How well do saliency-based features perform for shape retrieval?

[1]**Flora P. Tasse,** [2]**Jiří Kosinka,** [3]**Neil A. Dodgson**

[1]The Computer Laboratory, University of Cambridge, United Kingdom
[2]Johann Bernoulli Institute, University of Groningen, The Netherlands
[3]School of Engineering and Computer Science, Victoria University of Wellington, New Zealand

Abstract: Sparse features have been successfully used in shape retrieval, by encoding feature descriptors into global shape signatures. We investigate how sparse features based on saliency models affect retrieval and provide recommendations on good saliency models for shape retrieval. Our results show that randomly selecting points on the surface produces better retrieval performance than using any of the evaluated salient keypoint detection, including ground-truth. We discuss the reasons for and implications of this unexpected result.

Keywords: Shape retrieval, salient features, keypoints, bag of features.

1 Introduction

Large-scale shape retrieval typically consists of three steps: detection of local features, encoding of these features into a global descriptor, and comparison of shapes with a distance metric. This paper evaluates the influence of the first step on shape retrieval performance.

Shape saliency is a measure of perceived importance of points on a 3D surface. There is recent interest in shape analysis for computational models of saliency [2, 3, 6, 7]. However, there is no study on how features detected by these saliency models perform on shape retrieval. This paper evaluates selected saliency models. For each method, we compute surface feature points by extracting local maxima from a saliency map. The global descriptor of a 3D shape is then a distribution of quantized salient features. We investigate how well features based on selected saliency models perform on shape retrieval benchmarks. Our results show that randomly selecting points on the surface produces better retrieval performance than using any of the evaluated saliency-based features. This surprising result also holds for ground-truth salient points obtained in a previous user study [1] on the SHREC'07 Watertight Models Track (SHREC07) [8].

We provide a fair comparison of salient features by fixing other variables such as feature descriptors, descriptors distance metrics, and encoding method. We evaluate retrieval performance on three benchmarks: SHREC07 [8], SHREC'15 Non-Rigid Shape Retrieval track [9], and SHREC'15 Range scan shape retrieval [10]. Figure 1 illustrates how saliency, keypoint detection and retrieval performance varies by saliency method on a "human" shape.

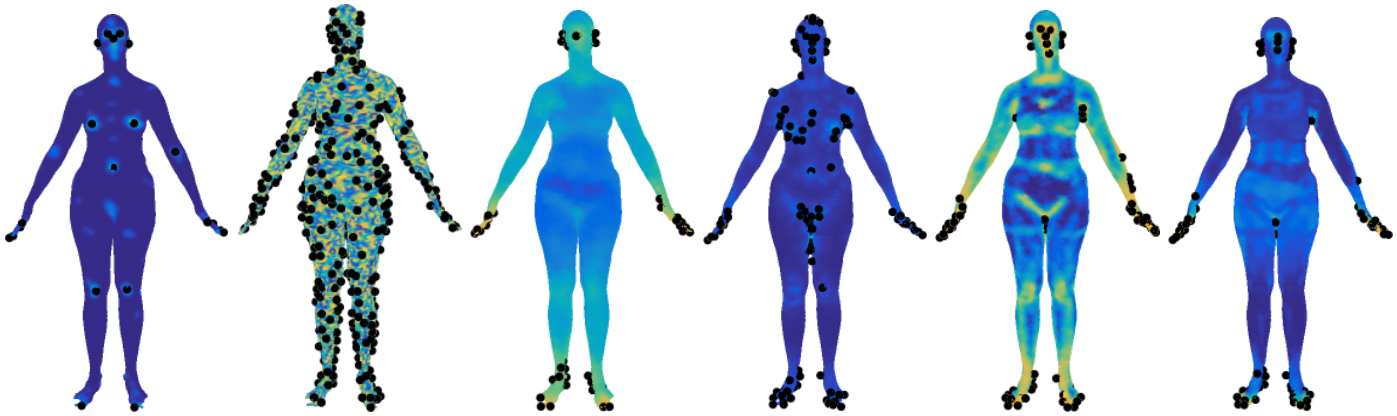
Our main contribution is the evaluation of six selected saliency models for shape retrieval.

2 Related Work

This section reviews state-of-the art in saliency detection, salient keypoint extraction, and local feature encoding for shape retrieval.

2.1 Saliency models

A wide range of 3D saliency models have been proposed in recent years, often inspired by analogous techniques in 2D saliency detection [2, 5, 6]. Several of these models compute a multi-scale representation of a mesh and observe how a local vertex property such as curvature changes at different scales [6]. Song et al. [2] propose using spectral properties of the log-Laplacian spectrum of a mesh at multiple scales. Both approaches require topological information and thus cannot support other shape representations such as point clouds. To address saliency detection on large point sets, Shtrom et al. [3] combine point distinctiveness at two scales with point association, a function that assigns higher saliency to regions near foci of attention. Point distinctiveness is computed by comparing points using the χ^2 distance between their Fast Point Feature Histograms (FPFH) [11]. Tasse et al. [5] achieve better fine-scale saliency detection and computation performance by segmenting point clouds into patches, and computing a patch saliency based on its descriptor distinctiveness and spatial distribution. FPFH have been successfully used in several saliency models [3, 5]. Authors et al. [4] show that applying PCA to these descriptors produces saliency maps that compare well with the state of the art.



$AP = 0.79$ Ground-truth [1] | $AP = 0.90$ RK (random) | $AP = 0.77$ MK [2] | $AP = 0.78$ LK [3] | $AP = 0.73$ PK [4] | $AP = 0.65$ CK [5]

Figure 1: Saliency maps and corresponding keypoints per saliency model on a shape in Dataset A (SHREC07). The retrieval performance of this shape is measured with Average Precision (AP). Higher AP means better retrieval performance. AP takes values between 0 and 1.

2.2 Salient keypoint detection

3D feature points can be detected by extracting local maxima of a map over a 3D shape [5, 12]. Dutagaci et al. [12] use this approach to compute interest points based on mesh saliency [6]. They also propose a benchmark that compares other feature detection algorithms such as local maxima of the Heat Kernel Signature (HKS) [13] to ground-truth interest points. Their benchmark shows that HKS-based keypoint detection has a higher false negative error rate and smaller false positive rate compared to all other tested methods. Tasse et al. [5] show, on the same benchmark, that using local maxima of their cluster-based point set saliency achieves better balance between false positive and false negative error rates. Chen et al. [1] extract consistent feature points on meshes in SHREC07, by taking the local maxima of ground-truth saliency aggregated from multiple participants. Based on this data, they propose a probabilistic model for extracting feature points from any mesh. Salti et al. [14] move away from saliency-based approach and cast the keypoint detection as a binary classification problem, to learn keypoints that are distinctive, according to a specific descriptor.

2.3 Feature encoding

Encoding of local features is a popular technique borrowed from image and video retrieval [15]. Local features, sparse or dense, are extracted from the whole dataset and each feature is represented by a multi-dimensional descriptor. Encoding local features of an input shape typically consists of evaluating the distribution of quantised features to form a global descriptor. Tabia et al. [16] use Histogram Encoding [15], based on counting quantized features, to encode their covariance descriptors. Bronstein et al. [17] use a more descriptive method, Soft Quantisation [18], that consists of summing softly-quantised features to encode sparse HKS-based features.

Fisher Vectors retain more information about shape features by recording the statistics of differences between local features and clusters of the descriptor space [19]. Savelonas et al. [20] present shape retrieval based on Fisher encoding of novel local descriptors derived from FPFH. We use Fisher Vectors to encode local features.

After encoding, global descriptors are typically compared using their normalized scalar product (cosine of angle) [16].

3 Experimental setup

3.1 Datasets

We use three datasets to evaluate retrieval performance, each targeting a different type of shape retrieval.

Dataset A: SHREC'07 Watertight Models [8] The dataset consists of 20 classes, each containing 20 different watertight 3D meshes. Figure 2 illustrates a few models in the dataset and their classes. It is a generic dataset, due to the diverse number of classes and the variety of objects within a class. We use this dataset, instead of others such as the Princeton Shape Benchmark, because ground-truth salient points are available for it. Chen et al. [1] present a user study that asks users to select points, on shapes from this dataset, that are likely to be selected by other users. From the collected user input, they identify ground-truth keypoints. We use these to evaluate shape retrieval based on human-perceived saliency.

Dataset B: SHREC'15 Non-Rigid Shape Retrieval [9] The benchmark objective is checking shape retrieval invariance to non-rigid shape transformations. The dataset contains 1200 watertight triangle meshes, obtained

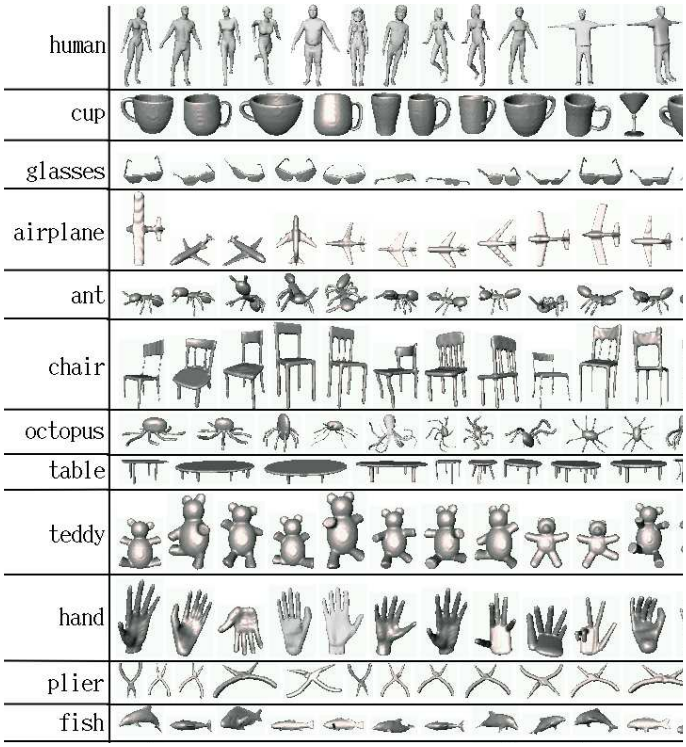


Figure 2: Dataset A: Examples of watertight models, divided into classes [8].

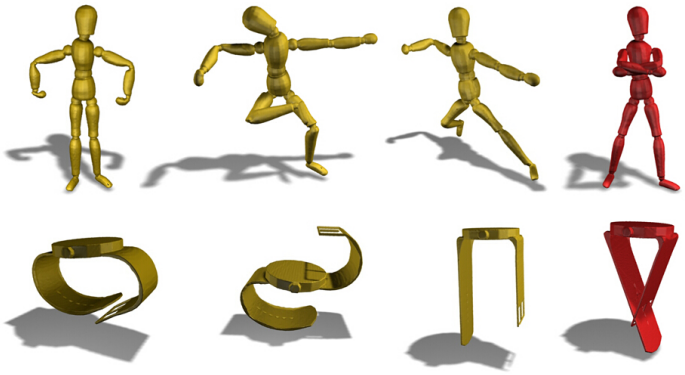


Figure 3: Dataset B: Examples of non-rigid models [9].

Figure 4: Dataset C: Example of range scans in the query set [10].

by deforming 60 models from 50 classes. Examples of non-rigid models are shown in Figure 3.

Dataset C: SHREC'15 Range Scans Shape Retrieval [10] The aim of this benchmark is testing shape retrieval robustness to partial queries. The dataset is divided into two: target models and query range scans. The target set contains 1200 complete 3D models from 60 classes. The query set consists of 180 range images acquired from 3 to 4 range scans of 60 models. Figure 4 shows examples of these 3D scans.

We now discuss other factors studied in our analysis such as local descriptors and saliency models.

3.2 Local descriptors

Local descriptors help describe the local neighbourhood or support area $\mathcal{N}$ of a point. $\mathcal{N}(p)$ is the set of neighbours q of p with $\|p-q\| < r$, where r is the radius of the neighbourhood often referred to as *support radius*. To support any surface representation, we focus on point-based descriptors. Few previous works report their choice of r for such descriptors. Rusu et al. [21] compute FPFH using $r = 0.5cm$ on point sets with average radius $R = 3cm$; thus they set $r = R/6$. Shtrom et al. [3] use $r = R/10$ when computing FPFH for the purpose of detecting high-level saliency. We also set $r = R/10$, where R is the underlying shape radius.

We test four local descriptors: Point Feature Histogram (PFH) [11], Fast Point Feature Histogram (FPFH) [11], Spin Images [22] and SHOT [23]. We use a PCL implementation [24] of all four descriptors.

PFH generalizes the mean curvature around a point using a multi-dimensional histogram of angular variations between all pairs of oriented points in a local neighbourhood [11]. FPFH is a faster version of PFH that averages histograms of angular variations between a point and its neighbourhood [11]. Spin Images bin 2D projections of nearby points in a cylindrical coordinate system defined by the oriented point of interest [22]. SHOT uses a set of local histograms over 3D volumes defined by a voxelization of the support area [23].

Shape retrieval has better performance when PFH is used as the local descriptor, compared to the alternatives as discussed in Section 5.1. Thus, we use PFH in experiments where we need a single robust local descriptor.

3.3 Saliency models

We test shape retrieval based on local features extracted with keypoint detectors based on six selected saliency models. Four of these saliency models were discussed in Section 2.1, and selected based on availability of code and data. They are:

1. Spectral mesh saliency (MK) [2]

2. Large point set saliency (LK) [3]

3. Cluster-based point set saliency (CK) [5]

4. PCA-based saliency (PK) [4]

We include in our evaluation 2 baseline models:

5. Random keypoints (RK): We are also interested in how points sampled uniformly at random on a shape affect retrieval performance. Some retrieval systems use dense random points rather than sparse features [25, 16] for shape representation, at the expense of a higher computational cost. We want to know how this approach compares with using salient features. We define a random scalar field on the shape, referred to as Random Saliency. We detect keypoints based on local maxima of the random saliency map, as previously discussed. The number of points detected depends on the shape sampling. An alternative to this is extracting a fixed number of random surface points per shape. This alternative random keypoint detector is explored in Section 5.3.

6. Ground-truth keypoints (GK): To see how these keypoint detectors compare to ground-truth when used for shape retrieval, we also evaluate retrieval performance based on ground-truth points collected on Dataset A [1].

Local keypoints are each represented by a descriptor. We then use Fisher Vectors (FV), previously mentioned in Section 2.3, to encode local descriptors on a given shape into a global descriptor. FV encode means and variances between a set of local descriptors and clusters of the space containing all local shape descriptors [19].

3.4 Evaluation metrics

Given a set of 3D models assigned to classes, a shape retrieved based on a query is relevant if both target and query belong to the same class. This interpretation of relevance is standard in shape retrieval benchmarks [8, 9, 10]. These benchmarks also use the following standard metrics to evaluate retrieval performance: Precision-Recall (PR) curve, Average precision (AP), First Tier (FT), Second Tier (ST), and Discounted Cumulative Gain (DCG).

To compute the performance of retrieval on a dataset, we proceed as follows: for each shape in the query set (or the dataset if there is no query set), we generate a list of all models in the target set, ranked from the most similar to the least similar. The ranked lists are used to compute the above metrics for each query. Each metric is then averaged over all queries to produce overall scores. Finally, we use the Wilcoxon rank-sum test [26], a non-parametric alternative to the two-samples t-test, at a 0.05 significance level to report statistically significant differences between AP performances of competing methods.

A performance metric that is often used in benchmarks is Nearest Neighbour. We did not observe any statistically significant difference in Nearest Neighbour performance across the evaluated methods, contrary to the other five metrics described above. Thus, we do not report its performance results.

Table 1: Dataset A: Performance per keypoint detector. Parameters: descriptor=PFH, $r = 0.1R$, $K = 100$.

	FT	ST	DCG	AP
RK	$\mathbf{0.63 \pm 0.03}$	$\mathbf{0.73 \pm 0.03}$	$\mathbf{0.84 \pm 0.02}$	$\mathbf{0.70 \pm 0.03}$
MeshDOG	0.58 ± 0.03	0.70 ± 0.03	0.82 ± 0.02	0.65 ± 0.03
ISS	0.58 ± 0.03	0.69 ± 0.03	0.82 ± 0.02	0.65 ± 0.03
GK	0.55 ± 0.03	0.68 ± 0.03	0.80 ± 0.02	0.62 ± 0.03
LK	0.53 ± 0.03	0.66 ± 0.03	0.80 ± 0.02	0.61 ± 0.03
PK	0.50 ± 0.03	0.62 ± 0.03	0.77 ± 0.02	0.57 ± 0.03
MK	0.48 ± 0.03	0.62 ± 0.03	0.76 ± 0.02	0.55 ± 0.03
CK	0.47 ± 0.03	0.59 ± 0.03	0.75 ± 0.02	0.54 ± 0.03

Table 2: Dataset B: Performance per keypoint detector. Parameters: descriptor=PFH, $r = 0.1R$, $K = 100$.

	FT	ST	DCG	AP
RK	$\mathbf{0.86 \pm 0.01}$	$\mathbf{0.91 \pm 0.01}$	$\mathbf{0.96 \pm 0.00}$	$\mathbf{0.90 \pm 0.01}$
MeshDOG	0.84 ± 0.01	0.91 ± 0.01	0.96 ± 0.00	0.89 ± 0.01
ISS	0.84 ± 0.01	0.90 ± 0.01	0.96 ± 0.00	0.88 ± 0.01
MK	0.83 ± 0.01	0.90 ± 0.01	0.95 ± 0.00	0.88 ± 0.01
LK	0.80 ± 0.01	0.86 ± 0.01	0.94 ± 0.01	0.85 ± 0.01
PK	0.78 ± 0.01	0.85 ± 0.01	0.94 ± 0.01	0.83 ± 0.01
CK	0.69 ± 0.01	0.78 ± 0.01	0.91 ± 0.01	0.75 ± 0.01

Table 3: Dataset C: Performance per keypoint detector. Parameters: descriptor=PFH, $r = 0.1R$, $K = 100$.

	FT	ST	DCG	AP
RK	$\mathbf{0.05 \pm 0.01}$	$\mathbf{0.10 \pm 0.03}$	$\mathbf{0.35 \pm 0.02}$	$\mathbf{0.06 \pm 0.01}$
MeshDOG	0.04 ± 0.02	0.07 ± 0.02	0.34 ± 0.02	0.05 ± 0.02
CK	0.04 ± 0.01	0.08 ± 0.02	0.34 ± 0.01	0.05 ± 0.01
ISS	0.04 ± 0.01	0.08 ± 0.02	0.34 ± 0.01	0.05 ± 0.01
LK	0.03 ± 0.01	0.06 ± 0.01	0.34 ± 0.01	0.05 ± 0.01
MK	0.03 ± 0.01	0.07 ± 0.02	0.33 ± 0.01	0.04 ± 0.01
PK	0.02 ± 0.01	0.06 ± 0.02	0.32 ± 0.01	0.04 ± 0.01

4 Evaluation of salient keypoints

We present PR curves of retrieval systems based on 6 salient keypoint detectors in Figure 5. Note that we use PFH as local descriptors and cluster size $K = 100$ to partition the set of all descriptors. We use FV as the encoding method, and compare the generated global shape descriptors with the cosine angle. We do not include GK in our evaluation on datasets B and C, because these datasets do not have ground-truth saliency data.

4.1 Comparison of selected saliency-based keypoint detectors

Dataset A Table 1 presents more details on retrieval performance by keypoint detector on Dataset A. RK is significantly better than all other methods on each performance metric. There is no statistically significant difference between GK and LK, implying that on this dataset, LK is as good as ground-truth when used for shape retrieval. There is also no significant difference between the worst-performing models PK, MK and CK. The top performance

of RK is not surprising since it generates on average 250 points per shape, while GK only produces an average of 33 points per shape. A large number of random surface points provides more coverage of the surface and thus captures more information about shape. Surprisingly we see in Figure 7 (bottom) that even when we choose only $n = 30$ random points per shape, we *still* get better average performance than using GK. The difference however is not statistically significant. Choosing $n = 50$ random points per shape produces a significantly better retrieval performance, meaning that we are getting better results with 50 randomly selected points than with 33 carefully selected salient points. An in-depth analysis of the effect of saliency and random sampling size is presented in Section 5.3. RK has comparable performance to the state-of-the-art in shape retrieval [16], which is based on dense random features represented by covariance descriptors, with performance metrics FT = 0.623, ST = 0.737 and DCG = 0.864.

Dataset B Table 2 shows retrieval performance on Dataset B. All pair-wise differences between keypoint detectors are significant, with the exception of LK and PK. RK remains the top-performing saliency model for retrieval. MK now has second place. This is explained by the fact that MK saliency model is based on spectral properties, which makes it more robust to non-rigid transformations, compared to other methods. CK is the worst performing method, since its saliency model relies on spatial distribution of FPFH and thus is not robust to large deformations. Lian et al. [9] evaluate recent retrieval systems on dataset B. RK is outperformed by methods, featured in their benchmark, that are based on isometry-invariant descriptors, or volumetric feature representations.

Dataset C Table 3 shows performance per keypoint detectors. Performance metrics on this dataset are low, and there are no statistically significant pair-wise differences between keypoint detectors. The low performance suggests that encoding 3D local features is not an adequate method for describing range scans. This is supported by an evaluation of recent retrieval methods on dataset C, which shows similar poor performance for methods using 3D features, and significantly better results for methods based on 2D rendered views [10].

Summary RK provides the best retrieval performance, out of the evaluated keypoint detectors, on datasets of whole models. It outperforms ground-truth salient points (GK). Furthermore, using a small number of random keypoints (as small as $n = 30$) still outperforms salient features on a generic dataset. This shows that sparse salient features are not appropriate for shape representation. Rather than using RK, which generates a large number of random keypoints and is thus expensive (see Section 6), we recommend using a small random sampling size, such as $n = 50$, per shape.

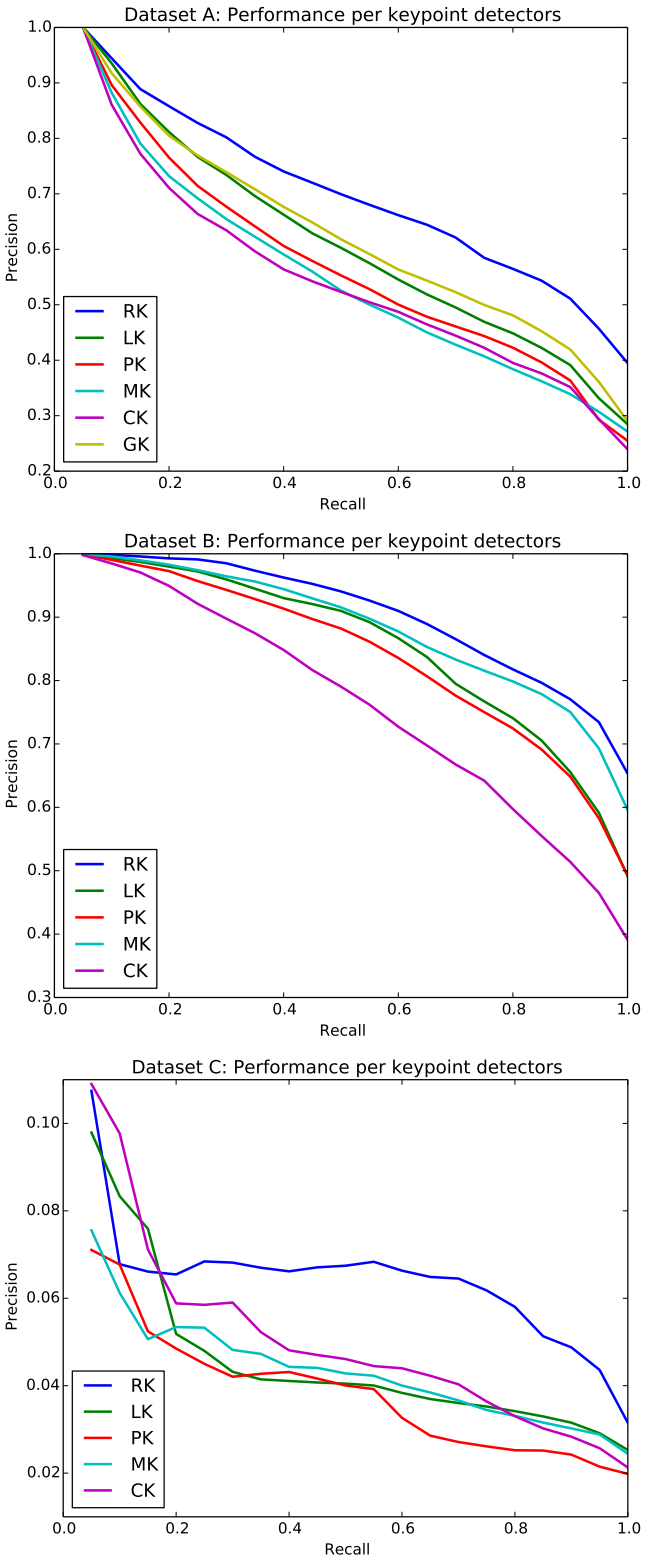


Figure 5: Precision recall curves for retrieval performance based on selected keypoint detectors.

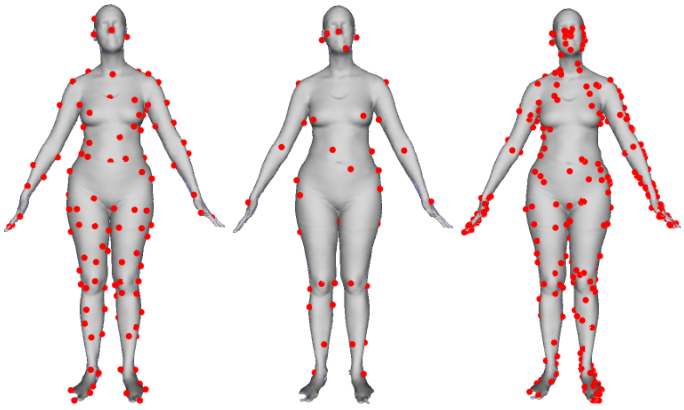
4.2 Comparison against other state-of-the-art keypoint detectors

Thus far, we have focused on keypoint detectors that are based on local maxima of saliency algorithms. There is a large body of work on interest point detection beyond the ones we have evaluated above. Tombari et al. [27] present a qualitative evaluation of major keypoint detectors in the literature, in terms of repeatability, distinctiveness and computational efficiency. Repeatability is the ability to find the same keypoints on different instances of a given shape. Their evaluation reports that Intrinsic Shape Signatures detector (ISS) [28] is highly efficient and provides a good trade-off between absolute and relative repeatability, making it a good choice for shape retrieval. Note that ISS is fixed-scale, and thus detects keypoints at a specific scale, provided via a support radius parameter. The authors then show that for object recognition, the adaptive-scale detector MeshDOG [29] produces the best results. In this section, we compare ISS and MeshDOG against GK and RK.

ISS [28] computes saliency of a point based on the eigenvalues of the scatter matrix of points within a support radius. Saliency is the magnitude of the smallest eigenvalue. Keypoint detection is based on this saliency value (to include points with large variations against each direction), and the ratio between the second and third eigenvalues (to avoid points with a similar spread along principal directions). On the other hand, MeshDOG is based on a scale-space representation of a mesh, built by applying the Difference-of-Gaussians (DoG) operator on a scalar function defined over the mesh. Points are ranked by their saliency values, thresholded to keep the number of detected points below a fixed percentage of the number of vertices in the shape, and retained only if they exhibit corner characteristics.

We use the same parameters for both detectors as Tombari et al. [27], including the selection of mean curvature as the scalar function in MeshDOG. They evaluate ISS at scales $\{6m, 10m, 12m, 18m\}$, where m is the mesh resolution computed by taking the average edge length. However, we noticed that on our datasets, ISS at scale m generates more keypoints than detection at higher scales, including $6m$, and achieves better performance. This is illustrated in Figure 6. Thus, in the results reported below, ISS is computed at scale m.

Table 1 shows retrieval performance on Dataset A, when using ISS or MeshDOG keypoint detectors against RK and GK. For more details on the number of keypoints detected per method, we refer the reader to Appendix C. The results show that the adaptive-scale MeshDOG not only produces more keypoints than ISS and GK, but it is also produces significantly better retrieval performance. However, RK outperforms MeshDOG. This supports the argument that more keypoints provide more coverage of the surface, which leads to better retrieval.



$AP = 0.70$ ISS (scale m) | $AP = 0.65$ ISS (scale $6m$) | $AP = 0.77$ MeshDOG

Figure 6: Keypoints on a shape in Dataset A, based on ISS and MeshDOG detectors.

5 Evaluation of parameter choices

We investigate the impact of various design choices on shape retrieval performance. We do this analysis on Datasets A, B and C unless specified otherwise. We pay closer attention to our results on Dataset A since it is the most diverse collection out of the three datasets, in terms of number of classes and variability within classes. The default design choices are RK as the keypoint detector, PFH as local descriptors, and cluster size $K = 100$. Changes in these default design choices are specified explicitly in each section.

5.1 Choice of local descriptors

In addition to PFH we evaluate 3 other feature descriptors: FPFH [11], SHOT [23] and Spin Images [22]. Results are summarized in Figure 7 (top) and Table 4. PFH is significantly better than other descriptors. It is more robust to sampling compared to SHOT and Spin Images, and captures more information than FPFH as it is 4 times larger. These results are based on the RK feature detector. Further experiments in Appendix A use alternative detectors and support the claim that PFH is a better descriptor.

Table 4 also includes performance when using a global descriptor, as opposed to a Bag of features approach, with no dependency on keypoint detection. We obtain such global descriptor by computing a single PFH descriptor with an infinite support radius, which means that all points are included in the histogram of angular variations. We denote the global descriptor by GPFH. Results show the global descriptor is outperformed by encoding local descriptors. This indicates that local information is important for retrieval.

5.2 Number of clusters

By default we set the cluster size $K = 100$, after exploring a range of values for K. Figure 7 (middle) shows retrieval performance for different K. There is no statistically significant

Table 4: Dataset A: Effect of local feature descriptors and PFH-based global descriptor GPFH on retrieval. Parameters: detector=RK, $r = 0.1R$, $K = 100$.

	FT	ST	DCG	AP
PFH	$\mathbf{0.63 \pm 0.03}$	$\mathbf{0.73 \pm 0.03}$	$\mathbf{0.84 \pm 0.02}$	$\mathbf{0.70 \pm 0.03}$
GPFH	0.58 ± 0.03	0.71 ± 0.03	0.84 ± 0.02	0.66 ± 0.03
FPFH	0.57 ± 0.03	0.68 ± 0.03	0.81 ± 0.02	0.64 ± 0.03
SHOT	0.54 ± 0.03	0.68 ± 0.03	0.80 ± 0.02	0.62 ± 0.03
SPIN	0.53 ± 0.03	0.65 ± 0.03	0.78 ± 0.02	0.60 ± 0.03

difference in performance for various values of K. Cluster size has little effect on FV, since the encoding retains some information about the original shape feature descriptors.

5.3 Number of random points

Section 4 shows that RK keypoint detector produces better retrieval. RK selects random points on the surface by extracting keypoints from a random saliency map. The number of random points depends on the shape sampling. We investigate the effect of choosing a specific number of random points per shape. We denote this number by n. Local features are n surface points randomly selected with uniform distribution, and shifted using Lloyds relaxation [30] to provide a better coverage of the whole surface. Figure 7 (bottom) shows retrieval performance for varying n. There is no statistically significant difference between using |RK| or choosing n random points per shape, for $n \geq 50$. This implies that selecting a small fixed number of points at random is good enough. To assess whether ground-truth salient points have any useful effect we tried restricting the random point selection to areas that are not close to a ground-truth salient point. This does reduce average retrieval performance, but not significantly. This suggests that the distribution of non-salient features on the surface is important for shape retrieval. We discuss this in more depth in Section 7.

5.4 Support radius r

The support radius is used to indicate the scale of local neighbourhood, and it is a key parameter in computing local descriptor. The above experiments use a fixed $r = 0.1R$. In this section, we investigate the influence of r on retrieval, by looking at values ranging from $0.05R$ to $0.5R$. Tables 5 and 6 show retrieval results for different values of r combined with RK30 (random keypoint detector with $n = 30$). We do not report results on Dataset C since they were not statistically significant. Results show that for Dataset A there is no performance improvement for $r \geq 0.2$. As for Dataset B, performance increases with the support radius, until $r = 0.2$ and decreases for larger r. This shows that support radius does indeed affect retrieval, and when chosen too small or too large, it can decrease retrieval performance.

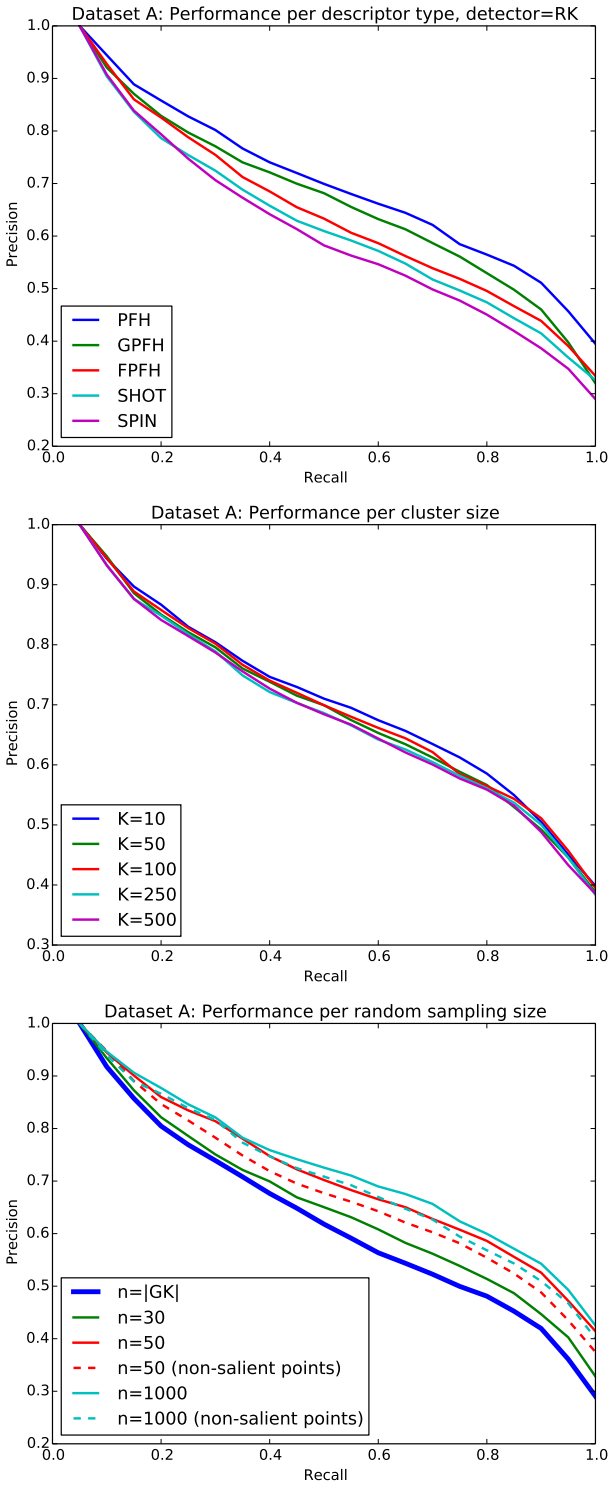


Figure 7: Performance on Dataset A for various design choices. The PR curves show that PFH is a better feature descriptor (top), cluster size does not have an impact on retrieval (middle), and retrieval performance increases with larger number of random features (bottom). $n = |\text{GK}|$ is the number of ground-truth salient points. We also investigate the effect of only choosing non-salient random points.

Table 5: Dataset A: Performance based on support radius $r = tR$. Parameters: detector=RK30, descriptor=PFH, $K = 100$.

	FT	ST	DCG	AP
t=0.05	0.41 ± 0.03	0.57 ± 0.03	0.71 ± 0.02	0.48 ± 0.03
t=0.1	0.57 ± 0.03	0.72 ± 0.03	0.83 ± 0.02	0.65 ± 0.03
t=0.2	0.62 ± 0.03	0.75 ± 0.03	0.85 ± 0.02	0.70 ± 0.03
t=0.3	$\mathbf{0.64 \pm 0.03}$	$\mathbf{0.77 \pm 0.03}$	$\mathbf{0.86 \pm 0.02}$	$\mathbf{0.71 \pm 0.03}$
t=0.4	0.62 ± 0.03	0.77 ± 0.03	0.86 ± 0.02	0.71 ± 0.03
t=0.5	0.61 ± 0.03	0.75 ± 0.03	0.85 ± 0.02	0.69 ± 0.03

Table 6: Dataset B: Performance based on support radius $r = tR$. Parameters: detector=RK30, descriptor=PFH, $K = 100$.

	FT	ST	DCG	AP
t=0.05	0.52 ± 0.01	0.64 ± 0.01	0.81 ± 0.01	0.57 ± 0.02
t=0.1	0.78 ± 0.01	0.88 ± 0.01	0.95 ± 0.00	0.84 ± 0.01
t=0.2	$\mathbf{0.79 \pm 0.01}$	$\mathbf{0.88 \pm 0.01}$	$\mathbf{0.95 \pm 0.00}$	$\mathbf{0.85 \pm 0.01}$
t=0.3	0.72 ± 0.01	0.82 ± 0.01	0.92 ± 0.01	0.78 ± 0.01
t=0.4	0.66 ± 0.01	0.76 ± 0.01	0.89 ± 0.01	0.72 ± 0.01
t=0.5	0.61 ± 0.01	0.72 ± 0.01	0.87 ± 0.01	0.67 ± 0.01

5.5 Performance per shape class

Finally, we study how retrieval performance varies with the shape class. Table 7 shows AP for each of the 20 classes in Dataset A, and for three keypoint selection methods: $n = 1000$ random points, $n = 1000$ non-salient random points, and ground-truth salient features. "Armadillo" and "plier" have perfect retrieval performance, while other classes such as "spring" and "bearing" have poor performance (below 30%). Note that classes with poor AP typically consist of tube-like features. We argue that local features on such shape classes do not contain enough information to differentiate them from one another, which leads to poor retrieval performance.

Table 7: Dataset A: AP per shape class, using $n = 1000$ random points, $n = 1000$ random points in non-salient regions, and ground-truth keypoints GK. Parameters: descriptor=PFH, $r = 0.1R$, $K = 100$.

	$n = 1000$	$n = 1000$ (non-salient)	GK
armadillo	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
plier	1.00 ± 0.01	1.00 ± 0.00	0.87 ± 0.08
ant	0.99 ± 0.00	0.99 ± 0.00	0.98 ± 0.01
teddy	0.99 ± 0.01	0.98 ± 0.02	0.95 ± 0.04
fish	0.95 ± 0.03	0.95 ± 0.03	0.64 ± 0.05
glasses	0.93 ± 0.03	0.89 ± 0.05	0.91 ± 0.06
mechanic	0.91 ± 0.09	0.90 ± 0.09	0.77 ± 0.12
buste	0.89 ± 0.05	0.80 ± 0.06	0.80 ± 0.06
airplane	0.87 ± 0.04	0.84 ± 0.05	0.84 ± 0.03
hand	0.86 ± 0.07	0.79 ± 0.07	0.68 ± 0.07
human	0.82 ± 0.08	0.80 ± 0.09	0.55 ± 0.05
four leg	0.70 ± 0.08	0.70 ± 0.07	0.41 ± 0.05
chair	0.68 ± 0.07	0.62 ± 0.08	0.57 ± 0.07
table	0.53 ± 0.11	0.52 ± 0.11	0.51 ± 0.10
cup	0.53 ± 0.08	0.53 ± 0.08	0.44 ± 0.08
bird	0.47 ± 0.05	0.46 ± 0.05	0.32 ± 0.05
octopus	0.46 ± 0.04	0.44 ± 0.05	0.49 ± 0.05
vase	0.34 ± 0.05	0.35 ± 0.04	0.26 ± 0.03
bearing	0.27 ± 0.03	0.27 ± 0.04	0.25 ± 0.03
spring	0.22 ± 0.06	0.21 ± 0.06	0.24 ± 0.07
Average	0.72 ± 0.03	0.70 ± 0.03	0.62 ± 0.03

Table 8: Dataset A: Computational cost. These include computation of local descriptors, clustering, encoding and retrieval. RK50 refers to choosing $n = 50$ random points per shape.

	Timing (secs)	AvgNumKeypoints
RK	306	248 ± 11
MeshDOG	302	236 ± 12
ISS	178	106 ± 7
LK	167	69 ± 4
CK	152	56 ± 4
PK	148	53 ± 3
RK50	145	50 ± 0
GK	121	32 ± 1
MK	127	29 ± 2

6 Computational cost

Table 8 shows computation time spent on training a retrieval system (computing descriptors and encodings) and testing it (computing similarities between pairs of shapes in the database). We record computation times for retrieval based on every keypoint detector. Results show that although RK produces similar or better retrieval performance than other evaluated methods, retrieval computation is at least 2 times more costly than the alternatives. On the other hand, choosing $n = 50$ random points produces similar retrieval performance, with half the computational cost of RK.

7 Discussion

Our most surprising result is that random points outperform real salient points for shape retrieval on a generic dataset (Section 4 and 5.3). Even when the random points are restricted to non-salient regions they still outperform salient points (Section 5.3), implying that the distribution of non-salient features is important in recognizing 3D shapes. We provide possible explanations for these results.

The Bag of features approach loses spatial information. The approach encodes the distribution of local features, but does not include where these features are spatially located. This may be desirable for deformation invariance. However, there are several methods for dealing with this, using for instance diffusion distances [17]. Adding spatial information in the encoding may better differentiate between shapes with similar local features but different relative positions of these features. Further work should investigate how a spatially-sensitive approach will affect retrieval performance given salient points.

Salient points are often symmetric, in other words they are symmetric to another keypoint with an identical local neighbourhood. Examples of symmetric keypoints are human eyes, airplane wings, and table legs. Chen et al. [1] show that salient point sets selected by people are highly

symmetric (76% of all selected points). Thus, GK contains less information than we think, since it generates lists of duplicate features that may be redundant. Random points, on the other hand, are not symmetric and thus capture more useful information for retrieval.

Moreover, although salient features may contain some class-differentiating information, the global shape plays a more important role in 3D recognition than a few interesting points. However, we see in Section 5.1 that a global descriptor that looks at the shape as whole with no focus on local regions underperforms compared to local features encoding. This indicates that encoding local surface patches provides a better global description of a shape. Salient features may be more useful for specific tasks such as shape correspondence [31], where two shapes typically have a common class, and the problem is matching corresponding local neighbourhoods.

Saliency computes the relative importance of points in a shape. Thus saliency-based keypoints are not inherently class-specific. Our series of experiments examine how much salient points are representative of shape classes. Our results show that random points outperform keypoints, suggesting that the latter are not important in discriminating between shapes from different classes. Rather it may be more effective to compute class-specific features, and given a shape encode how it matches these representative features. This shape representation could not only tell us how similar two shapes are, but also help us better understand the similarity. This may be possible by extracting meaningful combinations of common features between the two shapes and matching these combinations to known classes. An investigation of such class-specific features for retrieval is an interesting direction for future work.

8 Conclusion

We evaluated keypoint detectors on their performance in shape retrieval based on selected saliency models. Using a random saliency model, RK, leads to better retrieval performance compared to other saliency models including ground-truth, although more computationally expensive. For a low computational cost, with similarly good retrieval performance, we recommend selecting a small fixed number of random points per shape.

Acknowledgement

Tasse was supported by a Google European Doctoral Fellowship and an IDB Cambridge International Scholarship. Kosinka was supported by the Engineering and Physical Sciences Research Council [EP/H030115/1]. The authors thank Victor Jouffrey and Daniel Rowlands for valuable discussions regarding the paper.

Table 9: Dataset A: Effect of GK-based local descriptors on retrieval. Parameters: detector=GK, $r = 0.1R$, $K = 100$.

	FT	ST	DCG	AP
PFH	$\mathbf{0.55 \pm 0.03}$	$\mathbf{0.68 \pm 0.03}$	$\mathbf{0.80 \pm 0.02}$	$\mathbf{0.62 \pm 0.03}$
FPFH	0.51 ± 0.03	0.63 ± 0.03	0.77 ± 0.02	0.57 ± 0.03
SHOT	0.44 ± 0.03	0.58 ± 0.03	0.72 ± 0.02	0.51 ± 0.03
SPIN	0.44 ± 0.03	0.56 ± 0.03	0.72 ± 0.02	0.50 ± 0.03

Table 10: Dataset A: Effect of LK-based local descriptors on retrieval. Parameters: detector=LK, $r = 0.1R$, $K = 100$.

	FT	ST	DCG	AP
PFH	$\mathbf{0.53 \pm 0.03}$	$\mathbf{0.66 \pm 0.03}$	$\mathbf{0.80 \pm 0.02}$	$\mathbf{0.61 \pm 0.03}$
FPFH	0.51 ± 0.03	0.63 ± 0.03	0.78 ± 0.02	0.57 ± 0.03
SHOT	0.46 ± 0.03	0.60 ± 0.03	0.75 ± 0.02	0.53 ± 0.03
SPIN	0.45 ± 0.03	0.57 ± 0.03	0.73 ± 0.02	0.52 ± 0.03

A Local feature descriptors

We show in Section 5.1 that combining RK feature detector with the PFH local descriptor produces better results than alternative descriptors. To show that PFH is indeed the better descriptor, we carry out a similar experiment using other keypoint detectors. We focus on LK and GK. Tables 9 and 10 both show that PFH performs better than other descriptors (FPFH, SHOT and SPIN) with different detectors. This supports our choice of PFH as the default local descriptor.

B Number of random points

Section 5.3 discusses the influence of the number of random points on shape retrieval, using results on Dataset A as a test case. It shows that using as few as $n = 50$ random points per shape outperforms human-selected keypoints (GK) on this dataset. Although there is no ground-truth keypoint data available for Dataset B and C, we can still analyze the effect of keypoint sampling size on these datasets. Tables 11 and 12 show retrieval performance for various sampling sizes, on Datasets A and B. Results show that for these 2 datasets consisting of whole models, retrieval performance increases with the random sampling size, up to $n = 50$ where performance stops improving significantly. There is no significant difference in performance on Dataset C (not reported), due to the poor performance of Bag of features for partial shape retrieval.

C Number of salient points

Finally we report the number of salient points generated by the evaluated saliency-based keypoint detection algorithms, ISS and MeshDOG. For each dataset, we also include the retrieval system that performs best on it thus far, according to the literature [16, 9, 10]. Tables 13– 15 present this data along with DCG and AP performance. These results support the argument that for the detectors evaluated in this

Table 11: Dataset A: Effect of random sampling size on retrieval. Parameters: descriptor=PFH, $r = 0.1R$, $K = 100$.

	FT	ST	DCG	AP
GK	0.55 ± 0.03	0.68 ± 0.03	0.80 ± 0.02	0.62 ± 0.03
RK	0.63 ± 0.03	0.73 ± 0.03	0.84 ± 0.02	0.70 ± 0.03
n=20	0.51 ± 0.03	0.66 ± 0.03	0.79 ± 0.02	0.58 ± 0.03
n=30	0.57 ± 0.03	0.72 ± 0.03	0.83 ± 0.02	0.65 ± 0.03
n=50	0.62 ± 0.03	0.76 ± 0.03	0.85 ± 0.02	0.70 ± 0.03
n=100	0.64 ± 0.03	$\mathbf{0.77 \pm 0.03}$	$\mathbf{0.86 \pm 0.02}$	$\mathbf{0.72 \pm 0.03}$
n=300	$\mathbf{0.64 \pm 0.03}$	0.76 ± 0.03	0.86 ± 0.02	0.72 ± 0.03
n=500	0.64 ± 0.03	0.76 ± 0.03	0.85 ± 0.02	0.72 ± 0.03

Table 12: Dataset B: Effect of random sampling size on retrieval. Parameters: descriptor=PFH, $r = 0.1R$, $K = 100$.

	FT	ST	DCG	AP
RK	0.86 ± 0.01	0.91 ± 0.01	0.96 ± 0.00	0.90 ± 0.01
n=20	0.66 ± 0.01	0.78 ± 0.01	0.90 ± 0.01	0.73 ± 0.01
n=30	0.78 ± 0.01	0.88 ± 0.01	0.95 ± 0.00	0.84 ± 0.01
n=50	0.85 ± 0.01	0.92 ± 0.01	0.97 ± 0.00	0.90 ± 0.01
n=100	$\mathbf{0.88 \pm 0.01}$	$\mathbf{0.93 \pm 0.01}$	$\mathbf{0.97 \pm 0.00}$	$\mathbf{0.92 \pm 0.01}$
n=300	0.87 ± 0.01	0.92 ± 0.01	0.96 ± 0.00	0.91 ± 0.01
n=500	0.87 ± 0.01	0.92 ± 0.01	0.96 ± 0.00	0.91 ± 0.01

paper, large number of local features and better coverage of the surface lead to better retrieval performance.

Table 13: Dataset A: Performance and average number of detected points per keypoint detector. Parameters: descriptor=PFH, $r = 0.1R$, $K = 100$. The reported best retrieval system on the dataset is based on covariance descriptors for local features[16].

	DCG	AP	n
best: [16]	**0.86**	N/A	600 (random)
RK	0.84 ± 0.02	0.70 ± 0.03	248 ± 11
MeshDOG	0.82 ± 0.02	0.65 ± 0.03	236 ± 12
ISS	0.82 ± 0.02	0.65 ± 0.03	106 ± 7
LK	0.80 ± 0.02	0.61 ± 0.03	69 ± 4
CK	0.75 ± 0.02	0.54 ± 0.03	56 ± 4
PK	0.77 ± 0.02	0.57 ± 0.03	53 ± 3
GK	0.80 ± 0.02	0.62 ± 0.03	32 ± 1
MK	0.76 ± 0.02	0.55 ± 0.03	29 ± 2

Table 14: Dataset B: Performance and average number of detected points per keypoint detector. Parameters: descriptor=PFH, $r = 0.1R$, $K = 100$. The reported best retrieval system on the dataset is based on an encoding of Localized Statistical Feature vectors [9].

	DCG	AP	n
best: [9]	**0.99**	N/A	$5K$ (random)
RK	0.96 ± 0.00	0.90 ± 0.01	312 ± 8
MeshDOG	0.96 ± 0.00	0.89 ± 0.01	150 ± 1
ISS	0.96 ± 0.00	0.88 ± 0.01	94 ± 1
LK	0.94 ± 0.01	0.85 ± 0.01	176 ± 5
CK	0.91 ± 0.01	0.75 ± 0.01	145 ± 5
PK	0.94 ± 0.01	0.83 ± 0.01	174 ± 5
MK	0.95 ± 0.00	0.88 ± 0.01	142 ± 4

Table 15: Dataset C: Performance and average number of detected points per keypoint detector. Parameters: descriptor=PFH, $r = 0.1R$, $K = 100$. The reported best retrieval system on the dataset uses similarity metric learning to generate rankings of 3D models for the given range scan queries

	DCG	AP	n	n_{scans}
best: [10]	**0.64**	**0.78**	N/A	N/A
RK	0.35 ± 0.02	0.06 ± 0.01	134 ± 8	136 ± 10
MeshDOG	0.34 ± 0.02	0.05 ± 0.02	90 ± 9	652 ± 74
ISS	0.34 ± 0.01	0.05 ± 0.01	28 ± 3	395 ± 42
LK	0.34 ± 0.01	0.05 ± 0.01	97 ± 6	54 ± 4
CK	0.34 ± 0.01	0.05 ± 0.01	78 ± 5	47 ± 4
PK	0.32 ± 0.01	0.04 ± 0.01	145 ± 13	155 ± 24
MK	0.32 ± 0.01	0.04 ± 0.01	224 ± 23	334 ± 48

References

[1] X. Chen, A. Saparov, B. Pang, T. Funkhouser, Schelling points on 3D surface meshes, ACM Trans. Graph. 31 (4) (2012) 29:1–29:12.

[2] R. Song, Y. Liu, R. R. Martin, P. L. Rosin, Mesh saliency via spectral processing, ACM Trans. Graph. 33 (1) (2014) 6:1–6:17.

[3] E. Shtrom, G. Leifman, A. Tal, Saliency detection in large point sets, in: ICCV 2013, 2013, pp. 3591–3598.

[4] A. Anonymous, Shape co-saliency and quantitative analysis of saliency models, under review (2015).

[5] F. Tasse, J. Kosinka, N. Dodgson, Cluster-based point set saliency, in: ICCV 2015, 2015, pp. 163–171.

[6] C. H. Lee, A. Varshney, D. W. Jacobs, Mesh saliency, ACM Trans. Graph. 24 (3) (2005) 659–666.

[7] G. Leifman, E. Shtrom, A. Tal, Surface regions of interest for viewpoint selection, in: CVPR 2012, 2012, pp. 414–421.

[8] D. Giorgi, S. Biasotti, L. Paraboschi, Shape retrieval contest 2007: Watertight models track (2007).

[9] Z. Lian, J. Zhang, S. Choi, H. ElNaghy, J. El-Sana, T. Furuya, A. Giachetti, R. A. Guler, L. Lai, C. Li, H. Li, F. A. Limberger, R. Martin, R. U. Nakanishi, A. P. Neto, L. G. Nonato, R. Ohbuchi, K. Pevzner, D. Pickup, P. Rosin, A. Sharf, L. Sun, X. Sun, S. Tari, G. Unal, R. C. Wilson, Non-rigid 3D Shape Retrieval, in: Eurographics Workshop on 3D Object Retrieval, 2015.

[10] A. Godil, H. Dutagaci, B. Bustos, S. Choi, S. Dong, T. Furuya, H. Li, N. Link, A. Moriyama, R. Meruane, R. Ohbuchi, D. Paulus, T. Schreck, V. Seib, I. Sipiran, H. Yin, C. Zhang, Range Scans based 3D Shape Retrieval, in: Eurographics Workshop on 3D Object Retrieval, 2015.

[11] R. B. Rusu, N. Blodow, M. Beetz, Fast point feature histograms (FPFH) for 3D registration, in: ICRA, 2009, pp. 1848–1853.

[12] H. Dutagaci, C. P. Cheung, A. Godil, Evaluation of 3D interest point detection techniques via human-generated ground truth, Vis. Comput. 28 (9) (2012) 901–917.

[13] J. Sun, M. Ovsjanikov, L. Guibas, A concise and provably informative multi-scale signature based on heat diffusion, in: SGP 2009, 2009, pp. 1383–1392.

[14] S. Salti, F. Tombari, R. Spezialetti, L. Di Stefano, Learning a descriptor-specific 3d keypoint detector, in: The IEEE International Conference on Computer Vision (ICCV), 2015.

[15] J. Sivic, A. Zisserman, Video Google: A text retrieval approach to object matching in videos, in: ICCV, 2003.

[16] H. Tabia, H. Laga, D. Picard, P.-H. Gosselin, Covariance descriptors for 3D shape matching and retrieval, in: CVPR, 2014.

[17] A. M. Bronstein, M. M. Bronstein, L. J. Guibas, M. Ovsjanikov, Shape Google: Geometric words and expressions for invariant shape retrieval, ACM Trans. Graph. 30 (1) (2011) 1:1–1:20.

[18] J. Philbin, O. Chum, M. Isard, J. Sivic, A. Zisserman, Lost in quantization: Improving particular object retrieval in large scale image databases, in: CVPR'08, 2008, pp. 1–8.

[19] F. Perronnin, J. Sánchez, T. Mensink, Improving the fisher kernel for large-scale image classification, in: ECCV'10, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 143–156.

[20] M. A. Savelonas, I. Pratikakis, K. Sfikas, Partial 3d object retrieval combining local shape descriptors with global fisher vectors, in: Eurographics Workshop on 3D Object Retrieval, 2015.

[21] R. B. Rusu, A. Holzbach, G. Bradski, M. Beetz, Detecting and segmenting objects for mobile manipulation, in: Proceedings of IEEE Workshop on Search in 3D and Video (S3DV), held in conjunction with the 12th IEEE International Conference on Computer Vision (ICCV), Kyoto, Japan, 2009.

[22] A. E. Johnson, M. Hebert, Surface matching for object recognition in complex 3-d scenes, Image and Vision Computing 16 (1998) 635–651.

[23] F. Tombari, S. Salti, L. Di Stefano, Unique signatures of histograms for local surface description, in: Proceedings of the 11th European Conference on Computer Vision Conference on Computer Vision: Part III, ECCV'10, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 356–369.

[24] R. B. Rusu, S. Cousins, 3D is here: Point Cloud Library (PCL), in: IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China, 2011.

[25] G. Lavoué, Combination of bag-of-words descriptors for robust partial shape retrieval, The Visual Computer 28 (9) (2012) 931–942.

[26] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics Bulletin 1 (6) (1945) 80–83.

[27] F. Tombari, S. Salti, L. Di Stefano, Performance evaluation of 3d keypoint detectors, Int. J. Comput. Vision 102 (1-3) (2013) 198–220.

[28] Y. Zhong, Intrinsic shape signatures: A shape descriptor for 3d object recognition, in: Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on, 2009, pp. 689–696.

[29] A. Zaharescu, E. Boyer, K. Varanasi, R. Horaud, Surface feature detection and description with applications to mesh matching, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2009, pp. 373–380.

[30] S. P. Lloyd, Least squares quantization in pcm, IEEE Transactions on Information Theory 28 (1982) 129–137.

[31] H. Zhang, A. Sheffer, D. Cohen-Or, Q. Zhou, O. van Kaick, A. Tagliasacchi, Deformation-driven shape correspondence, Computer Graphics Forum 27 (5) (2008) 1431–1439.