Eurographics Workshop on 3D Object Retrieval (2010)
I. Pratikakis, M. Spagnuolo, T. Theoharis, and R. Veltkamp (Editors)

SHREC'10 Track: Non-rigid 3D Shape Retrieval

Z. Lian[†,1,2], A. Godil[†,1], T. Fabry[3], T. Furuya[4], J. Hermans[3],
R. Ohbuchi[4], C. Shu[5], D. Smeets[3], P. Suetens[3], D. Vandermeulen[3], S. Wuhrer[5]

[1]National Institute of Standards and Technology, Gaithersburg, USA
[2]Beihang University, Beijing, PR China
[3]Katholieke Universiteit Leuven, Belgium
[4]University of Yamanashi, Japan
[5]National Research Council, Canada

Abstract

Non-rigid shape matching is one of the most challenging fields in content-based 3D object retrieval. The aim of the 3D Shape Retrieval Contest 2010 (SHREC'10) track on non-rigid 3D shape retrieval is to evaluate and compare the effectiveness of different methods run on a non-rigid 3D shape benchmark consisting of 200 watertight triangular meshes. Three groups with six methods have participated in this track and the retrieval performance was evaluated using six commonly-used metrics.

Categories and Subject Descriptors (according to ACM CCS): I.5.4 [Pattern Recognition]: Computer Vision—
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—

1. Introduction

After years of rapid development, 3D shape retrieval has become an important area of research in several fields including computer graphics, pattern recognition, machine vision, and so on. However, previous efforts have been mainly devoted to the retrieval of rigid 3D models, and thus how to efficiently and effectively compare non-rigid shapes is still a challenging problem in the field of content-based 3D object retrieval. As we know, non-rigid objects are commonly seen in our surroundings. Take Figure 1 for an example, a human being can appear in quite different postures, which might be identified as different kinds of objects using traditional rigid-shape analyzing techniques. Up to now, just a few methods have been reported to address this problem, generally, they utilized local features (e.g. [SF07] [OOFB08]), or global isometric-invariant features (e.g. [EK03] [RpSS]), or the combination of above two features (e.g. [GSCO07]) to generate 3D shape descriptors, that are invariant or approximately invariant under isometric transformations (e.g. bending and articulation).

To promote the investigation of non-rigid 3D shape retrieval, we organized the 3D Shape Retrieval Contest 2010(SHREC'10) track on non-rigid 3D shape retrieval. In this track, we provided a non-rigid shape database, that contains 200 watertight 3D meshes with randomized name indexes, based on the McGill Shape Benchmark [ZKCS05]. Each participant was asked to submit up to five dissimilarity matrices calculated using their methods in one week. Finally, six matrices were submitted by three groups, and then the retrieval performance was measured and compared using six standard evaluation metrics.

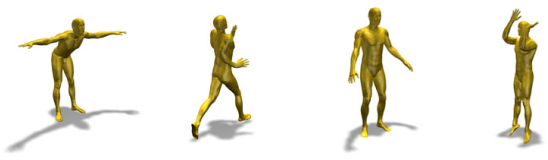


Figure 1: *Examples of non-rigid models.*

The remainder of the paper is organized as follows. Section 2 shows the data collection of a non-rigid shape benchmark. Section 3 indicates how to measure the retrieval per-

† Organizer of this SHREC track.

DOI: 10.2312/3DOR/3DOR10/101-108

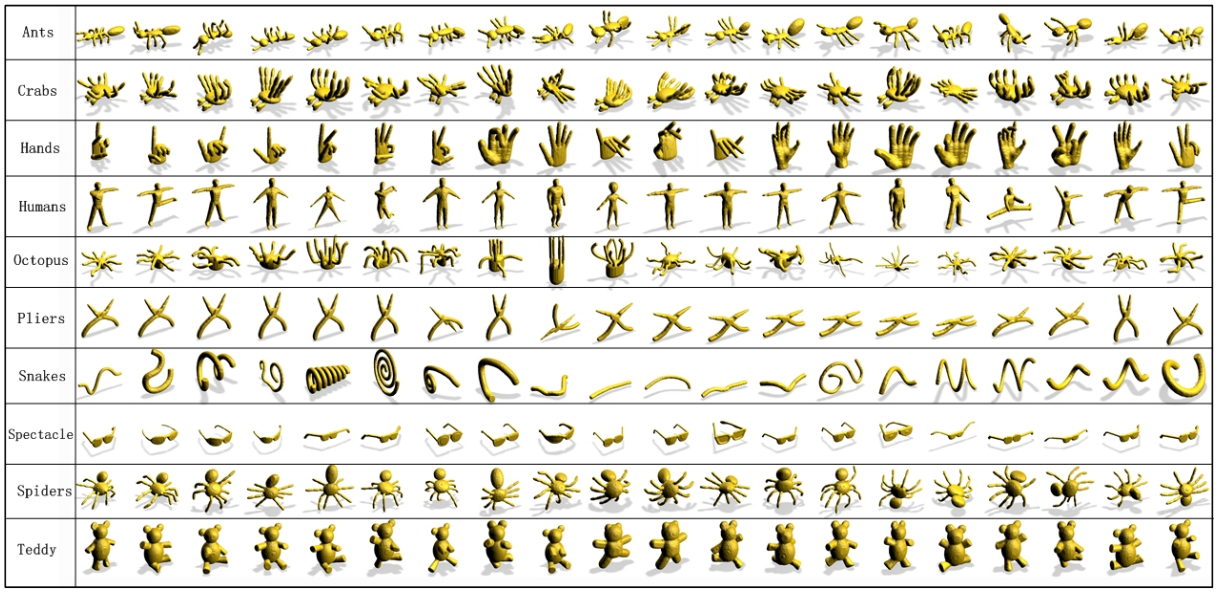


Figure 2: *Models of our database that is divided into 10 classes.*

formance based on the benchmark, while participants' information is described in Section 4. Afterwards, Section 5 discusses briefly the methods tested in this track and Section 6 demonstrates the evaluation results. Finally, the conclusion of this paper is provided in Section 7.

2. Data Collection

The database used in this track consists of 200 watertight 3D triangular meshes, which are selected from the McGill Articulated Shape Benchmark database [ZKCS05] [SZM*08]. Eleven of these original models are modified (or replaced by new models created by us) to make sure that all meshes in our database are watertight and they don't have any topology errors. We then equally classify these 200 models into 10 categories based on their semantic meanings, as shown in Figure 2, each class contains 20 objects with distinct postures.

3. Evaluation

Participants are asked to apply their methods to calculate the dissimilarity between every two objects and then generate corresponding dissimilarity matrices. Since our database contains 200 objects, the matrix is composed of 200×200 floating point numbers, where the number at position (i, j) represents the dissimilarity between models i and j.

Given a dissimilarity matrix, the retrieval performance of the method can be quantitatively evaluated based on the following statistics: 1) Nearest Neighbor (NN), 2) First Tier (FT), 3) Second Tier (ST), 4) E-measure (E), and 5) Discounted Cumulative Gain (DCG). In order to give the reader a more intuitive feeling about the results, precision-recall curves are also computed for each class as well as the whole database. More details about these standard evaluation metrics can be found in [SMKF04]. In fact, we directly used the reliable source code provided by the Princeton Shape Benchmark [SMKF04] to carry out the evaluation.

4. Participants

There are three groups participating in this track and, totally, six dissimilarity matrices have been submitted,

1. *MR-BF-DSIFT-E* and *BF-DSIFT-E* run by R. Ohbuchi and T. Furuya, from University of Yamanashi, Japan.
2. *DMEVD_run1*, *DMEVD_run2*, and *DMEVD_run3* run by D. Smeets, T. Fabry, J. Hermans, D. Vandermeulen and P. Suetens, from Katholieke Universiteit Leuven, Belgium.
3. *CF* run by S. Wuhrer and C. Shu, from National Research Council, Canada.

5. Methods

5.1. Bag-of Densely-Sampled Local Visual Features, by R. Ohbuchi and T. Furuya

The algorithm compares 3D shapes based on their appearances, that are, range images of the object rendered from multiple viewpoints. The algorithm is designed so that it could handle (1) a diverse range of shape representations, including polygon soup, point set, or B-rep solid, and (2) models having articulation or deformation.

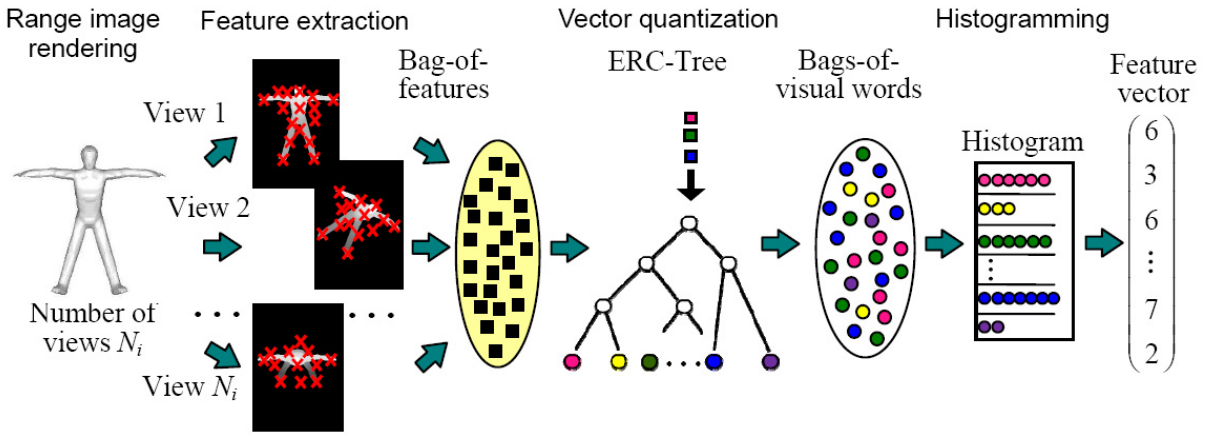


Figure 3: *Flow of the Bag-of-Feature Dense-SIFT with ERC-Tree* (BF-DSIFT-E) *algorithm.*

Appearance based comparison gives the algorithm its ability to handle diverse shape representation. Multiple-viewpoint rendering of dozens of range images coupled with (2D) rotation invariant image feature gives the algorithm its rotation invariance. Invariance to articulation and/or global deformation is achieved through the use of a set of multi-scale, local, visual features integrated into a feature vector per 3D model by using *Bag-of-Features (BoF)* approach. A feature vector per 3D model makes the cost of comparison among a pair of 3D models much lower than comparing sets of features, each consisting of thousands of local features. The algorithm is called *Bag-of- Features Dense-SIFT with Extremely randomized tree (BF-DSIFT-E)*. Please refer to the paper by Furuya et al. [FO09] for the details. While most 3D model retrieval algorithms deal with invariance to similarity transformation, very few algorithms achieve invariance to articulation while being able to accept 3D models having diverse set of 3D shape representations.

Figure 3 shows the processing flow of the BF-DSIFT- E algorithm. After normalizing the model for its position and scale, dozens of range images of the model are generated by using multiple virtual cameras looking inward at the model sitting at the coordinate origin. From each range image, our algorithm densely and randomly samples a few hundreds of local, multiscale image feature *Scale Invariant Feature Transform (SIFT)* by David Lowe [Low04]. Salient point detector of the SIFT is disabled for the dense sampling. A SIFT feature, typically having 128 dimensions, encodes position, orientation, and scale of gray-scale gradient change of the image about the sample point.

A 3D model is rendered into 42 depth images, each one of which then is sampled at 300 or so random locations. Thus, a 3D model is described by a set of about 12k SIFT features. The set of thousands of visual features is integrated into a feature vector per model by using BoF (e.g., [CDF*04] [SZ03]). The BoF approach vector quantizes, or encodes, a SIFT feature into a representative vector, or a "visual word", using a previously learned codebook. Visual words are accumulated into a histogram, which is the feature vector for the 3D model. The optimal dimension of the histogram, that is, the dimension of BoF-integrated feature vector, depends on the database. The dimension of feature vector is experimentally chosen as about 30k for the non-rigid model.

To extract this many features quickly, a fast GPU-based implementation of SIFT algorithm called SiftGPU by Wu [Wu] is applied. The Extremely Randomized Clustering Tree, or ERC-Tree, by Guerts et al. [GEW06], is also used for both feature set clustering during codebook learning and for vector quantization of SIFT features. With a small penalty in retrieval performance, the ERC-Tree is much faster than k-means clustering during codebook learning and naive linear search during VQ.

To derive ranked list of retrieval results given a query, two methods are employed: simple distance computation using Kullback-Leibler Divergence (KLD), and a distance-metric learning approach called Manifold Ranking (MR) [ZWG*03] with a small modification. In the retrieval experiment, the version that used KLD is named BF-DSIFT-E, while the one used MR is named MR-BF-DSIFT-E. The KLD below performs well for comparing two histogram-based feature vectors x and y.

$$d_{KLD}(x,y) = \sum_{i=1}^{n} (y_i - x_i) ln \frac{y_i}{x_i}$$

The MR first estimates the distribution of features in a low-dimensional subspace, or "manifold" approximated by a mesh. The mesh is created based on the proximity of feature points, and its edges are weighted based on the distance

among the feature points. It then computes similarity among the features on the manifold using a process similarity solving a diffusion equation. A relevance rank is diffused from the source, that is, the query. At an equilibrium state, the concentration of the relevance rank at a feature point indicates its closeness to the query. KLD is utilized to form the manifold mesh on which the diffusion takes place.

5.2. Inelastic Deformation Invariant Modal Representation, by D. Smeets, T. Fabry, J. Hermans, D. Vandermeulen and P. Suetens

In this section, a 3D shape recognition method which is invariant for inelastic deformations is proposed, while not requiring explicit point correspondences for shape comparison.

5.2.1. Inelastic deformation invariant representations

The 3D shapes are represented by inelastic deformation invariant matrices: the geodesic distance matrix (GDM) and the diffusion distance matrix (DDM). The former contains the geodesic distance between each pair of points on the surface. This distance is the length of the shortest path on the object surface between two points on the object. Geodesic distances are calculated by solving the Eikonal equation, $|\nabla T(\mathbf{P})| = 1$, on the surface, which can be achieved with a fast marching algorithm for triangulated meshes [PC08]. Isometric deformations, and thus also inelastic deformations, leave these geodesic distances unchanged.

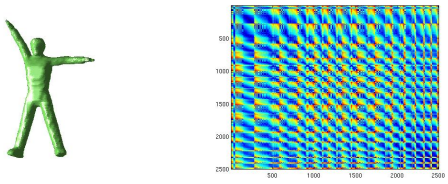


Figure 4: *The geodesic distance matrix as an elastic deformation invariant representation of the object shown left.*

The diffusion distance matrix contains the average diffusion distance between each pair of points on the surface. The average diffusion distance is the probability that a particle, started in one point, arrives at the other point after a diffusion process ran for a certain time t_D. This distance is calculated solving the heat equation, $\partial_t u = \alpha \Delta_X u$, on the surface. Practically this can be achieved by solving the generalized eigen decomposition problem of the discretized Laplace-Beltrami operator, $L\Phi = \lambda A\Phi$. A is a term proportional to the area of the triangles in the mesh and L is the laplacian of the surface [BBK*09].

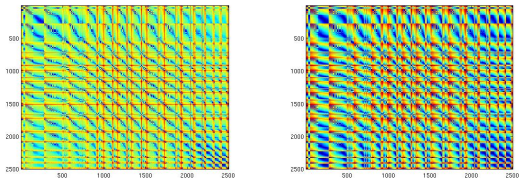


Figure 5: *The diffusion distance matrix as an elastic deformation invariant representation, after a diffusion time of $t_D = 200$ (left) and $t_D = 800$ (right).*

5.2.2. Sampling order invariant modal representations

Using the elastic deformation invariant distance matrices, a shape descriptor is needed that is invariant to the sampling order of points on the surfaces represented. It can be proved that the modal representation of these distance matrices, i.e. the singular value matrix, is invariant to the sampling order under the condition that each point on one surface has one corresponding point on the other surface [SFH*09]. Therefore, the surfaces are resampled such that each surface has the same number of points assuming a one to one correspondence. However, as the shape descriptor is sampling order invariant, this correspondence should not be known. As such, object recognition reduces to direct comparison of the shape descriptors without the need to establish explicit point correspondences. For computational reasons, only the k largest singular values are calculated. For the GDM, k is chosen to be 19, for the DDM 11.

5.2.3. Shape descriptor comparison

The modal representations of the GDM and DDM, calculated for $t_D = 800$, are then compared using the normalized Euclidean distance and fused together with the product rule.

5.3. Non-Rigid Shape Retrieval Using Canonical Forms, by S. Wuhrer and C. Shu

In this section, a method is presented that evaluates the dissimilarity between two Non-Rigid 3D models $S^{(0)}$ and $S^{(1)}$ represented by triangular meshes. The method proceeds by deforming the meshes into pose-invariant canonical forms $X^{(0)}$ and $X^{(1)}$ and by computing the dissimilarity between $S^{(0)}$ and $S^{(1)}$ using the Euclidean distance between $X^{(0)}$ and $X^{(1)}$.

Canonical forms were introduced by Elad and Kimmel [EK03]. The canonical form $X^{(i)}$ of a surface $S^{(i)}$ is the mapping of $S^{(i)}$ to $R^{(d)}$, such that the Euclidean distances between the mapped vertices approximate the geodesic distances between the original vertices well. Here, d is the dimension of the target space. We set $d = 3$ in our experiments. The canonical form is computed via multi-dimensional scaling with the geodesic distances between vertices on the

triangular manifold as dissimilarities. Hence, the canonical form of a non-rigid body is pose-invariant. The fast marching [KS98] is first used to compute geodesic distances on $S^{(i)}$, and then the least-squares multi-dimensional scaling [BG97] is applied to compute the canonical form in R^3. For increased space efficiency, the canonical form is computed using a coarse-to-fine strategy as outlined by Wuhrer et al. [WSAB07]. The approach by Wuhrer et al. consists of two steps. First, n' vertices of $S^{(i)}$ are used to compute a canonical form at low resolution. Second, the remaining vertices of $S^{(i)}$ are added to the canonical form one by one by minimizing a least-squares energy function. In our experiments, $n' = 500$.

To compute the dissimilarity δ, the average distance of points on $X^{(0)}$ to their nearest neighbors on $X^{(1)}$ is calculated and vice versa. More precisely, the dissimilarity is computed as,

$$\delta = \frac{\sum_{p \in X^{(0)}} ||N_{X^{(1)}}(p), p||}{n^{(0)}} + \frac{\sum_{p \in X^{(1)}} ||N_{X^{(0)}}(p), p||}{n^{(1)}}$$

where $n^{(i)}$ is the number of vertices of $X^{(i)}$, $p \in X^{(i)}$ means that p is a vertex of $X^{(i)}$, $N_{X^{(i)}}(p)$ is the vertex of $X^{(i)}$ that is closest to p , and $||.,.||$ denotes the Euclidean distance between two vertices. Kd-trees is applied to speed up the nearest neighbor search [AM93].

6. Results

In this section, we present the results of the three groups that submitted six runs, which are denoted as *MR-BF-DSIFT-E, BF-DSIFT-E, DMEVD_run1, DMEVD_run2, DMEVD_run3,* and *CF*, respectively, to this track. Given the six 200×200 dissimilarity matrices obtained using these methods, we carry out evaluations not only on the average performance of the whole database, but also on the result corresponding to each specific class. The performance measures adopted here are the five quantitative statistics *(i.e. NN, FT, ST, E, and DCG)* and the *Precision-recall plot* mentioned in Section 3.

Table 1 and Figure 6 show the retrieval performance of all six methods evaluated on the whole database. As we can see from Table 1 where quantitative retrieval statics are reported, these methods obtain good results on the given database, probably because they are all specifically designed for non-rigid 3D shapes. Considering the values of NN and DCG, Smeets's *DMEVD_run1* give the best results, while if we based the evaluation on FT, ST, and E-measure, Ohbuchi's MR-BF-DSIFT-E would take the first place. Similar observations can be made from Figure 6, where the precision-recall curves of the best two runs submitted by Ohbuchi et al. and Smeets et al. are both very close to the ideal horizontal line (i.e. the perfect result). Yet, it is hard to say clearly which method obtains the best performance.

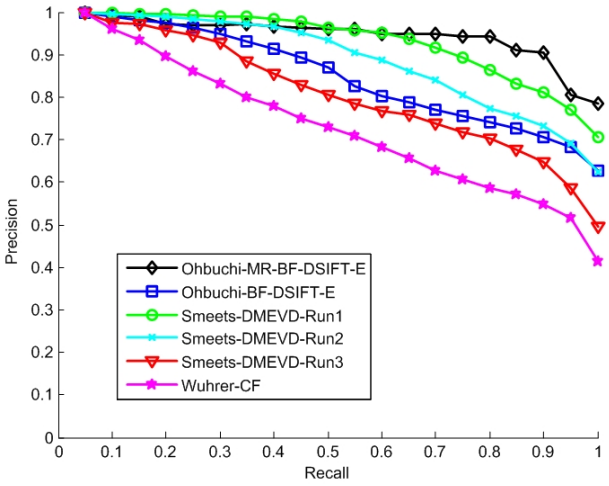


Figure 6: *Precision-recall curves of all runs evaluated for the whole database.*

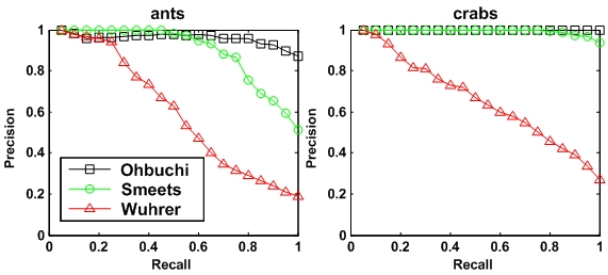


Figure 7: *Precision-recall curves of the best runs of each participant evaluated for two classes in the non-rigid database.*

Figure 7 and Figure 8 display the Precision-recall curves of the best runs (i.e. Ohbuchi's MR-BF-DSIFT-E, Smeets's DMEVD_run1, and Wuhrer's CF) of each group measured for every class in the non-rigid database. We observe that, none of these methods could guarantee better performance against others for all kinds of models. For example, Wuhrer's CF obtained the best result when searching for teddy bears, and Smeets's DMEVD_run1 clearly outperforms other methods considering the retrieval of spiders, while Ohbuchi's MR-BF-DSIFT-E is without doubt the best for hands, octopuses, and pliers. Generally speaking, algorithms developed by Ohbuchi et al. and Smeets et al. work considerably well for every class in this track, as their precision-recall curves all appear in the top right parts of these figures.

In addition to the calculation of these performance statistics, we have also created a web interface [NIS], which displays the retrieved models for all objects and methods, to analyze the results. Using the web interface, users are able

Table 1: *Retrieval performance of all runs evaluated using five standard measures evaluated for the whole database.*

PARTICIPANT	METHOD	NN	FT	ST	E	DCG
Ohbuchi, Ryutarou	BF-DSIFT-E	0.9800	0.7658	0.8924	0.6447	0.9409
	MR-BF-DSIFT-E	0.9850	0.9092	0.9632	0.7055	0.9763
Smeets, Dirk	DMEVD_run1	1.0000	0.8611	0.9571	0.7012	0.9773
	DMEVD_run2	0.9950	0.7884	0.9442	0.6796	0.9612
	DMEVD_run3	0.9600	0.7189	0.8505	0.6157	0.9203
Wuhrer, Stefanie	CF	0.9200	0.6347	0.7800	0.5527	0.8781

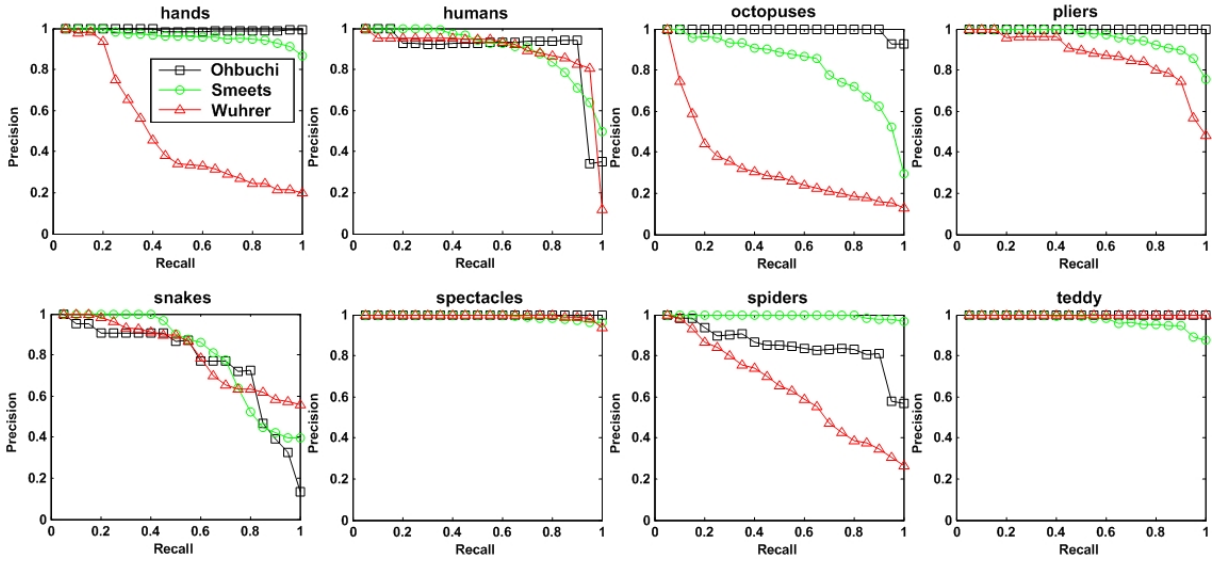


Figure 8: *Precision-recall curves of the best runs of each participant evaluated for other eight classes in the non-rigid database.*

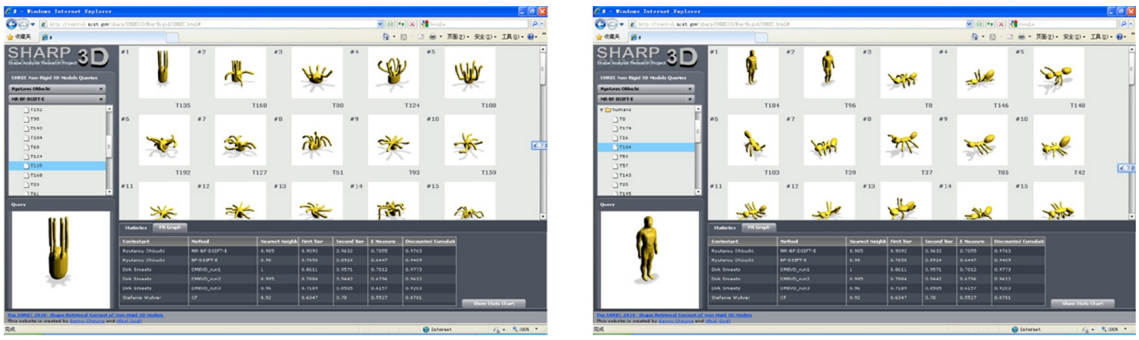
to examine the responses of the participants' methods to individual query models and visually inspect the effectiveness of the methods for every object in the database. Several retrieval examples are shown and compared in Figure 9, where two queries are provided to search for similar models using three different methods. We find that, no method always outperforms others for every query object and they all have their own strengths as well as weaknesses for particular models.

Base on the criterion of method classification described in Section 1, the participants' algorithms can be divided into two categories, 1) methods using local features: MR-BF-DSIFT-E and BF-DSIFT-E; 2) methods using global isometric-invariant features: DMEVD and CF. We speculate that the combination of these two kinds of invariant features could give better retrieval performance for non-rigid models. It should also be pointed out that, the MR-BF-DSIFT-E method apply an unsupervised machine learning algorithm (i.e. manifold ranking) which can be applied to enhance the results of any other approaches. However, the time consuming manifold ranking might not well suited for the retrieval application in large-scale databases.

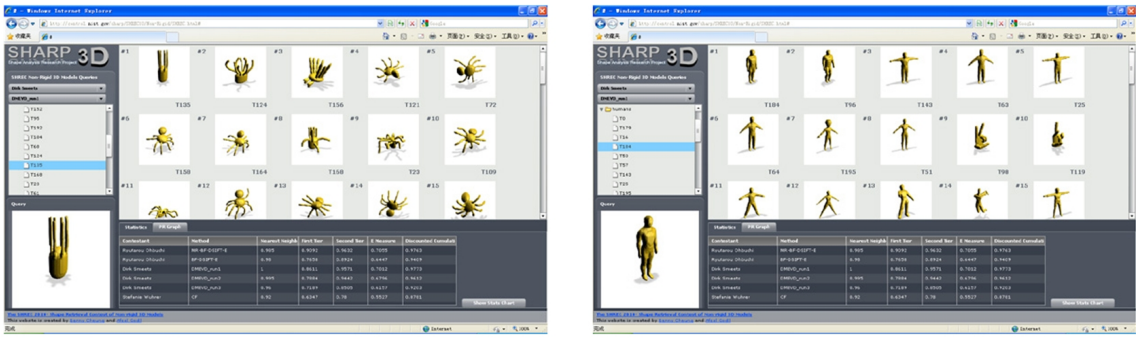
7. Conclusion

In this paper, we first described the database of non-rigid 3D shapes and then mentioned how to measure retrieval performance for the SHREC'10 track on non-rigid 3D shape retrieval. Afterwards, we briefly introduced the six methods used by three groups who participated in this track. Finally, experimental results were presented to compare the effectiveness of different algorithms.

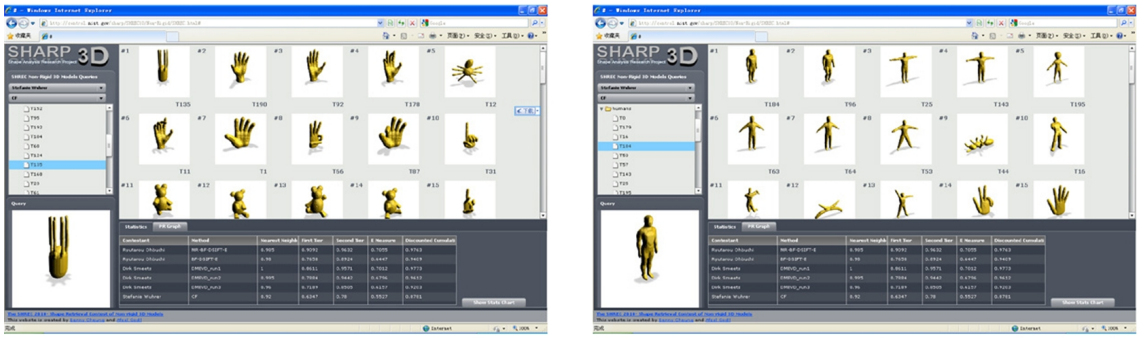
The non-rigid track organized this year is the first attempt in the history of SHREC to specifically focus on the performance evaluation of non-rigid 3D shape retrieval algorithms, and it is far from perfect. On the one hand, until recently just a few such kinds of practical methods have been reported, therefore, we only had three groups taking part in this track. On the other hand, because both creating and collecting large numbers of articulated watertight models are



(a) Two retrieval examples of Ohbuchi's MR-BF-DSIFT-E



(b) Two retrieval examples of Smeets's DMEVD_run1



(c) Two retrieval examples of Wuhrer's CF

Figure 9: *Retrieval examples of three methods using the web interface of the SHREC non-rigid track. The left bottom in the web interface displays the thumbnail of a query and the right top part shows retrieved models.*

not trivial tasks, we had to evaluate the participates' methods based on a small database consisting of 200 models. Hopefully, there will be more researchers working on non-rigid 3D shape retrieval in the future, and more groups will participate in the next SHREC non-rigid track, where a large-scale database could be available.

ACKNOWLEDGMENTS

This work has been supported by the SIMA and the IDUS program. We would like to thank the Shape Analysis Group in McGill University for providing the McGill 3D Shape Benchmark database, and we are also grateful to Benny Cheung for creating the web interface mentioned in this paper.

References

[AM93] ARYA S., MOUNT D. M.: Approximate nearest neighbor queries in fixed dimensions. In *ACM-SIAM Symposium on Discrete Algorithms* (1993), pp. 271–280.

[BBK*09] BRONSTEIN A. M., BRONSTEIN M. M., KIMMEL

R., MAHMOUDI M., SAPIRO G.: A Gromov-Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching. *International Journal of Computer Vision* (2009). published online.

[BG97] BORG I., GROENEN P.: *Modern Multidimensional Scaling Theory and Applications*. Springer, 1997.

[CDF*04] CSURKA G., DANCE C. R., FAN L., WILLAMOWSKI J., BRAY C.: Visual categorization with bags of keypoints. In *Proc. ECCV'04 workshop on Statistical Learning in Computer Vision* (2004), pp. 59–74.

[EK03] ELAD A., KIMMEL R.: On bending invariant signatures for surface. *IEEE Trans. Pattern Analysis and Machine Intelligence 25*, 10 (2003), 1285–1295.

[FO09] FURUYA T., OHBUCHI R.: Dense sampling and fast encoding for 3d model retrieval using bag-of-visual features. In *Proc. ACM CIVR 2009* (2009).

[GEW06] GUERTS P., ERNST D., WEHENKEL L.: Extremely randomized trees. *Machine Learning 36*, 1 (2006), 3–42.

[GSCO07] GAL R., SHAMIR A., COHEN-OR D.: Pose-oblivious shape signature. *IEEE Trans. Visualization and Computer Graphics 13*, 2 (2007), 261–271.

[KS98] KIMMEL R., SETHIAN J.: Computing geodesic paths on manifolds. In *Proc. the National Academy of Science* (1998), pp. 8431–8435.

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2 (2004), 91–110.

[NIS] NIST: The SHREC'10 Track on Non-rigid 3D Shape Retrieval. *http://www.itl.nist.gov/iad/vug/sharp/contest/2010/NonRigidShapes/*.

[OOFB08] OHBUCHI R., OSADA K., FURUYA T., BANNO T.: Salient local visual features for shape-based 3D model retrieval. In *Proc. IEEE international Conference on Shape Modeling and Applications (SMI'08)* (2008), pp. 93–102.

[PC08] PEYRÉ G., COHEN L. D.: Heuristically driven front propagation for fast geodesic extraction. *IJVCB 1*, 1 (2008), 55–67.

[RpSS] RUGGERI M. R., PATANE G., SPAGNUOLO M., SAUPE D.: Spectral-driven isometry-invariant matching of 3D shapes. *International Journal of Computer Vision (in press)*.

[SF07] SHILANE P., FUNKHOUSER T.: Distinctive regions of 3d surfaces. *ACM Transactions on Graphics 26*, 2 (2007).

[SFH*09] SMEETS D., FABRY T., HERMANS J., VANDERMEULEN D., SUETENS P.: Isometric deformation modelling for object recognition. In *CAIP '09: Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns* (2009), Jiang X., Petkov N., (Eds.), vol. 5702 of *LNCS*, pp. 757–765.

[SMKF04] SHILANE P., MIN P., KAZHDAN M., FUNKHOUSER T.: The princeton shape benchmark. In *Proc. Shape Modeling Applications (SMI'04)* (2004), pp. 167–178.

[SZ03] SIVIC J., ZISSERMAN A.: Video google: A text retrieval approach to object matching in videos. In *Proc. ICCV 2003* (2003), vol. 2, pp. 1470–1477.

[SZM*08] SIDDIQI K., ZHANG J., MAXRINI D., SHOKOUFANDEH A., BOUIX S., DICKINSON S.: Retrieving articulated 3d models using medial surfaces. *Machine Vision and Applications 19*, 4 (2008), 261–274.

[WSAB07] WUHRER S., SHU C., AZOUZ Z. B., BOSE P.: Posture invariant correspondence of incomplete triangular manifolds. *International Journal of Shape Modeling 13*, 2 (2007), 139–157.

[Wu] WU C.: SiftGPU: A GPU implementation of david lowe's scale invariant feature transform (SIFT). *http://cs.unc.edu/ ccwu/siftgpu/*.

[ZKCS05] ZHANG J., KAPLOW R., CHEN R., SIDDIQI K.: The mcgill shape benchmark. *http://www.cim.mcgill.ca/ shape/benchMark/* (2005).

[ZWG*03] ZHOU D., WESTON J., GRETTON A., BOUSQUET O., SCHOLKOPF B.: Ranking on data manifolds. In *Proc. the Conference on Advances in Neural Information Processing Systems (NIPS)* (2003).