Eurographics Workshop on 3D Object Retrieval (2013)
S. Biasotti, I. Pratikakis, U. Castellani, T. Schreck, A. Godil, and R. Veltkamp (Editors)

SHREC'13 Track: Retrieval of Objects Captured with Low-Cost Depth-Sensing Cameras

J. Machado[†1], A. Ferreira[†1], P. B. Pascoal[†1,2], M. Abdelrahman[6], M. Aono[3],

M. El-Melegy[6], A. Farag[6], H. Johan[5], B. Li[4], Y. Lu[4], A. Tatsuma[3]

[1]INESC-ID/IST/Technical University of Lisbon
[2]Microsoft Language Development Center, Portugal
[3]Toyohashi University of Technology, Japan
[4]Department of Computer Science, Texas State University, San Marcos, USA
[5]Fraunhofer IDM@NTU, Singapore
[6]Computer Vision and Image Processing Laboratory, University of Louisville, Louisville, USA

Abstract

The SHREC'13 Track: Retrieval of Objects Captured with Low-Cost Depth-Sensing Cameras is a first attempt at evaluating the effectiveness of 3D shape retrieval algorithms in low fidelity model databases, such as the ones captured with commodity depth cameras. Both target and query set are composed by objects captured with a Kinect camera and the objective is to retrieve the models in the target set who were considered relevant by a human-generated ground truth. Given how widespread such devices are, and how easy it is becoming for an everyday user to capture models in his household, the necessity of algorithms for these new types of 3D models is also increasing. Three groups have participated in the contest, providing rank lists for the set of queries, which is composed of 12 models from the target set.

Categories and Subject Descriptors (according to ACM CCS): H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—Relevance feedback I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Geometric algorithms, languages, and systems

1. Introduction

The advent of low-cost scanners in the consumer market, such as the Microsoft Kinect, has made this technology available to the everyday user and is fast becoming a staple in many households. While designed for a different purpose, such devices have proven able to digitize 3D objects in real time with acceptable quality [NIH*11], at least considering a myriad of contexts where before the presence of 3D capturing devices was virtually null. As a result, the proliferation of 3D models on the Internet is growing and expected to keep on that path as new and innovative ways of capturing and sharing 3D information are trusted to develop in the future.

Up to this moment, little research has been made regarding the retrieval of 3D models captured with commodity depth sensing cameras, being this the first attempt at such an endeavor in the Shape REtrieval Contest (SHREC). Previous versions of SHREC had their evaluations mostly focused on well-defined geometric or semantic classification of objects contained in the dataset, along with their ground-truth.

In this track, we propose a method based on the human classification of the original set of objects, using the real models that are used for the retrieval contest. We pretended to test the algorithms against real, although subjective, human expectations of the queries they were presented with.

† Track organizers

DOI: 10.2312/3DOR/3DOR13/065-071

Therefore, there are some challenges that must be addressed which are prone to skew the results of retrieval approaches. The first difficulty is the aforementioned subjectivity of the human evaluation. A second is the low degree of accuracy low-cost depth cameras present [KE12], which can be thousands lower when compared to some of the more expensive range scanners available today.

2. The Dataset

Our dataset is composed of 192 scanned models, which were acquired through the real-time capture of 224 collected objects. Of these, 32 were rejected due to low quality or material incompatibility. The range images were captured using a Microsoft Kinect camera (Fig. 2) and the ReconstructMe software for image capture. Some post-processing was done to extract the meshes and make them watertight (Fig. 3). The collection is presented in three different ASCII file formats: PLY, OFF and STL, representing the scans in a single triangular mesh.

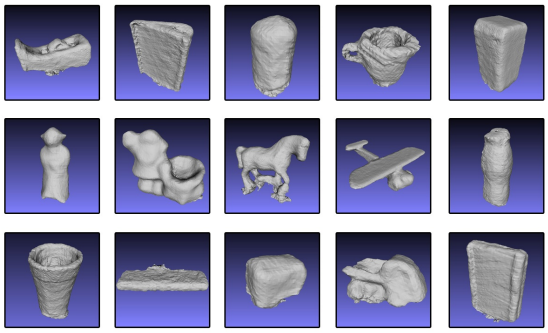

Figure 1: *Sample from the collection*

2.1. Target Set

The target database is composed by the collection of 192 models with varying degrees of accuracy over the original respective objects. Of these, those with higher degree of unique features tend to present much better digitizations. Samples can be seen in Figures 1 and 4.

The collection itself is uncategorized and the objects were collected with unrestricted regulations. All of these were kindly lent by 26 distinct collaborators from their households to suit the track's theme of ubiquity. The dataset, along with other details on the collection is available at http://3dorus.ist.utl.pt/research/BeKi/.

2.2. Query Set

The query set is simply a subset of the target set. It features 12 significantly distinct models to which we constructed a human-generated ground truth in a series of user tests. The

Figure 2: *Capture setup*

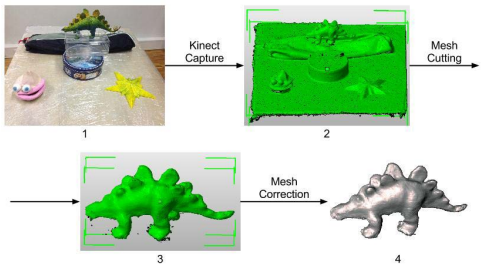

Figure 3: *Capture process*

query objects can be seen in the Figure 4. To evaluate the user agreement, we calculated the kappa statistic [F*71] for our tests. Since the users were asked to retrieve the top results amongst the complete database, the general agreement is overwhelmingly high, given the high rate of accordance on non-relevant retrievals ($\kappa \approx 0.995\%$). Therefore, we track only the relevant category of the results, which can be seen in Table 1. Queries with higher percentage of agreement are naturally expected to yield better retrieval results.

Table 1: κ *values per query:*

Query	κ	Query	κ
17	41,38%	117	62,68%
52	47,97%	145	74,45%
55	65,83%	160	56,31%
64	48,32%	172	73,08%
83	44,26%	200	53,91%
100	60,97%	202	54,75%
Average κ	56,99%		

3. Evaluation

All participants submitted, for the requested queries, at least one rank listing (one for each run). Each rank list has the length of the size of the target database. We employed the following evaluation measures on the results: Nearest Neighbor (**NN**), First-Tier (**FT**), Second-Tier (**ST**) and Discounted

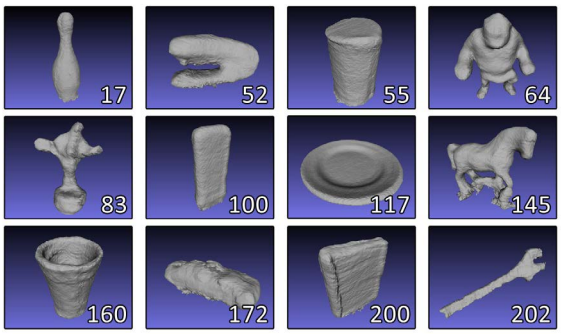

Figure 4: *Query list*

Cumulative Gain (**DCG**) [SMKF04]. These measures are based on the Precision and Recall evaluations of the queries and were chosen to give a general overview of the proposed methods in this first approach to this problem. As an additional visual indicator, the precision-recall curves were plotted as well.

Precision quantifies the ratio of retrieved models that are relevant to the search. For example, for a given search that returns 6 valid results in the first 12, the precision is 0.5 or 50%. **Recall** represents the ratio of relevant results retrieved against the total of valid results. For the previous query, if the class size is 10, the Recall would be 0.6 or 60%. The NN, FT and ST evaluations try the recall at different search depths. The Discounted Cumulative Gain is a measure that effectively grades the relevance of a result according to its position in the retrieval list. Top relevant results have a higher gain than models retrieved in a lower position.

4. Submissions

For this contest, three different groups participated with their respective methods.

- A. Tatsuma and M. Aono from the Toyohashi University of Technology have participated with a shape feature called Local Feature Correlation Descriptor (LCoD), producing just one run.
- B. Li, Y. Lu (Texas State University) and H. Johan from Nanyang Technological University present several approaches on Hybrid Shape Descriptors largely based on the ZFDR [LJ13]. They submitted five sets of lists each using a different combination of features: 1) ZFDR, 2) ZF, 3) ZFD, 4) ZFR and 5) ZFDSR.
- M. Abdelrahman, M. El-Melegy and A. Farag from the University of Louisville consider the 3D models captured with a commodity low-cost depth scanner as non-rigid, deformed objects, and propose an approach based on Scale Invariant Heat Kernels (SI-HKS) [BK10] for which they have submitted one run.

4.1. Local Feature Correlation Descriptor (LCoD)

A. Tatsuma and M. Aono propose a new 3D shape feature called Local Feature Correlation Descriptor (LCoD). The overview of how the method defines the proposed **LCoD** is illustrated in Figure 5. They developed this algorithm on the premise that in the field of image classification, the methods that consider high-order statistics of local features obtain a higher accuracy [PSM10, PG11]. Based on that, they expected that the shape feature based on the correlation of local features achieves high search performance. **LCoD** consists of the correlation of the local features extracted from depth-buffer images.

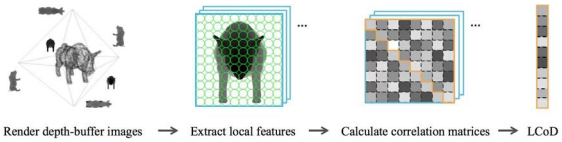

Figure 5: *Overview of the Local Feature Correlation Descriptor (LCoD).*

In **LCoD**, the first step is pose normalization, since 3D objects are usually defined by different authors with distinct authoring tools, which makes the position, size, and orientation of 3D objects quite different from each other. To solve this problem, they used their own [TA09] Point SVD that aligns the centroid and principal axes by generating random points on the surface of 3D shape objects, and Normal SVD that aligns the surface normals with respect to principal axes. In **LCoD**, a combination of Point SVD and Normal SVD is adopted for pose normalization.

Once pose normalization is done, the 3D object is enclosed within a regular octahedron where, from each vertex and midpoint of each edge, a depth-buffer image rendering with 256×256 resolution is performed. Note that a total of 18 viewpoints are defined.

After image rendering, Scale Invariant Feature Transform (SIFT) features [Low04] are extracted as local features from each depth-buffer image, and regular dense sampling [LP05] is employed on the interest point detection. **SIFT** features are extracted from 80×80 pixel patches arranged every 8 pixels.

A. Tatsuma and M. Aono then calculate the correlation matrix of local features for each depth-buffer image. Let $I_1, \ldots, I_{18}$ be 18 depth-buffer images rendered from the 3D object, and $\mathbf{v}_i^{(m)} \in \mathbb{R}^d (i = 1, \ldots, n)$ be the d-dimensional local features extracted from a depth-buffer image I_m. The correlation matrix $R^{(m)}$ is obtained of local features as follows:

$$R^{(m)} = \frac{1}{n} \sum_{j,k=1}^{n} \mathbf{v}_j^{(m)} {\mathbf{v}_k^{(m)}}^T. \quad (1)$$

The vector $\mathbf{r}^{(m)}$ consists of concatenating the elements in the upper triangular part of the correlation matrix $R^{(m)}$:

$$\mathbf{r}^{(m)} = [R^{(m)}_{1,1}, \ldots, R^{(m)}_{1,d}, R^{(m)}_{2,2}, \ldots, R^{(m)}_{2,d}, \ldots, R^{(m)}_{d,d}]. \quad (2)$$

The vector $\mathbf{f}$ is generated, consisting of vector $\mathbf{r}^{(m)}$ calculated for each depth-buffer image:

$$\mathbf{f} = [\mathbf{r}^{(1)} \ldots, \mathbf{r}^{(18)}]^T. \quad (3)$$

Finally, to obtain the proposed **LCoD** feature, the vector $\mathbf{f}$ is normalized with the power-norm and the ℓ_2-norm [PSM10].

For **LCoD** similarity between two 3D objects, a simple calculation of inner product is required.

LCoD consists of concatenating the correlation matrix of the local features extracted from each depth-buffer image. This definition of **LCoD** leads to high dimensional shape feature. Since the dimension of **SIFT** extracted as a local feature is $d = 128$, the total dimension of **LCoD** becomes $18 \times (d(d+1)/2) = 148,608$.

4.2. ZFDR

3D models reconstructed from 3D images captured by low-cost cameras, such as Microsoft Kinect, are only approximate representations of real objects. The accuracy is highly dependent on the cameras and the 3D reconstruction algorithms employed. Therefore, compared to the 3D models in existing benchmarks, there are many errors in the geometrical properties of these models, such as normals, curvatures and connectivity. Topological errors are also easy to be found. Because of these issues, compared to view-based retrieval approaches, it will be relatively more challenging for many geometry-based and topology-based 3D model retrieval approaches to deal with the retrieval of these models. On the other hand, most view-based methods and many hybrid techniques are more robust to the errors in geometry or topology. Motivated by this, they mainly adopt a view-based approach to extract visual information-based features, such as Zernike moments, Fourier descriptors and 2D Fourier Transform coefficients features, to retrieve these models.

Their algorithms and the corresponding five runs are largely based on the hybrid shape descriptor **ZFDR** proposed in [LJ13], which comprises both visual and geometrical features of a 3D model: **Z**ernike moments and **F**ourier descriptor features of 13 sample silhouette views, **D**epth information of six depth buffer views, and **R**ay-based features of the model based on a set of ray-based feature vectors shooting from the center to the utmost intersections on the surface of the model. Based on **ZFDR** and for a comparative evaluation, they further test its three reduced versions: **ZF**, **ZFD** and **ZFR**, which will partially or completely reduce the contribution of geometrical features. **D** and **R** are two components of the hybrid shape descriptor **DESIRE** (also mentioned as **DSR**, that is **D**+**S**+**R**) proposed by Vranic [Vra04].

The third component **S** denotes the **S**ilhouette-based component shape descriptor which extracts 1D Fourier transform features of the three canonical silhouette views of a 3D model. Similarly, to compete with the above descriptors, they also test the shape descriptor **ZFDSR** which combines **ZF** and **DSR**. They graphically demonstrate their feature extraction process in Figure 6. Some details are mentioned below.

They normalize the 3D models by utilizing the Continuous Principle Component Analysis (CPCA) [Vra04] algorithm before feature extraction. Their cube-based view sampling approach samples 13 views for an aligned 3D model with CPCA by setting cameras on the 4 top corners, 3 adjacent face centers and 6 middle edge points of a cube. For each sample view, they compute 35 Zernike moments [KH90] in total and its first 10 centroid distance-based Fourier descriptors [ZL01]. They utilize the executable file [Vra04] to extract the features of **D**, **R** and **S**.

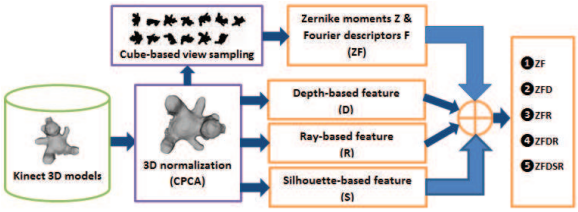

Figure 6: *Flowchart of computing five hybrid shape descriptors: **ZFDR**, **ZF**, **ZFD**, **ZFR** and **ZFDSR**.*

After obtaining the component shape descriptors **Z**, **F**, **D**, **R** and **S**, they assign appropriate distance metrics to measure the component distances d_Z, d_F, d_D, d_R and d_{DSR} between two models. These component distances are linearly combined accordingly to form five hybrid descriptor distances d_{ZFDR}, d_{ZF}, d_{ZFD}, d_{ZFR} and d_{ZFDSR}, which correspond to their five funs: **ZFDR**, **ZF**, **ZFD**, **ZFR** and **ZFDSR**. For more details about the feature extraction and retrieval processes, please refer to [LJ13] and [LGA*12].

4.3. Scale Invariant Heat Kernels (SI-HKS)

M. Abdelrahman, M. El-Melegy and A. Farag faced the contest by considering the models captured with a commodity low-cost depth scanner as deformed objects, which in itself is a challenging problem as it needs more work to compensate for the degrees of freedom resulting from local deformations. They quote Reuter et al [RWSN09] who used the Laplacian spectra as intrinsic shape descriptors, and employed the Laplace-Beltrami spectra as 'shape-DNA' or a numerical fingerprint of any 2D or 3D manifold (surface or solid). That publication proved that 'shape-DNA' is an isometry-invariant shape descriptor. Recently Sun et al. [SOG09] proposed heat kernel signatures (HKS)

as a deformation-invariant descriptors based on diffusion of multi-scale heat kernels. **HKS** is a point based signature satisfying many of the good descriptor properties, but suffers from sensitivity to scale. Bronstein et al [BK10] solved the **HKS** scale problem through a series of transformations. The same research group has recently introduced the Shape Google approach [BBGO11] based on the scaled-invariant **HKS**. The idea is to use **HKS** at all points of a shape, or alternatively at some shape feature points, to represent the shape by a Bag of Features (BoF) vector. Sparsity in the time domain is enforced by preselecting some values of the time.

In this work, the participants propose an approach for shape matching and retrieval based on scale invariant heat kernel signature (SI-HKS). Sun et al. [SOG09] proposed to use the **HKS** as a local shape descriptor

$$h(x,t) = H_t(x,x) = \sum_{k=1}^{\infty} e^{-\lambda_i t} \phi_i(x)^2 \quad (1)$$

where λ_i and ϕ_i are the eigenvalues and eigenfunctions of the Laplace-Beltrami operator. **HKS** has several desired properties [SOG09]: it is intrinsic and thus isometry-invariant (two isometric shapes have equal **HKS**), multi-scale and thus capture both local features and global shape structure, and also informative: under mild conditions, if two shapes have equal heat kernel signatures, they are isometric. The proposed descriptor in this work is based on BoF representation of the **HKS** in frequency domain combined with the first 15 normalized eigenvalues of the Laplace-Beltrami operator. The novelty introduced by the proposed method is to achieve scale-invariance of **HK** which is shown to be noise-robust.

Scale invariance is a desirable property of the shape descriptor, which can be achieved by many ways. A novel local scale normalization method is proposed based on simple operations. It was shown [BBGO11] that scaling a shape by a factor β results in changing $H(x,t)$ to $\beta^2 H(x,\beta^2 t)$. The participants propose to apply the Fourier transform (FT) directly

$$H'(w) = \beta^2 H(w) exp(j2\pi ws). \quad (2)$$

Then taking the amplitude of the FT,

$$|H'(w)| = \beta^2 |H(w)| \quad (3)$$

The effect of the multiplicative constant β^2 is eliminated by normalizing the $|H'(w)|$ by the sum of the amplitudes of the FT components. The amplitudes of the first significant FT components (normally 6) are employed to construct the scale-invariant shape descriptor. This proposed method eliminates the scale effect without having to use the noise-sensitive derivative operation or the logarithmic transformation that both were used in [BBGO11]. This method is simpler, more computational-efficient and more robust to noise. Eventually the classification is done with the L1-Norm.

5. Results

The three groups of participants of the SHREC'13 Retrieval of objects captured with low-cost depth-sensing cameras contest have submitted 7 sets of rank lists in total. The results for these submissions are summarized in Figure 7 and in the precision-recall curves in Figure 8. Figures 9, 10 and 11 shows the individual results for the **LCoD**, **ZFDR** and **SI-HKS** shape descriptors, respectively.

In the Local Feature Correlation Descriptor (LCoD), the participants use a view-based approach to the problem using Dense SIFT to perform the feature extraction, which seems to be an appropriate candidate. This descriptor proved to be averagely the most effective of the 3 submissions in all the evaluated retrieval measures. It is important to note that the results are being compared against potentially idiosyncratic evaluations by human subjects, so it stands to reason to conclude this approach is the one that best suited the human expectations for the results.

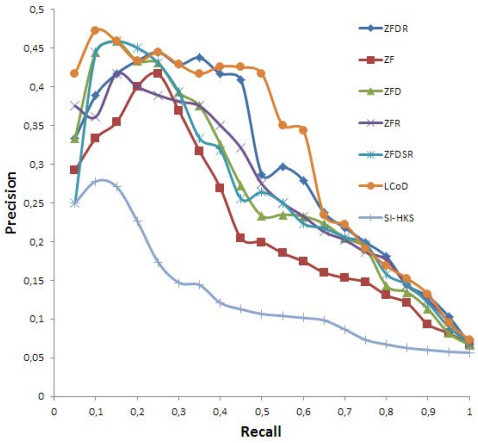

Figure 8: *Precision-recall curves*

For the hybrid approach, the participants submitted 5 different runs, composed by distinct linear combinations of their 5 shape descriptors, while every run included the Zernike moments and Fourier descriptor features. From these runs, **ZFDR** produced the best overall results while **ZF** had the lowest scores, which hints against underestimating the contribution of geometrical features in such approaches. Ray-based features of the models seem to also play an important part on the retrieval of these models, as **ZFR** comes close to **ZFDR** in the comparison. Interesting to note that the DCG outcomes are stable across the different implementations.

The Scale Invariant Heat Kernel Signatures (SI-HKS) presented the lowest average scores in all categories, save for queries 117 (plate) and 202 (wrench). These exceptions can be explained by the contents of the target set, which includes apparently scaled-down variants of the mentioned queries

Participant	Method	NN	FT	ST	DCG
A. Tatsuma & M. Aono	LCoD	**0,833333**	**0,458333**	**0,575**	**0,716436**
B. Li, Y. Lu and H. Johan	ZFDR	0,666667	0,416667	0,566667	0,688601
	ZF	0,583333	0,325	0,425	0,627619
	ZFD	0,666667	0,358333	0,483333	0,684041
	ZFR	0,75	0,383333	0,533333	0,664129
	ZFDSR	0,5	0,35	0,525	0,685519
M. Abdelrahman, M. El-Melegy and A. Farag	SI-HKS	0,5	0,191667	0,291667	0,524218

Figure 7: *Retrieval performances of the algorithms*

Query:	1-Tier	2-Tier	NN	DCG
17	0,6000	0,7000	0,0000	0,6311
52	0,1000	0,2000	1,0000	0,4672
55	0,2000	0,2000	1,0000	0,4983
64	0,4000	0,5000	0,0000	0,6614
83	0,4000	0,4000	1,0000	0,6059
100	0,4000	0,7000	1,0000	0,6904
117	0,4000	0,7000	1,0000	0,8101
145	0,5000	0,6000	1,0000	0,8438
160	0,7000	0,8000	1,0000	0,8674
172	0,6000	0,8000	1,0000	0,8766
200	0,4000	0,4000	1,0000	0,7432
202	0,8000	0,9000	1,0000	0,9016
Average	0,4583	0,5750	0,8333	0,7164

Figure 9: *Individual results for LCoD*

Query:	1-Tier	2-Tier	NN	DCG
17	0,5000	0,7000	0,0000	0,7088
52	0,1000	0,1000	1,0000	0,4447
55	0,2000	0,3000	0,0000	0,5409
64	0,3000	0,4000	1,0000	0,6407
83	0,2000	0,4000	1,0000	0,5901
100	0,4000	0,6000	1,0000	0,6448
117	0,5000	0,6000	1,0000	0,8744
145	0,6000	0,8000	1,0000	0,7756
160	0,6000	0,7000	1,0000	0,8661
172	0,5000	0,8000	0,0000	0,5748
200	0,5000	0,6000	0,0000	0,6993
202	0,6000	0,8000	1,0000	0,9031
Average	0,4167	0,5667	0,6667	0,6886

Figure 10: *Individual results for ZFDR*

(smaller plates and wrenches). This algorithm seems to work well in the context of transformations of non-rigid objects, which is not the case with this dataset, where every model is unique.

Although it would be logical to consider that the queries that yield better agreement among human judges, would also have slightly better results across runs, such fact could not be correlated with the results from this track. Comparing the numbers from Table 1 and Figures 10 and 11, a direct match

Query:	1-Tier	2-Tier	NN	DCG
17	0,1000	0,2000	0,0000	0,4078
52	0,1000	0,3000	1,0000	0,4890
55	0,1000	0,1000	0,0000	0,4263
64	0,1000	0,1000	1,0000	0,4617
83	0,0000	0,1000	0,0000	0,3262
100	0,1000	0,1000	0,0000	0,3686
117	0,6000	0,8000	1,0000	0,9045
145	0,2000	0,3000	0,0000	0,4837
160	0,0000	0,0000	0,0000	0,3000
172	0,3000	0,3000	1,0000	0,7117
200	0,1000	0,3000	1,0000	0,5359
202	0,6000	0,9000	1,0000	0,8754
Average	0,1917	0,2917	0,5000	0,5242

Figure 11: *Individual results for SI-HKS*

between agreement and algorithm performance can not be extrapolated and further study on this topic is required.

6. Conclusions

In this paper, we have described and compared the algorithms from each of the three different research groups that participated in the SHREC'13 Track: Retrieval of Objects Captured with Low-Cost Depth-Sensing Cameras. Each participant was presented with a subset of the target collection to pose as the query set, and asked to submit a full-depth list of results for each of their respective algorithms and possible variants.

While the levels of precision reached by these submissions are relatively low, that was to be expected, both by the subjectivity of the proposed ground truth, and by the lower quality of the dataset, when compared with all other existing 3D-shape benchmarks. The state of the art of low-cost depth scanners shows that, although promising in its potential of 3D model scanning, it still lacks a degree of accuracy that lets its results be usable for many different purposes. As it is, it seems more interesting to understand its usability as a fast scan and query device than as a benchmark modeler, as it was used for this track.

The method that demonstrated best overall performance was the Local Feature Correlation Descriptor (LCoD). From

the set of different configurations of Hybrid Descriptors presented, **ZFDR** had the best results in average, while **ZF** shows promising numbers while the search-depth is still low. Finally, the **SI-HKS** was able to match the previous algorithms for a small number of queries, while providing the worst average values overall. Generally, view-based and hybrid approaches seem to be better choices for 3D-shape retrieval of objects captured with low-cost depth sensing cameras than topological or geometrical feature algorithms.

This is just a first step into this topic of research. Other approaches can be considered, such as the retrieval of models in a larger and more accurate database, using full queries captured with low-cost depth cameras like the ones in this benchmark, or just range scans captured with these devices. Such work could provide grounds for the use of low-cost cameras in object retrieval and environment recognition in real-time settings.

Acknowledgements

The work described in this paper is supported by the following institutions:

- Partially by the Portuguese Foundation for Science and Technology (FCT) through the project 3DORuS, reference PTDC/EIA-EIA/102930/2008, and the INESC-ID multiannual funding, reference PEst-OE/EEI/LA0021/2011.
- Partially by the European Commission's Seventh Framework Programme, through project Marie Curie Golem (ref.251415, FP7-PEOPLE-2009-IAPP).
- The work of Bo Li, Yijuan Lu and Henry Johan is supported by the Texas State University Research Enhancement Program (REP), Army Research Office grant W911NF-12-1-0057, and NSF CRI 1058724 to Dr. Yijuan Lu.

References

[BBGO11] BRONSTEIN A. M., BRONSTEIN M. M., GUIBAS L. J., OVSJANIKOV M.: Shape google: Geometric words and expressions for invariant shape retrieval, Feb. 2011. URL: http://doi.acm.org/10.1145/1899404.1899405, doi:10.1145/1899404.1899405. 5

[BK10] BRONSTEIN M. M., KOKKINOS I.: Scale-invariant heat kernel signatures for non-rigid shape recognition. In *CVPR* (2010), IEEE, pp. 1704–1711. 3, 5

[F*71] FLEISS J., ET AL.: Measuring nominal scale agreement among many raters. *Psychological Bulletin 76*, 5 (1971), 378–382. 2

[KE12] KHOSHELHAM K., ELBERINK S. O.: Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors 12*, 2 (2012), 1437–1454. URL: http://www.mdpi.com/1424-8220/12/2/1437, doi:10.3390/s120201437. 2

[KH90] KHOTANZAD A., HONG Y.: Invariant image recognition by Zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence 12*, 5 (1990), 489–497. 4

[LGA*12] LI B., GODIL A., AONO M., BAI X., FURUYA T., LI L., LÓPEZ-SASTRE R. J., JOHAN H., OHBUCHI R., REDONDO-CABRERA C., TATSUMA A., YANAGIMACHI T., ZHANG S.: SHREC'12 track: Generic 3D shape retrieval. In *3DOR* (2012), Spagnuolo M., Bronstein M. M., Bronstein A. M., Ferreira A., (Eds.), Eurographics Association, pp. 119–126. 4

[LJ13] LI B., JOHAN H.: 3D model retrieval using hybrid features and class information. *Multimedia Tools Appl. 62*, 3 (2013), 821–846. 3, 4

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *International Journal Computer Vision 60*, 2 (Nov. 2004), 91–110. 3

[LP05] LI F.-F., PERONA P.: A bayesian hierarchical model for learning natural scene categories. In *Proceedingsedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2005), vol. 2 of *CVPR '05*, pp. 524–531. 3

[NIH*11] NEWCOMBE R. A., IZADI S., HILLIGES O., MOLYNEAUX D., KIM D., DAVISON A. J., KOHLI P., SHOTTON J., HODGES S., FITZGIBBON A.: Kinectfusion: Real-time dense surface mapping and tracking. In *Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality* (Washington, DC, USA, 2011), ISMAR '11, IEEE Computer Society, pp. 127–136. URL: http://dx.doi.org/10.1109/ISMAR.2011.6092378, doi:10.1109/ISMAR.2011.6092378. 1

[PG11] PICARD D., GOSSELIN P. H.: Improving image similarity with vectors of locally aggregated tensors. In *Proceedings of the 18th IEEE International Conference on Image Processing* (2011), pp. 669–672. 3

[PSM10] PERRONNIN F., SÁNCHEZ J., MENSINK T.: Improving the fisher kernel for large-scale image classification. In *Proceedings of the 11th European Conference on Computer Vision: Part IV* (2010), ECCV '10, pp. 143–156. 3, 4

[RWSN09] REUTER M., WOLTER F.-E., SHENTON M., NIETHAMMER M.: Laplace-beltrami eigenvalues and topological features of eigenfunctions for statistical shape analysis. *Computer-Aided Design 41*, 10 (2009), 739–755. URL: http://dx.doi.org/10.1016/j.cad.2009.02.007, doi:10.1016/j.cad.2009.02.007. 4

[SMKF04] SHILANE P., MIN P., KAZHDAN M., FUNKHOUSER T.: The Princeton shape benchmark. In *Shape Modeling International* (June 2004). 3

[SOG09] SUN J., OVSJANIKOV M., GUIBAS L.: A concise and provably informative multi-scale signature based on heat diffusion, 2009. URL: http://dl.acm.org/citation.cfm?id=1735603.1735621. 4, 5

[TA09] TATSUMA A., AONO M.: Multi-fourier spectra descriptor and augmentation with spectral clustering for 3D shape retrieval. *The Visual Computer 25*, 8 (2009), 785–804. 3

[Vra04] VRANIC D.: *3D Model Retrieval*. PhD thesis, University of Leipzig, 2004. 4

[ZL01] ZHANG D., LUO G.: A comparative study on shape retrieval using Fourier Descriptors with different shape signatures. In *Proc. of International Conference on Intelligent Multimedia and Distance Education (ICIMADE01)* (2001), pp. 1–9. 4