Eurographics Workshop on 3D Object Retrieval (2011)
H. Laga, T. Schreck, A. Ferreira, A. Godil, I. Pratikakis, and R. Veltkamp (Editors)

SHREC'11 Track: Generic Shape Retrieval

H. Dutagaci[1], A. Godil[1], P. Daras[2], A. Axenopoulos[2], G. Litos[2], S. Manolopoulou[2], K. Goto[3], T. Yanagimachi[3], Y. Kurita[3], S. Kawamura[3], T. Furuya[3], R. Ohbuchi[3]

[1]National Institute of Standards and Technology, Gaithersburg, USA
[2]Informatics and Telematics Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece
[3]University of Yamanashi, Yamanashi-ken, Japan

Abstract

In this paper we present the results of the 3D Shape Retrieval Contest 2011 (SHREC'11) track on generic shape retrieval. The aim of this track is to evaluate the performance of 3D shape retrieval algorithms that can operate on arbitrary 3D models. The benchmark dataset consists of 1000 3D objects classified in 50 categories. The 3D models are mainly classified based on visual shape similarity and each class has equal number of models to reduce the possible bias in evaluation results. Two groups have participated in the track with six methods in total.

Categories and Subject Descriptors (according to ACM CCS): I.5.4 [Pattern Recognition]: Applications—Computer vision; H.3.3 [Computer Graphics]: Information Systems—Information Search and Retrieval

1. Introduction

In this paper, we report the results of six 3D retrieval algorithms tested in the generic shape retrieval track of SHREC 2011, held in conjunction with the fourth Eurographics Workshop on 3D Object Retrieval. The Generic 3D Shape Benchmark, we provided for this track, is suitable for 3D retrieval algorithms that do not require any assumption on the query and target models, besides being complete. Many of the models in large 3D repositories and those circulating in the Internet don't meet criteria such as being oriented, manifold or watertight triangular meshes. For some models, due to their nature, and modeling necessities, these criteria are impossible to meet, such as a CAD model of a car with many disconnected components and internal structures. Therefore, there is an ongoing need for effective retrieval algorithms which can operate on all kinds of complete 3D meshes, including those referred to as "polygon soups" [FMK*03].

2. Dataset

The dataset consists of 1000 models, acquired from major 3D repositories on the Internet. The dataset is based on two previous NIST efforts to establish a Generic 3D Shape Benchmark. 800 models are from the work described in [FGLW08], which were also used in the SHREC'09 track: Generic Shape Retrieval [GDA*09], and the 200 models are from the dataset of the SHREC'10 Track: Generic 3D Warehouse [VGD*10]. We have obtained permission to redistribute the models for research purposes. There are 50 classes defined with respect to their semantic categories (Table 1), and each class contains the same number of 3D models (20 models). This reduces the possible bias in evaluation results; i.e. one method performing better for certain types of models is not favored due to the higher number of models in that category. The file format to represent the 3D models is the ASCII Object File Format (*.off).

Table 1 gives the labels of the categories, most of which correspond to man-made objects. Mostly, the models are not acquired via scanning and reconstruction, but created using modeling tools such as CAD. Therefore, the resolution vary significantly among models, surface normals are not consistent among and within models, and the models are far from being manifold or watertight surfaces. This is typical for most of the models residing in online repositories such as Google 3D Warehouse.

3. The Task and Performance Evaluation

The participants submit a 1000×1000 distance matrix per method. The matrix gives the pairwise dissimilarity figures of all the possible model pairs in the dataset. Using the dissimilarity matrices provided by the participants, we

DOI: 10.2312/3DOR/3DOR11/065-069

Bird	Fish	NonFlyingInsect
FlyingInsect	Biped	Quadruped
ApartmentHouse	Skyscraper	SingleHouse
Bottle	Cup	Glasses
HandGun	SubmachineGun	Guitar
Mug	FloorLamp	DeskLamp
Sword	Cellphone	DeskPhone
Monitor	Bed	NonWheelChair
WheelChair	Sofa	RectangleTable
RoundTable	Bookshelf	HomePlant
Tree	Biplane	Helicopter
Monoplane	Rocket	Ship
Motorcycle	Car	MilitaryVehicle
Bicycle	Bus	ClassicPiano
Drum	HumanHead	ComputerKeyboard
TruckNonContainer	PianoBoard	Spoon
Truck	Violin	

Table 1: *50 classes of the generic dataset.*

based our evaluations on six standard metrics widely used by 3D model retrieval community: Precision-Recall curve, Nearest Neighbor (NN), First-Tier (FT), Second-Tier (ST), E-measure (E), and Discounted Cumulative Gain (DCT) [SMKF04].

4. Participants

Two groups have participated in SHREC'11 track on generic shape retrieval. The participants and their methods can be listed as follows:

- P. Daras, A. Axenopoulos, G. Litos, and S. Manolopoulou from Centre for Research and Technology Hellas, Greece, participated with two methods (described in Section 5): 1) COMBINED-CMVD-STT-DSR, 2) COMBINED-CMVD-STT-DSR-LE
- K. Goto, T. Yanagimachi, Y. Kurita, S. Kawamura, T. Furuya, and R. Ohbuchi from University of Yamanashi, Japan, participated with four methods (described in Section 6: 1) DSIFT, 2) E0VF, 3) EVF, 4) EVF-MR

5. 3D Object Retrieval combining View-based and Volumetric Information

The proposed unified framework is a combination of three 3D object retrieval approaches: the Compact Multi-View Descriptor (CMVD [DZTS09], the Spherical Trace Transform (STT) [ZDA*07] and the Depth-Silhouette-Radial-EXTent descriptor (DSR) [Vra04]. Moreover, two novel features are introduced in order to improve the retrieval performance: The first is a new method for rotation estimation and the second is a manifold learning approach based on Laplacian Eigenmaps. In the following subsections, a brief description of each method is given.

5.1. Rotation Estimation

During preprocessing, both CMVD and DSR require a rotation estimation step, since the 3D object may have an arbitrary orientation. In the proposed framework, a new Combined Pose Estimation (CPE) method is introduced, which intuitively merges the well-known Continuous PCA (CPCA) [Vra04] with plane symmetry and rectilinearity. It must be noted here that STT does not require rotation estimation since it is a rotation-invariant descriptor.

As a first step, CPCA is applied to the input 3D object to produce a first pose estimation. Then, the reflection symmetry for the three CPCA-coordinate planes $(0_{xy}, 0_{xz}, 0_{yz})$ is computed [CVB09]. If symmetry is observed in two or three coordinate planes, the transformation is kept as it is and the process terminates. In case symmetry is observed in only one or zero coordinate planes, then, the algorithm proceeds to a correction step based on rectilinearity [LRS10]. The outcome of this step is finally kept as the result of the CPE method.

5.2. Compact Multi-View Descriptor

After the pre-processing step, a set of uniformly distributed views are extracted. The viewpoints are chosen to lie at the 18 vertices of a regular 32-hedron. The 2D image types are binary images (black/white images). Three rotation-invariant functionals are applied to the views to produce the descriptors: 1) 2D Polar-Fourier Transform, 2) 2D Zernike Moments, and 3) 2D Krawtchouk Moments. A more detailed description of the extraction of these 2D functionals is available in [DZTS09]. The number of descriptors per view, ND is determined experimentally, and is equal to $N_D = N_{FT} + N_{Zern} + N_{Kraw}$, where $N_{FT} = 78$, $N_{Zern} = 56$, and $N_{Kraw} = 78$. The CMVD framework measures the distance between two 3D objects by summing up the L1-distances between the descriptors of their corresponding pairs of views.

5.3. Spherical Trace Transform

Every 3D object is expressed in terms of a binary volumetric function. In order to achieve translation invariance, the center of mass of the 3D object is calculated and the model is translated so that its center of mass coincides with the origin of the coordinates system. Scaling invariance is also accomplished by scaling the object in order to fit inside the unit sphere. Then, a set of concentric spheres is defined. For every sphere, a set of planes which are tangential to the sphere is also defined. Further, the intersection of each plane with the object's volume provides a spline of the object, which can be treated as a 2D image.

Next, 2D rotation invariant functionals F are applied to this 2D image, producing a single value. Thus, the result of these functionals when applied to all splines, is a set of functions defined on every sphere whose range is the results

of the functional. Finally, a rotation invariant transform T is applied on these functions, in order to produce rotation invariant descriptors. For the needs of the SHREC, the implemented functionals F are the 2D Krawtchouk moments, and the 2D Zernike Moments, while the T function is the Spherical Fourier Transform.

A more detailed description of the extraction of these descriptors is available in [ZDA*07]. The dimension of descriptor vectors is $N_{Zernike} = 1080$ for the descriptors based on the 2D Zernike Moments and $N_{Krawtchouk} = 1080$ for the descriptors based on the Krawtchouk 2D functional. For descriptor matching, the Minkowski L1 distance is computed for a pair of descriptor vectors.

5.4. Depth-Silhouette-Radial-Extent descriptor

The DSR descriptor was introduced in [Vra04]. It combines the Depth Buffer descriptor, the Silhouette descriptor and the Radialized Spherical Extent function.

As a preprocessing step, CPCA is applied to the 3D object. In order to extract the 2D views, the 3D object is projected perpendicularly on the coordinate hyperplanes. Three silhouette images and six depth buffer images are extracted. In the case of silhouette images, a 1D FFT transform is applied to the contour which approximates the silhouette. This descriptor is invariant to rotation of the 2D view. In the case of depth buffer images, a 2D FFT transform is applied to the depth image. In the Radialized Spherical Extent descriptor, the 3D model is described by a spherical function which decomposes the model into a sum of concentric shells and gives the maximal distance of the model from the center of mass as a function of angle and the radius of the equivalent shell. The spherical function is represented by spherical harmonics coefficients. The above descriptors are concatenated in order to form a single descriptor vector for each 3D object.

5.5. Manifold Learning based on Laplacian Eigenmaps

The overall dissimilarity between two 3D objects A and B is a weighted sum of the dissimilarities of each descriptor separately:

$$\begin{aligned} dis(A,B) \quad = \quad & w_{CMVD}\, dis_{CMVD}(A,B) \\ & + w_{STT}\, dis_{STT}(A,B) \\ & + w_{DSR}\, dis_{DSR}(A,B) \end{aligned} \tag{1}$$

where $w_{CMVD} = 0.5$, $w_{STT} = 0.2$, $w_{DSR} = 0.3$.

The $dis(A,B)$ for all pairs of models in the SHREC database were used to create the dissimilarity matrix. This dissimilarity matrix corresponds to the output of the combined method referred to as COMBINED-CMVD-STT-DSR.

The method referred to as COMBINED-CMVD-STT-DSR-LE has an additional step involving a manifold learning method based on Laplacian Eigenmaps, which is used to improve the retrieval accuracy. The method creates an adjacency matrix W as follows:

$$W_{ij} = \begin{cases} 1 & \textit{if } j \in k-\textit{neighbours of } i, \\ 0 & \textit{otherwise}\ . \end{cases} \tag{2}$$

The Laplacian Eigenmaps method [BN03] creates a feature space of lower-dimension, where each 3D object is represented by an l-dimensional point. In our case, the feature space is of dimension $l = 40$. Then the dissimilarity between two 3D objects of the database is calculated by applying L2 distance to their 40-dimensional feature vectors.

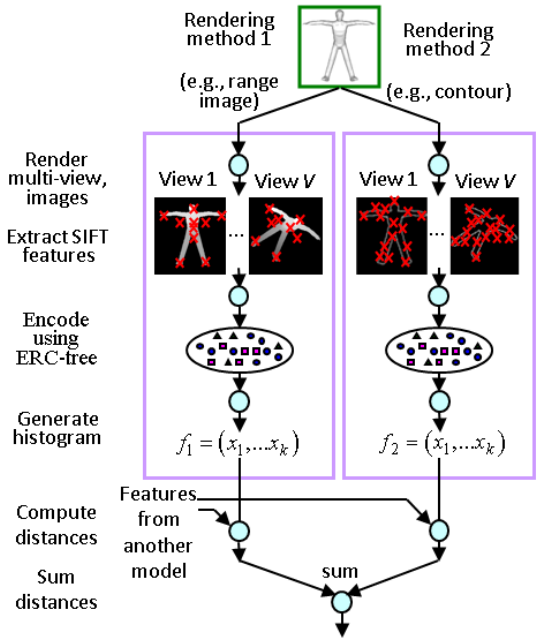

Figure 1: *The EVF is based on the BF-DSIFT [FO09]. Unlike BF-DSIFT, which used depth image only, the EVF uses more than one kind of rendering methods with the intet of extracting richer local, multi-scale, visual features.*

6. Expressive Visual Features (EVF)

The Expressive Visual Features (EVF) algorithm is an appearance-based method for comparing and retrieving 3D shapes. The EVF is based on the BF-DSIFT algorithm by Furuya et al [FO09].

The BF-DSIFT is designed for a diverse range of shape representations, including B-Rep solid, point set and polygon soup. So far as a shape representation can be rendered as surfaces, it can be compared. In addition, the BF-DSIFT is able to handle non-rigid or articulated shapes in addition to rigid shapes without any change. This is because a set of

local, multi-scale, rotation invariant visual features Scale Invariant Feature Transform (SIFT) [Low04] extracted from rendered images are integrated into a feature vector per 3D model by using bag-of-features method. The bag-of-features integration does not employ location information attached to each local feature.

While quite capable, the BF-DSIFT, sometimes fails to perform well on a database containing certain kind of shape models. One of the reasons is the interaction between the SIFT feature and depth images used for feature extraction. SIFT feature extracts scale, position, and orientation of change in gray level values in the image. A range image of a 3D model has clear and well defined contours, but often lacks features, that are, change in pixel values, inside the shape. The lack of feature inside a 3D model is exactly the reason to employ dense sampling at random locations on the image, instead of using the interest point detector of built into the original SIFT. Dense, random sampling forces SIFT features to be extracted from around the points at which no significant change in intensity is found. Even with the dense and random sampling of BF-DSIFT, however, there are features that are not easy to extract.

The EVF aims to alleviate the problem stated above by introducing multiple rendering schemes. The EVF renders a 3D model by using multiple rendering techniques for capturing rich and diverse set of features.

For SHREC 2011 Generic Shape Retrieval track, the depth and silhouette renderings are employed. The four different methods tested on the Generic Dataset are 1) DSIFT which corresponds to the algorithm described in [FO09], 2) E0VF which uses silhouette images only, 3) EVF which is a combination of silhouette and range images, and 4) EVF-MR, which uses manifold ranking [ZBL*04] on the EVF feature for the distance metric learning [OF10].

The EVF combines the two feature vector per 3D model by using distances computed from them, as indicated in the Figure 1. Note that the EVF (as well as the DSIFT included as a reference) are still based on bag-of local visual features, and retains the invariance against articulation and deformation. If we were to add a global feature, as in the 1SIFT described in [OF10], the retrieval performance could have been higher.

7. Results

Table 2 gives the scalar performance measures of the six methods. The COMBINED-CMVD-STT-DSR-LE method proposed by Daras et al. outperforms all the other methods with respect to the scalar measures other than the Nearest Neighbor. In terms of Nearest Neighbor, the COMBINED-CMVD-STT-DSR achieves the best performance. The COMBINED-CMVD-STT-DSR-LE method is followed by COMBINED-CMVD-STT-DSR.

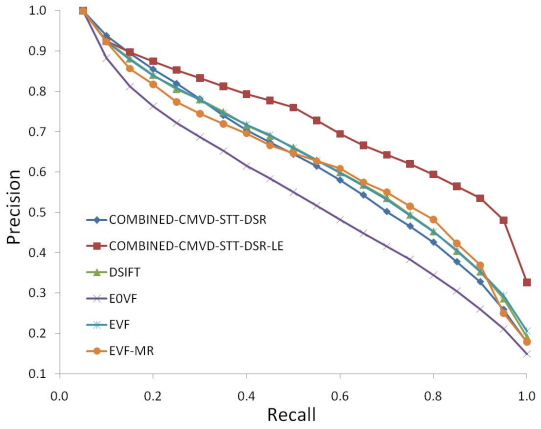

Figure 2: *Precision-Recall curves of the methods.*

In Figure 2, the precision-recall diagrams of all participating methods are presented. The COMBINED-CMVD-STT-DSR has similar performance to DSIFT and EVF methods. More specifically, the COMBINED-CMVD-STT-DSR is slightly better for recall values up to 0.3, while DSIFT and EVF are slightly better for recall values higher than 0.4. When the Laplacian Eigenmaps method is applied to the combination of CMVD, STT and DSR, the performance is significantly improved.

As for the SIFT-based methods, silhouette images (E0VF) by itself is outperformed by the original DSIFT, but the combination of silhouette and range images (EVF) is comparable to the DSIFT for this dataset and queries. For this benchmark, the EVF performed about as well as DSIFT. For other dataset, however, EVF could perform better.

8. Conclusions

In this paper, we have described and compared the performance of six algorithms submitted by two research groups that participated in this track. Based on all the performance evaluation measures, except for the NN, Daras et al.'s COMBINED-CMVD-STT-DSR-LE method performed best.

9. Acknowledgements

The work by Dutagaci and Godil from NIST, was supported by the SIMA program and the Shape Metrology IMS.

The work by Daras et al. from Informatics and Telematics Institute, Centre for Research and Technology Hellas, was supported by the EC project I-SEARCH (http://www.isearch-project.eu/).

PARTICIPANTS	METHODS	NN	FT	ST	E	DCG
Daras, Axenopoulos,	COMBINED-CMVD-STT-DSR	0.896	0.542	0.670	0.476	0.822
Litos and Manolopoulou	COMBINED-CMVD-STT-DSR-LE	0.889	0.661	0.771	0.558	0.862
	DSIFT	0.869	0.553	0.695	0.494	0.825
Goto, Yanagimachi, Kurita,	E0VF	0.812	0.470	0.594	0.420	0.764
Kawamura, Furuya, and Ohbuchi	EVF	0.870	0.551	0.691	0.493	0.824
	EVF-MR	0.877	0.569	0.667	0.483	0.806

Table 2: *Retrieval performance of the methods.*

References

[BN03] BELKIN M., NIYOGI P.: The 3D shape impact descriptor. *Neural Computation* (2003). 3

[CVB09] CHAOUCH M., VERROUST-BLONDET A.: Alignment of 3D models. *IEEE Transactions on Multimedia 71* (2009), 63–76. 2

[DZTS09] DARAS P., ZARPALAS D., TZOVARAS D., STRINTZIS M. G.: A 3D shape retrieval framework supporting multimodal queries. *SPRINGER, International Journal of Computer Vision* (2009). 2

[FGLW08] FANG R., GODIL A., LI X., WAGAN A.: A new shape benchmark for 3d object retrieval. In *Advances in Visual Computing*, vol. 5358 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2008, pp. 381–392. 1

[FMK*03] FUNKHOUSER T., MIN P., KAZHDAN M., CHEN J., HALDERMAN A., DOBKIN D., JACOBS D.: A search engine for 3D models. *ACM Transactions on Graphics 22*, 1 (Jan. 2003), 83–105. 1

[FO09] FURUYA T., OHBUCHI R.: Dense sampling and fast encoding for 3D model retrieval using bag-of-visual features. In *Proceeding of the ACM International Conference on Image and Video Retrieval* (2009), CIVR '09, pp. 26:1–26:8. 3, 4

[GDA*09] GODIL A., DUTAGACI H., AKGÜL C. B., AXENOPOULOS A., BUSTOS B., CHAOUCH M., DARAS P., FURUYA T., KREFT S., LIAN Z., NAPOLEON T., MADEMLIS A., OHBUCHI R., ROSIN P. L., SANKUR B., SCHRECK T., SUN X., TEZUKA M., VERROUST-BLONDET A., WALTER M., YEMEZ Y.: Shrec'09 track: Generic shape retrieval. In *3DOR* (2009), pp. 61–68. 1

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision 60* (November 2004), 91–110. 4

[LRS10] LIAN Z., ROSIN P., SUN X.: Rectilinearity of 3D Meshes. *International Journal of Computer Vision 89*, 2 (2010), 130–151. 2

[OF10] OHBUCHI R., FURUYA T.: Distance metric learning and feature combination for shape-based 3d model retrieval. In *Proceedings of the ACM workshop on 3D object retrieval* (2010), 3DOR '10, pp. 63–68. 4

[SMKF04] SHILANE P., MIN P., KAZHDAN M., FUNKHOUSER T.: The Princeton Shape Benchmark. In *Shape Modeling International* (2004). 2

[VGD*10] VANAMALI T., GODIL A., DUTAGACI H., FURUYA T., LIAN Z., OHBUCHI R.: Shrec'10 track: Generic 3d warehouse. In *3DOR* (2010). 1

[Vra04] VRANIC D.: *3D Model Retrieval*. PhD thesis, University of Leipzig, 2004. 2, 3

[ZBL*04] ZHOU D., BOUSQUET O., LAL T. N., WESTON J., SCHÖLKOPF B.: Learning with local and global consistency. In *Advances in Neural Information Processing Systems 16*, Thrun S., Saul L., Schölkopf B., (Eds.). MIT Press, Cambridge, MA, 2004. 4

[ZDA*07] ZARPALAS D., DARAS P., AXENOPOULOS A., TZOVARAS D., STRINTZIS M. G.: 3D model search and retrieval using the spherical trace transform. *EURASIP Journal on Advances in Signal Processing 2007* (2007). 2, 3