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

# SHREC'10 Track: Range Scan Retrieval

H. Dutagaci[†,1], A. Godil[†,1], C. P. Cheung[†,1], T. Furuya[2], U. Hillenbrand[3], R. Ohbuchi[2]

[1]National Institute of Standards and Technology, USA
[2]Graduate School of Medical and Engineering Science, University of Yamanashi, Japan
[3]Institute of Robotics and Mechatronics, German Aerospace Center (DLR), Germany

**Abstract**

*The 3D Shape Retrieval Contest 2010 (SHREC'10) on range scan retrieval aims at comparing algorithms that match a range scan to complete 3D models in a target database. The queries are range scans of real objects, and the objective is to retrieve complete 3D models that are of the same class. This problem is essential to current and future vision systems that perform shape based matching and classification of the objects in the environment. Two groups have participated in the contest. They have provided rank lists for the query set, which is composed of 120 range scans of 40 objects.*

Categories and Subject Descriptors (according to ACM CCS): I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Range data, H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—

## 1. Introduction

There has been a considerable amount of work on 3D shape retrieval based on complete 3D query models [TV07]. There is also interest in object retrieval based on 2D sketches [MCF02, NS10]. However, range images of real objects are becoming common-place with the increase in accuracy, speed and portabilty of 3D scanning devices. These developments bring the necessity for algorithms that enable querying by the single view range scans of generic objects. Range image recognition is already being used in many applications, such as face recognition, automated inspection, and target detection. In these applications, the objects in question are limited to certain categories; examples are 3D faces, products in assembly lines, or specific targets in military applications.

The datasets in this 3D Shape Retrieval Contest track, on the other hand, are designed for algorithms that will enable recognizing novel 3D objects that belong to diverse categories. These algorithms are necessary for a navigating robot to explore and interpret its environment, classify the objects in the scene, and handle them if necessary.

The difference between traditional range image recognition and 3D shape retrieval of range scans was discussed in [RCSM03] and in the SHREC'09 paper [DGA*09]. In traditional range image recognition, the objective is to match the range image of an object to its another scan or its complete 3D model. The modifications are limited to pose variations, acquisition noise, and small deformations; such as facial expressions in the case of 3D face recognition. In the retrieval problem, a version of the input object is not necessarily available to the system. The objective is to retrieve other objects that belong to the same category of the query scan. The categories are determined with respect to their semantic content; hence the objects in the same category may present high geometric variations. Therefore, as mentioned in [DGA*09], the range queries should be processed with regard to the current issues raised by the 3D object retrieval community [TV07].

Another problem with a range query is that, besides containing the geometric information of only one view of an

† Organizer of this SHREC track

**Disclaimer:** Any mention of commercial products or reference to commercial organizations is for information only; it does not imply recommendation or endorsement by NIST nor does it imply that the products mentioned are necessarily the best available for the purpose.

object, that information is still noisy and incomplete due to the limitations of range scaners (Figure 3). The scans contain many holes, and unconnected regions. The scanner was unable to read some steep regions of the view. These imperfections of the view data pose a challenge to the existing shape retrieval algorithms.

The objective of last year's SHREC track on retrieval of partial models was to compare algorithms that perform 3D model retrieval based on range queries [DGA*09]. This year's contest is a continuation of that objective with a larger set containing more accurate range scans.

## 2. The Data Set

### 2.1. Target Set

The target database is the generic shape benchmark constructed at NIST and is desribed in [FGLW08]. It contains 800 complete 3D models, which are categorized into 40 classes. The classes are defined with respect to their semantic categories and are listed in Table 1. In each class there are 20 models. The file format to represent the 3D models is the ASCII Object File Format (*.off).

| | | |
|---|---|---|
| Bird | Fish | NonFlyingInsect |
| FlyingInsect | Biped | Quadruped |
| ApartmentHouse | Skyscraper | SingleHouse |
| Bottle | Cup | Glasses |
| HandGun | SubmachineGun | MusicalInstrument |
| 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 | | |

**Table 1:** *40 classes of the target database.*

### 2.2. Query Set

The query set is composed of 120 range images, which are acquired by capturing 3 range scans of 40 real objects from arbitrary view directions. These objects correspond to the 23 classes in the target database, with one exception. The 3 scans of that object were excluded from the evaluation; hence the evaluation is based on 117 scans. Table 2 gives a summary of the class labels of the scanned objects.

The range images are captured using a Minolta Laser Scanner (Figure 1). We removed the background noise manually and used ASCII Object File Format (*.off) to represent the scans in triangular meshes. Figure 2 and Figure 3 show examples of query objects and their range scans, respectively.

| Class name | # query models |
|---|---|
| Bed | 1 |
| Bicycle | 1 |
| Biped | 4 |
| Bird | 1 |
| Bookshelf | 1 |
| Bottle | 1 |
| Car | 2 |
| Cellphone | 1 |
| DeskPhone | 1 |
| Fish | 1 |
| FlyingInsect | 1 |
| Glasses | 2 |
| HandGun | 1 |
| HomePlant | 1 |
| MilitaryVehicle | 1 |
| Monoplane | 2 |
| Motorcycle | 1 |
| Mug | 3 |
| NonFlyingInsect | 2 |
| NonWheelChair | 1 |
| Quadruped | 5 |
| SingleHouse | 4 |
| SubmachineGun | 1 |
| NONE | 1 |

**Table 2:** *Class information of the query set.*

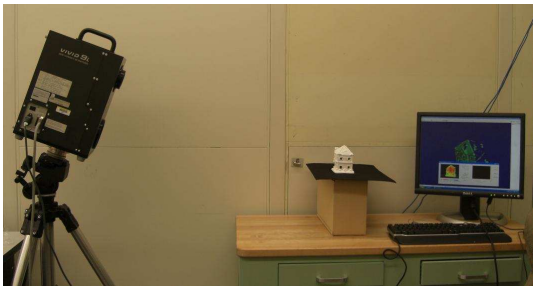

**Figure 1:** *Setup for range scanning.*

## 3. Evaluation Measures

The participants have submitted rank lists for the query inputs. The length of each rank list is equal to the size of the target database. Using the rank lists the following evaluation measures were calculated: 1) Nearest Neighbor (NN), 2) First Tier (FT), 3) Second Tier (ST), 4) E-measure (E), and 5) Discounted Cumulative Gain (DCG) [SMKF04]. In addition to these scalar performance measures, the precision-recall curves were also obtained.

## 4. Submissions

Two groups have participated in the SHREC'10 track Range Scan Retrieval. R. Ohbuchi and T. Furuya from University of Yamanashi have participated with a method based on bag of local 2D visual features. They submitted five sets of rank

**Figure 3:** *Examples from the query set.*

**Figure 2:** *Examples from the set of the objects that were scanned to obtain queries.*

lists corresponding to different choices of pre-processing of range inputs. We will refer to these five submissions as 1) BF-DSIFT-E, 2) Closing-3x3-BF-DSIFT-E, 3) Closing-6x6-BF-DSIFT-E, 4) Dilation-3x3-BF-DSIFT-E, and 5) Dilation-6x6-BF-DSIFT-E. The details are explained in Section 5.

U. Hillenbrand from Institute of Robotics and Mechatronics, German Aerospace Center has participated with a SURFLET-based approach. He submitted five sets of rank lists each using different statistics of SURFLET pairs. We will refer to these five submissions as 1) SURFLET-mean, 2) SURFLET-meanraw, 3) SURFLET-meansqrd, 4) SURFLET-median, and 5) SURFLET-mediansqrd. A description of Hillenbrand's method is given in Section 5. The reader may refer to [WHH03] for a detailed description of shape representation based on Surflet-pair relations.

## 5. Bag of Features - Scale Invariant Feature Transform (BF-SIFT)

R. Ohbuchi and T. Furuya entered the SHREC'10 track Range Scan Retrieval by using their earlier algorithm that employed multi-view rendering and bag-of-2D local visual features. The algorithm is based on their earlier 3D model comparison algorithm [OOFB08], but incorporates later improvements for 3D-to-3D comparison [FO09], and range-scan-to-3D comparison [OF09].

The shape comparison of the algorithm starts with multiple-viewpoint depth-image rendering (Figure 4). For each query, which is a 3D model based on single view range scan, a range image is rendered. Note that, as shown in Figure 5(a), the range image of the single view range-scan 3D mesh contains cracks, jagged edges, and noisy surfaces. For the query model, the method first renders a range image of size $1024 \times 1024$ pixels. It is down sampled with low-pass filtering to $256 \times 256$ pixels in order to reduce aliasing artifacts. For each 3D model in the target database, after normalizing for position and scale, a set of range images from $N_i = 42$ viewpoints are rendered. Range image size of $256 \times 256$ is used for the 3D models. GPU is used for depth image rendering.

The method then densely places sample points on each range image, and extracts *Scale Invariant Feature Transform (SIFT)* [Low04] feature at each sample point. Note that the interest point detector of the Lowe's SIFT algorithm is disabled. Instead, sample positions are assigned a priori. The sample point placement is not completely random. To concentrate samples on or near objects on each image, the samples are placed only on pixels having pixel value above a certain threshold in a large scale (i.e., low-frequency) image of the multi-scale image pyramid of the SIFT algorithm. To speed up the extraction of tens of thousands of SIFT fea-

tures, the GPU-based implementation of SIFT is employed, *siftGPU* [Wu], with necessary modifications.

Extracted SIFT features forms a bag for each view; a bag for a range scan query, and $N_i$ bags for a 3D model. The SIFT features in each bag are then Vector Quantized (VQed) using pre-learned codebook, and accumulated into a histogram for the view. The VQ codebook is learned from 50k SIFT features generated and then sub-sampled from the 3D models in the database. *Extremely Randomized Clustering Tree (ERC-Tree)* [GEW06] is used to learn the codebook and to vector quantize the SIFT features into visual words [FO09].

Then, a histogram is created of the visual words for the partial view query model. For each 3D model rendered into $N_i$ range images, $N_i$ histograms (i.e., feature vectors) are created, and a set of $N_i$ histograms describes a 3D model. To compare a query range scan with a (complete) 3D model, a feature vector of the query model is compared with $N_i$ feature vectors corresponding to $N_i$ views of a 3D model. A minimum of $N_i$ distances between a view of the query and $N_i$ views of the 3D model becomes the distance between the query and the 3D model.

The query contains range scan artifacts, such as cracks, jagged edges, and noisy surface normal vectors, which affect retrieval performance. Figure 5(a) shows the query, and Figure 5(b) shows its range images. Artifacts due to range scan, such as cracks (missing surfaces) and jagged edges at glazing angle are noticeable in the range image as well. To reduce effects of these artifacts, three approaches were employed.

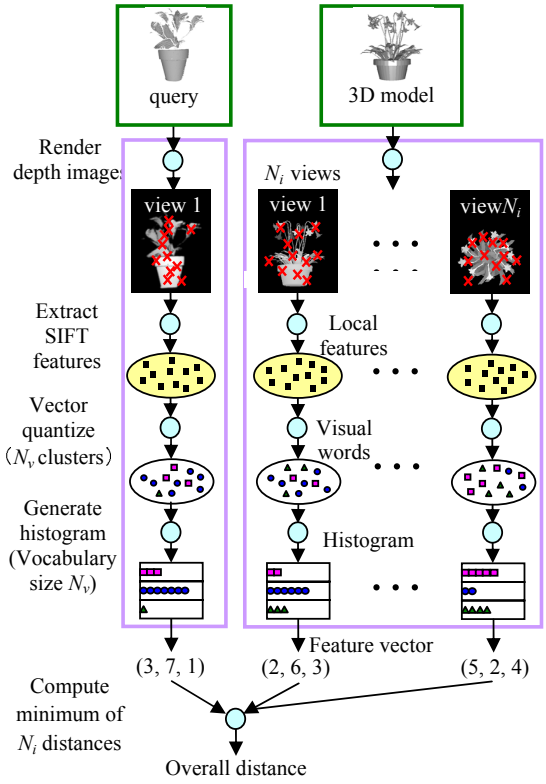


**Figure 4:** *Processing pipeline.*

(a) Range scan model with artifacts.

(b) Rendered range image.

(c) After dilation.

(d) After closing.

**Figure 5:** *Query 3D mesh (a) and its rendered range image (b). The range image after dilation (c) and closing (d) using circular structural element of radius 6 pixel.*

The first is abandoning of interest point detector in the SIFT algorithm; interest point tend to stick to high-contrast, corner-like features due to cracks and jagged edges. Instead, SIFT sample points are placed densely and randomly [FO09, OF09].

The second is *Lower Frequency Emphasis (LFE)* [OF09], which is an importance sampling of larger scale images. The LFE places more samples per pixel in the lower frequency (i.e., larger scale) images than in higher frequency (i.e., smaller scale) images of the multiresolution image pyramid of the SIFT algorithm. This produces a feature histogram that emphasizes larger scale features.

The third is morphological filtering, dilation and closing, to fill/reduce cracks in rendered range images. For dilation and closing, circular structural elements of radius 3 and 6 pixels are used on rendered range images prior to SIFT feature extraction.

## 6. Shape similarity from surflet pair relations

This entry to the range scan retrieval track of SHREC'10 investigates the performance of a variation of a technique that was applied earlier to a different 3D shape classification task [WHH03]. The method is based upon a feature called *surflet pair relations*. A surflet is a surface point with its local (directed) normal vector. The intrinsic geometric relation of two surflets can be described by four parameters, three angels and a distance; see Figure 6. These parameters can be understood as a generalization of surface curvature: while curvature describes the *local* change of surface normals, surflet pair relations describe the normal change across *various distances*.

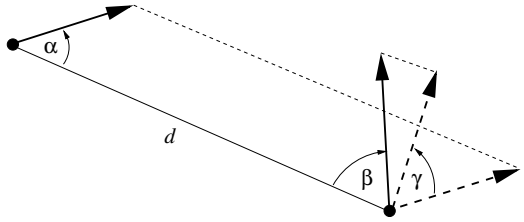


**Figure 6:** *Three angles and a distance for parameterizing a surflet pair relation. Disks are points, arrows are normal vectors, dashed arrows are normal vectors projected into plane orthogonal to point difference.*

For a pose-invariant description of object geometry, the associated density of surflet pair relations, or rather of their four parameters, is of interest. This density can be sampled from geometric object models.

The general strategy for shape similarity computation pursued here is thus:

1. sampling of surflet pair relations from the target models;
2. compact description of the sampled target feature densities;
3. at query time, sampling of surflet pair relations from the query models;
4. computing the distance between query feature samples and target density models.

Prior to any feature sampling, the object surfaces are uniformly point-sampled. The size of the models is normalized such that the median of point pair distances is always unity.

### 6.1. Sampling of surflet pair relations from target models

Since the query data are single-view object data, 1000 viewing directions were uniformly sampled from the sphere. For each viewing direction, 1000 surflet pairs were drawn uniformly among those surflets with orientation consistent with the viewing direction. A total sample of 1,000,000 surflet pairs were obtained from each target object. The surflet pair relations were computed from each pair.

### 6.2. Description of target feature densities

For comparing query to target features, a compact description of the target feature density is needed. To this end, a number of clusters are extracted from the target feature samples. In detail, 100,000 features were drawn randomly from the total sample of 1,000,000 target features for each object, and a $k$-means clustering was applied to the sub-sample with $k = 100$. This procedure was repeated 10 times, yielding a total of 1000 cluster centers for each target object, which may be somewhat sloppily regarded as describing modes of the underlying feature density.

### 6.3. Sampling of surflet pair relations from query models

At query time, 100,000 surflet pairs were randomly drawn from the low-resolution version of the query models. The surflet pair relations were computed from each pair.

### 6.4. Computing the distance between query feature samples and target density models

The all-over distance between shapes was computed as a simple statistics of the Euclidean distance between each individual query feature (from the sample of 100,000 features for each query model) and its nearest target cluster center in the 4D feature space. The following statistics were tried.

- Mean Euclidean distance (method *mean*).
- Median Euclidean distance (method *median*).
- Mean squared Euclidean distance (method *meansqrd*); if the cluster centers are understood as means of Gaussian density modes with equal and isotropic covariance, the resulting shape distance is essentially a negative likelihood.
- Median squared Euclidean distance (method *medsqrd*); this distance measure will be nearly the same as the square of the median Euclidean distances above, the rankings produced by the two median measures must hence be very similar, if not identical.

Moreover, for comparison a method was included (method *meanraw*) where the mean Euclidean distance was taken between the query features and their nearest target feature from the raw sample.

## 7. Results

The two participants of the SHREC'10 track Range Scan Retrieval submitted five sets of rank lists each. The results for the ten submissions are summarized in Figure 7 and in the precision-recall curves in Figure 8. The reader can find the retrieval results for the individual range queries at our Flex based Graphical User Interface [INT]. The interface shows the retrieved models for all the query models and for all the methods described in this paper. Figure 9 shows the models retrieved by Closing-6x6-BF-DSIFT-E in response to a

range scan of a toy bike, and Figure 10 shows the interface for a range scan of a toy fish and the SURFLET-mean algorithm.

| PARTICIPANT | METHOD | NN | FT | ST | E | DCG |
|---|---|---|---|---|---|---|
| | BF-DSIFT-E | 0.573 | 0.380 | 0.524 | 0.367 | 0.683 |
| Ohbuchi | Closing_3x3_BF-DSIFT-E | 0.598 | 0.393 | 0.535 | 0.382 | 0.696 |
| & Furuya | Closing_6x6_BF-DSIFT-E | 0.650 | 0.424 | 0.569 | 0.398 | 0.713 |
| | Dilation_3x3_BF-DSIFT-E | 0.675 | 0.405 | 0.557 | 0.392 | 0.713 |
| | Dilation_6x6_BF-DSIFT-E | 0.547 | 0.395 | 0.550 | 0.386 | 0.696 |
| | SURFLET - mean | 0.325 | 0.244 | 0.363 | 0.252 | 0.556 |
| | SURFLET - meanraw | 0.171 | 0.153 | 0.242 | 0.163 | 0.462 |
| Hillenbrand | SURFLET - meansqrd | 0.231 | 0.197 | 0.322 | 0.213 | 0.513 |
| | SURFLET - median | 0.282 | 0.226 | 0.325 | 0.224 | 0.528 |
| | SURFLET - mediansqrd | 0.282 | 0.226 | 0.325 | 0.224 | 0.528 |

**Figure 7:** *Retrieval performances.*

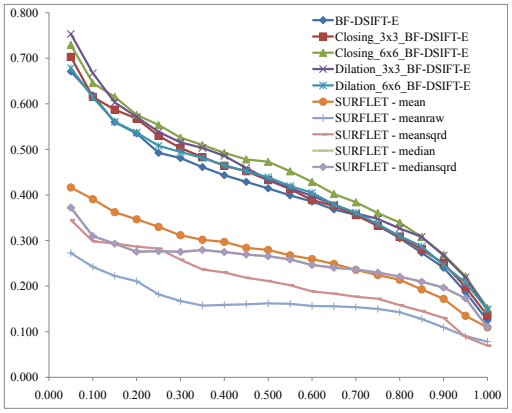


**Figure 8:** *Precision-recall curves.*

For the bag of features algorithm, the participants used the dense and random sampling combined with the LFE. Four out of five used morphological filtering; either dilation or closing, each with circular structural element of radius either 3 or 6 pixels. The evaluation results show that the morphological filtering worked to some extent. In terms of the First Tier (FT) or Discounted Cumulative Gains (DCG) measures, the variation using closing with structural element radius 6 pixel ("Closing-6x6-BF-DSIFT-E (LFE)") performed the best. In terms of the Nearest Neighbor (NN) measure, the method employing dilation with structural element radius 3 ("Dilation-3x3-BF-DSIFT-E (LFE)") performed the best.

For the surflet-based algorithm, the average performance across the 120 queries turns out to be rather poor, as seen in Figure 7 and 8. Looking at the distribution of quality measures for individual queries gives somewhat more insight. As an example, Figure 11 shows the histograms of first tier ratios, i.e., the frequencies of their discrete values, for the five variants of surflet-based method and the five variants of the bag of features algorithm. The first tier distribution for all methods is roughly uniform for most of the range of values. For surflet-based method, however, there is a set of about one third of queries with zero first tier ratios, indicating a complete break-down, while on a set of two thirds of queries its performance is comparable to the one based on bag of features.

The main reason for the frequent break-down of the surflet-based method likely lies in the fact that here, all views of the target objects have been superposed in a single model. Not keeping the features of different views separate may easily give rise to confusion for certain shapes. Since employing multiple view models generally incurs higher computational costs, the results reflect a kind of performance/cost trade-off. Problems might also derive from the rather coarse description of the feature densities used here for the target models. However, it is interesting to note that usage of the raw sample of target features does not improve the results.

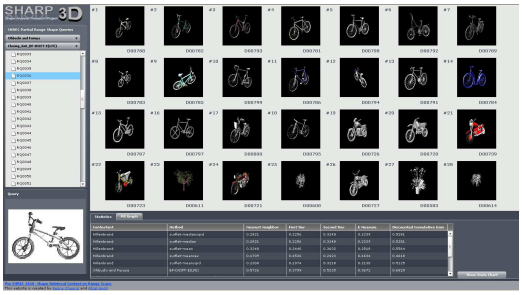


**Figure 9:** *A sample shot from the web-based interface [INT]. The query is the range scan of a toy bike. The method is Closing-6x6-BF-DSIFT-E.*

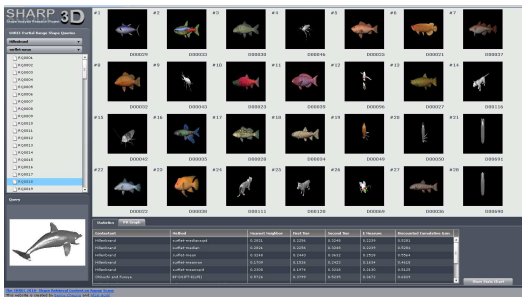


**Figure 10:** *A sample shot from the web-based interface [INT]. The query is the range scan of a toy fish. The method is SURFLET-mean.*

## 8. Conclusions

In this paper, we have described and compared two algorithms and their variants of two research groups that participated in the SHREC'10 Range Scan Retrieval track. The algorithms accept a range scan as the input and retrieve similar models from a database of complete 3D models. The method based on bag of features yielded better average performance as compared to the surflet-based algorithm.

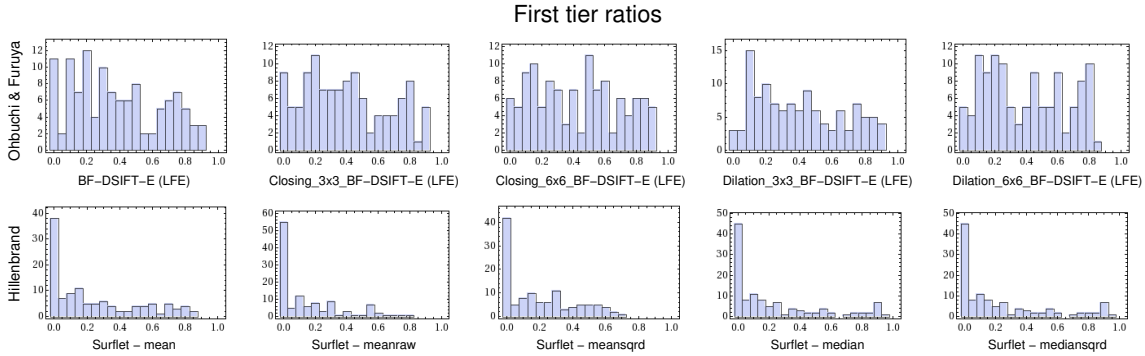


**Figure 11:** *Histograms of first tier ratios for all queries.*

## References


[DGA*09] DUTAGACI H., GODIL A., AXENOPOULOS A., DARAS P., FURUYA T., OHBUCHI R.: Shrec'09 track: Querying with partial models. In *3DOR* (2009).

[FGLW08] FANG R., GODIL A., LI X., WAGAN A.: A new shape benchmark for 3D object retrieval. In *ISVC (1)* (2008), pp. 381–392.

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

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

[INT] http://control.nist.gov/sharp/SHREC10/Partial-Range/SHREC .

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision 60*, 2 (2004), 91–110.

[MCF02] MIN P., CHEN J., FUNKHOUSER T.: A 2d sketch interface for a 3d model search engine. In *SIGGRAPH '02: ACM SIGGRAPH 2002 conference abstracts and applications* (New York, NY, USA, 2002), ACM, pp. 138–138.

[NS10] NAPOLÉON T., SAHBI H.: Sketch-driven mental 3d object retrieval. In *Proc. SPIE, Three-Dimensional Image Processing (3DIP) and Applications, Vol. 7526* (2010).

[OF09] OHBUCHI R., FURUYA T.: Scale-weighted dense bag of visual features for 3d model retrieval from a partial view 3d model. In *Proc. ICCV 2009 Workshop on Search in 3D and Video (S3DV)* (2009).

[OOFB08] OHBUCHI R., OSADA K., FURUYA T., BANNO T.: Salient local visual featuers for shape-based 3D model retrieval. In *Shape Modeling International* (2008).

[RCSM03] RUIZ-CORREA S., SHAPIRO L. G., MEILA M.: A new paradigm for recognizing 3-D object shapes from range data. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision* (2003), p. 1126.

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

[TV07] TANGELDER J., VELTKAMP R.: A survey of content based 3d shape retrieval methods. *Multimedia Tools and Applications* (2007).

[WHH03] WAHL E., HILLENBRAND U., HIRZINGER G.: Surflet-pair-relation histograms: A statistical 3d-shape representation for rapid classification. *Proc. International Conference on 3-D Digital Imaging and Modeling 0* (2003), 474–481.

[Wu] WU C.: A gpu implementation of david lowe's scale invariant feature transform (sift). http://cs.unc.edu/~ccwu/siftgpu/ .