *Proc. IEEE SMI 2009*.

Novotni, M., & Klein, R. (2003). 3d Zernike descriptors for content based shape retrieval. In *Proc. of the eighth ACM symposium on solid modeling and applications* (pp. 216–225). NY, USA, 2003. New York: ACM.

Ohbuchi, R., Otagiri, T., Ibato, M., & Takei, T. (2002). Shape-similarity search of three-dimensional models using parameterized statistics. In *Oroc. of Pacific graphics* (pp. 265–274). Beijng, China. Los Alamitos: IEEE Comput. Soc.

Ohbuchi, R., Minamitani, T., & Takei, T. (2003a). Shape similarity search of 3D models by using enhanced shape functions. In *Proc. TP.CG. 03* (pp. 97–104).

Ohbuchi, R., Nakazawa, M., & Takei, T. (2003b). Retrieving 3D shapes based on their appearance. In *Proc. ACM workshop on multimedia information retrieval (MIR) 2003* (pp. 39–46).

Ohbuchi, R., Osada, K., Furuya, T., & Banno, T. (2008). Salient local visual features for shape-based 3D model retrieval. In *Proc. of the IEEE international conference on shape modeling and applications (SMI 2008)* (pp. 93–102).

Osada, R., Funkhouser, T., Chazelle, B., & Dobkin, D. (2001). Matching 3D models with shape distributions. In *Proc. of the IEEE international conference on shape modeling and applications (SMI2001)* (pp. 154–166).

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.

Pu, J., & Ramani, K. (2005). An approach to drawing-like view generation from 3D models. In *Proc. of IDETC/CIE 2005*. ASME.

Real-time 3D models, http://www.3drt.com/.

Shih, J.-L., Lee, C.-H., & Wang, J. T. (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 *Proceedings of the shape modeling international (SMI '04)* (pp. 167–178). Genova, Italy, June 2004.

Tal, A., & Zuckerberger, E. (2006). Mesh retrieval by components. *International Conference on Computer Graphics Theory and Applications*, 142–149.

Tangelder, J., & Veltkamp, R. C. (2004). A survey of content based 3D shape retrieval methods. In *Proc. of IEEE shape modelling international* (pp. 145–156).

Teague, M. R. (1979). Image analysis via the general theory of moments. *Journal of Optical Society of America*, *70*, 920–930.

The VICTORY 3D Search Engine, http://www.victory-eu.org:8080/victory/results/search.html.

Tsatsaias, V., Daras, P., & Strintzis, M. G. (2007). 3D protein classification using topological, geometrical and biological information. In *Proc. of IEEE international conference on image processing (ICIP 2007)*. San Antonio, Texas, USA, 2007.

Tung, T., & Schmitt, F. (2004). Augmented reeb graphs for content-based retrieval of 3D mesh models. In *Proc. of the IEEE international conference on shape modeling and applications (SMI2004)* (pp. 157–166).

Tung, T., & Schmitt, F. (2005). The augmented multiresolution Reeb graph approach for content-based retrieval of 3D shapes. *International Journal of Shape Modeling (IJSM)*, *11*(1).

Vinanco, A. P., Ramirez, A. M., & Agustin, F. G. (2003). Digital image reconstruction by using Zernike moments. In *Proc. of SPIE* (pp. 281–289). Barcelona, Spain, Sept. 2003.

Vranic, D. V. (2003). An improvement of rotation invariant 3d-shape based on functions on concentric spheres. In *Proc. of IEEE ICIP* (3) (pp. 757–760).

Vranic, D. (2004). *3d model retrieval*. Ph.D. Dissertation, University of Leipzig.

Vranic, D., & Saupe, D. (2002). Description of 3d-shape using a complex function on the sphere. In *Proc. of the IEEE international conference on multimedia and Expo (ICME2002)* (Vol. 1, pp. 177–180).

Vranic, D., Saupe, D., & Richter, J. (2001). Tools for 3d-object retrieval: Karhunen-loeve transform and spherical harmonics. In *Proc. of the IEEE fourth workshop on multimedia signal processing* (pp. 293–29).

Wahl, E., Hillenbrand, U., & Hirzinger, G. (2003). Surflet-pair-relation histograms: A statistical 3D-shape representation for rapid classification. In *Proc. 3DIM 2003* (pp. 474–481).

Weyrich, T., Pauly, M., Keiser, R., Heinzle, S., Scandella, S., & Gross, M. (2004). Post-processing of scanned 3d surface data. In *Proc. of Eurographics*. Granada, Spain, 2004.

Yap, P. T., Paramesran, R., & Ong, S. H. (2003). Image analysis by Krawtchouk moments. *IEEE Transactions on Image Processing* *12*(11), 1367–1377.

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*. doi:10.1155/2007/23912. Volume 2007, Article ID 23912, 14 pages.

Zhang, D., & Lu, G. (2002). Shape-based image retrieval using generic Fourier descriptor. *ELSEVIER Signal Processing: Image Communication*, *17*(10), 825–848.