

Quantitative Analysis of Saliency Models

F. P. Tasse¹ and J. Kosinka^{1,2} and N. A. Dodgson^{1,3}

¹Rainbow Group, University of Cambridge, United Kingdom

²Scientific Visualization and Computer Graphics, University of Groningen, Netherlands

³Computer Graphics, Victoria University of Wellington, New Zealand

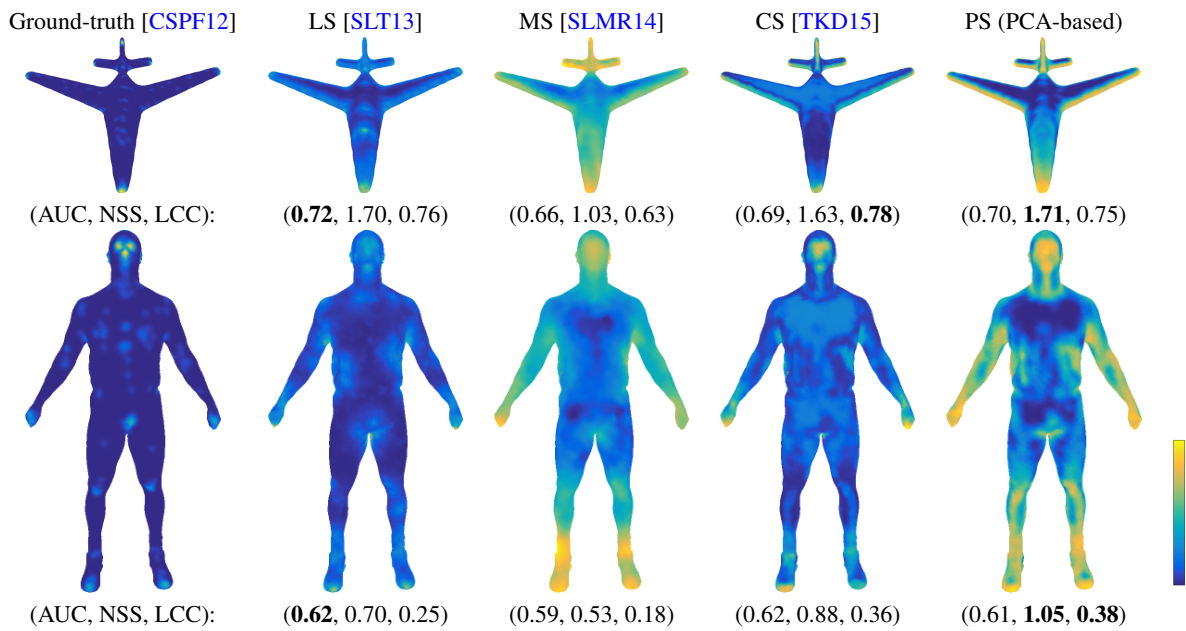


Figure 1: Examples of saliency maps generated by selected saliency models. The scores underneath each map are evaluation metrics that indicate how well the map compares to ground-truth. The metric for the top-performing saliency map is indicated in **bold**.

Abstract

Previous saliency detection research required the reader to evaluate performance qualitatively, based on renderings of saliency maps on a few shapes. This qualitative approach meant it was unclear which saliency models were better, or how well they compared to human perception. This paper provides a quantitative evaluation framework that addresses this issue. In the first quantitative analysis of 3D computational saliency models, we evaluate four computational saliency models and two baseline models against ground-truth saliency collected in previous work.

Categories and Subject Descriptors (according to ACM CCS):
I.3.5 [Computer Graphics]: Computational Geometry and Object
Modelling—Curve, surface, solid, and object representations

1. Introduction

We introduce three metrics for quantitative analysis and comparison of 3D saliency methods.

Despite recent interest in saliency detections of 3D surfaces, there are no evaluation metrics available, and no previous quantitative evaluation of 3D saliency models, making it impossible to ob-

jectively compare one method against another. Previous work evaluates results qualitatively, by showing 3D renderings of saliency maps of a few shapes to the reader. This approach does not objectively determine whether one saliency model is better than another. With the absence of a saliency evaluation benchmark, new techniques cannot be compared to previous methods on a common dataset. To the best of our knowledge, no previous work performs a quantitative evaluation of computational models to ground truth. We propose such a quantitative evaluation, based on surface feature points acquired from users by Chen et al. [CSPF12]. Inspired by Judd et al.’s benchmark of computational models to predict human eye fixations in images [JDT12], we evaluate 4 saliency models, and 2 baseline models against ground truth 3D saliency on watertight meshes and simulated range scans.

Quantitative studies on shape saliency are few. Howlett et al. [HHO05] empirically show, using eye fixations data collected from users, that mesh saliency exists and is useful in mesh simplification. Kim et al. [KVJG10] compare mesh saliency [LVJ05] to human eye movements on 5 meshes, and show that a computational saliency models human eye fixations significantly better than chance. A larger study by Chen et al. [CSPF12], on 400 models from Giorgi et al. dataset [GBP07], collects ground-truth saliency data via salient point selection by users. The study shows strong correlations between saliency as perceived by humans and mesh properties such as curvature.

An unexplored area of empirical research on 3D saliency detection is a saliency benchmark that compares saliency models against one another and human performance. Benchmarks for 3D keypoint detection have been proposed [DCG12], as well as quantitative analysis that explain how mesh properties such as curvature are correlated with ground-truth saliency using information theory statistics [CSPF12]. However, none of these works compare saliency performance using raw saliency maps. In contrast, there are well-established 2D image saliency benchmarks [JDT12, BSI13], which provide a framework for quantitative evaluation of different 2D saliency algorithms. Thus, we use available ground-truth saliency data [CSPF12] and present evaluation metrics, inspired by the 2D case, to compare computational saliency models.

Our contributions are three metrics for estimating the performance of saliency models, evaluated on six saliency models’ ability to predict ground truth saliency, on a dataset of 400 triangular meshes from the SHREC’07 Watertight models track [GBP07] (SHREC07) and a synthetic dataset of 4800 single-view 3D scans.

2. 3D Saliency methods

We review recent computational saliency models, with a particular emphasis on methods evaluated in this paper.

2.1. Saliency of large point sets (LS) [SLT13]

Shtrom et al. [SLT13] propose the first method that supports saliency detection on large points sets. Inspired by Leifman et al.’s [LST12] work on triangular meshes, saliency is a combination of point distinctiveness at two scales with point association, a function that assigns higher saliency to regions near foci of attention.

Formally, saliency of a point p_i is:

$$S(p_i) = \frac{1}{2}(D_{low}(p_i) + D_{high}(p_i)) + \frac{1}{2}A_{low}(p_i),$$

where D_{low} and D_{high} are the low-level and high-level distinctiveness, and $A_{low}(p_i)$ is the point association. The point distinctiveness of p_i is computed as the average dissimilarity between p_i and other points. This dissimilarity measure between two points is the χ^2 distance between their local descriptors, weighted by their Euclidean distance to give more influence to nearby points. Shtrom et al. use Fast Point Feature Histograms (FPFH) as local descriptors [RBB09] using two scales, denoted here with r_{low} and r_{high} , to compute local descriptors for low-level and high-level distinctiveness, respectively. Finally, foci of attention used in A_{low} are computed by taking 20% of the points with highest low-level distinctness. A_{low} attributes higher saliency to points near these foci of attention. This saliency model is described as generating plausible saliency maps for small and large point sets such as city scans.

2.2. Mesh saliency via spectral processing (MS) [SLMR14]

Song et al. [SLMR14] propose a spectral-based approach, described as more robust compared to previous mesh saliency methods that focused on analysing changes in local vertex properties [PKG03, LVJ05, CCFM08, UH08]. The new approach uses spectral properties of a mesh at multiple scales using the n lowest frequencies of its log-Laplacian spectrum L . The log-Laplacian spectrum amplifies variation in the low-frequency parts of the Laplacian spectrum and detects the most ‘fundamental’ saliencies. Single-scale saliency is computed by taking the absolute difference between L and a locally averaged log-Laplacian spectrum A , and mapping it back to the spatial domain. This single-scale method captures globally salient regions but ignores local details. The authors address this issue by computing saliency on a group of smoothed meshes at scales $\{\epsilon^2, 2\epsilon^2, 3\epsilon^2, 4\epsilon^2, 5\epsilon^2\}$ for a given ϵ . Finally, multi-scale saliency is obtained by applying the logarithm operator to a smoothed summation of the absolute differences of saliency maps at consecutive scales. This method is described as being able to capture both globally important regions and local saliencies.

2.3. Cluster-based point set saliency (CS) [TKD15]

Tasse et al. [TKD15] propose a cluster-based saliency model presented as being able to detect fine-scale saliency with better time complexity. They segment point sets into K clusters, and compute cluster saliency as a sum of cluster distinctiveness and spatial distribution. The point-level saliency is obtained by smoothing cluster-level saliency. Cluster distinctiveness is based on the mean FPFH of points belonging to that cluster, using a method similar to Shtrom et al.’s [SLT13] low-level distinctiveness. It compares a cluster descriptor to every other cluster, with more influence accorded to nearby clusters, so that a region is more distinctive if it is different from its surroundings. Cluster spatial distribution computes the spatial variance of geometrically similar clusters. Both heuristics capture both local and global saliencies. Local descriptors are computed at a single scale r , as opposed to Shtrom et al.’s two-scale method.

2.4. Other methods

Early saliency models compute a multi-scale representation of a mesh and observe how a local vertex property such as curvature, surface variation or normal displacement changes at different scales [PKG03, LVJ05, CCFM08, UH08]. These methods are tightly linked to local properties which are not robust against noise and topological changes.

Other saliency models achieve robustness and speed by first segmenting a mesh into patches represented by descriptors, followed by a ranking process that specifies patch distinctiveness [GCO06, WSZL13, TCL*14]. Patch distinctiveness may then be assigned to individual vertices by smoothing with a Gaussian kernel. This distinctiveness computation can also be done with robust vertex descriptors but this comes with a higher computational cost [GMGP05]. Leifman et al. [LST12] compute vertex distinctiveness at a fine and coarser scale, and combine this with the idea that vertices close to foci of attention, such as extremities, are salient. Saliency models discussed so far, with a few exceptions [PKG03, UH08], require topological information and thus cannot support other shape representations such as point clouds. Recent saliency models described in the above section, such as Shtrom et al. and Tasse et al., focus on point sets. They fill a gap in the literature that has become important owing to the proliferation of low-range scanning devices that produce point sets that may not always be suitable for full reconstruction. Thus, our paper evaluates selected saliency models on both typical watertight meshes and simulated range scans.

Few methods use learning for shape saliency detection. As part of their large-scale user study on mesh saliency, Chen et al. [CSPF12] propose two regression models, based on the collected data. The first model is trained over the whole dataset, and the second model only on meshes from the same class. The second model produces better results than the first. Supervised learning for saliency detection is challenging to apply to a wider range of datasets due to lack of training data. Shilane and Funkhouser [SF07] learn distinctive regions without a need for saliency training data, given a collection of meshes partitioned into classes. Meshes are split into regions and the saliency of a region corresponds to how well that region defines the mesh class and differentiates it from other classes. This method identifies large salient regions, such as the head of a four-legged animal, but does not detect finer-scale features.

3. PCA-based saliency

Both Tasse et al. [TKD15] and Shtrom et al. [SLT13]

use FPFH to compute saliency, under the premise that this descriptor captures the geometry of a local neighbourhood and can be used to best compute similarity between neighbourhoods. It is robust to noise and sampling density. We propose a new, simple, saliency method where saliency is the absolute value of the descriptor projected onto the largest principal axis after mean centering. It is interesting to investigate how this simple model compares with recent computational models described above. Before providing details on the analysis of PCA-based saliency and other methods, we first describe FPFH descriptors.

The FPFH of a point p is computed by taking a distance-weighted average of the Simplified Point Feature Histogram (SPFH) of nearby points within a support sphere of radius r . SPFH captures mean curvature around a point using a multi-dimensional histogram that captures the sampled surface variations. The SPFH of a point p is a histogram of angular variations between p and other points in its neighbourhood. Given a pair of points p_i and p_j with their respective unit normals n_i and n_j , the angular variations between the two are computed by first defining an orthonormal Darboux coordinate frame (u, v, w) by

$$u = n_i, \quad v = u \times \frac{p_j - p_i}{\|p_j - p_i\|}, \quad w = u \times v.$$

The relative difference between p_i and p_j with their estimated normals

is then encoded by the triple (α, ϕ, θ) , where

$$\alpha = v \cdot n_j, \quad \phi = u \cdot \frac{p_j - p_i}{\|p_j - p_i\|}, \quad \theta = \arctan2(w \cdot n_j, u \cdot n_j).$$

Note that we use the PCL library [RC11] to compute (α, ϕ, θ) and take their absolute values to get invariance to reflection [SLT13].

An 11-bin histogram is computed for each angular parameter, α , ϕ , θ , and the three histograms are then concatenated to form a 33-dimensional vector. This produces a scale-invariant descriptor robust to noise, which is a requirement for handling point sets from range scans. If the input are unoriented point sets, the normal at a point is approximated by analysing the eigenvectors of the covariance matrix created from its nearest neighbours. To consistently orient normals, we propagate a seed orientation through a minimum spanning tree of the Riemann graph constructed over the point set, as proposed by Hoppe et al. [HDD*92]. Given the above definition of SPFH, denoted by $\hat{H}(p)$, the final FPFH descriptor $H(p)$ is a 33-dimensional weighted average of SPFH of neighbouring points:

$$H(p) = \hat{H}(p) + \frac{1}{|\mathcal{N}(p)|} \sum_{q \in \mathcal{N}(p)} \frac{1}{\|p - q\|_2} \hat{H}(q),$$

where $\mathcal{N}(p)$ is the set of neighbours q of p with $\|p - q\| < r$.

Our PCA-based saliency approach (**PS**) is a variation of the above FPFH-based models that dramatically reduces the amount of information that needs to be compared when processing a shape. For each 3D model, we apply PCA to a large matrix consisting of the FPFHs of each point, and retain only the strongest component. This component should capture the distinctiveness of each point. This can be seen as saliency detection purely based on geometric similarity, with any spatial consideration ignored.

4. Methodology

We present our setup for a quantitative analysis of computational saliency models.

4.1. Datasets

We use two datasets to evaluate saliency performance on typical watertight meshes and simulated range scans.

4.1.1. Watertight meshes

Ground-truth saliency is obtained from Chen et al.’s user study [CSPF12], which comprises salient points, referred to by the authors as Schelling points. Users were asked to select points that were likely to be selected by other users. Each mesh in the SHREC07 dataset was annotated by at least 22 participants. Chen et al. also compute a scalar field over a mesh by smoothing, with a Gaussian filter, the frequency with which each vertex was selected by all participants. We use this scalar field as ground-truth saliency.

4.1.2. Simulated range scans

In addition to mesh data, we are also interested in how saliency models deal with challenging data such as single-view point scans. Using the SHREC07 dataset above, we generated a large dataset of 4800 synthetic scans by rendering 12 range images from each of 400 meshes, and converting the range images to point sets. Similarly to Sipiran et al.’s [SMB*13] generation of range images, each mesh is enclosed in a regular icosahedron, and each vertex of the icosahedron is used as a camera position. Scanned points are intersections between the mesh and rays shot from the camera. These points are in the same object coordinate system as their base mesh, making it possible to map a point on a scan to a point on the base mesh and record their ground-truth saliency value.

Because some of our tested saliency models require meshes as input [SLMR14], we reconstruct partial meshes from points sets using the Greedy Projection Triangulation method available in the PCL library. Triangulation is performed by projecting the local neighborhoods of points along their normals and linking unconnected points.

The above approach generates 4800 single-view range scans with ground-truth data that can be used in our evaluation.

4.2. Evaluation metrics

Saliency evaluation benchmarks in 2D images are well-established. We adapt evaluation scores used in these benchmarks to 3D saliency. Each saliency model is compared against ground-truth (GS) using the following 3 metrics:

Area under the ROC curve (AUC): The Receiver Operating Characteristic (ROC) curve is obtained by thresholding the saliency map into a binary mask that separates positive samples (salient points) from negative samples (non-salient points) and, for different threshold values, plotting the true positive rate against the false positive rate. This metric is commonly used to compare saliency models in the 2D case [JDT12, LB13]. The ideal saliency model has AUC of 1.0. AUC disregards regions with no saliency, and focuses on the ordering of the saliency values. Other, more selective, metrics are needed to support evaluation.

Normalized scanpath saliency (NSS): Also widely used in comparing 2D saliency maps to human eye fixations [LB13], NSS measures saliency values at fixation points along each user’s eye scanpath. In our 3D case, we consider points selected by users as fixation points. For each participant, a NSS score is computed by a weighted sum of the computational saliency at points selected by

the participant. The final NSS metric is the average over all participants. The higher this metric, the closer the evaluated computational model is to ground-truth since interest points selected by users should have large saliency values.

Linear correlation coefficient (LCC): This coefficient measures the strength of the linear relationship between two variables [BSI13]. The coefficient ranges between -1 and 1 , with values closer to 0 implying a weak relationship. We use its absolute value as the metric score. With X the ground-truth and Y the saliency map under consideration, the correlation coefficient is

$$LCC(X, Y) = \frac{|\text{cov}(X, Y)|}{\sigma_X \sigma_Y},$$

where σ_X and σ_Y are the standard deviation of X and Y , respectively, and $\text{cov}(X, Y)$ is the covariance between the two variables. When $\sigma_Y = 0$, we set $LCC(X, Y) = 0$. Note that one of the things captured by LCC is whether two distributions have peaks at the same place, but this is limited by the fact that LCC is greatly influenced by the shapes of these peaks. Despite its limitations, we include LCC since it is a popular metric for measuring linear relationships between distributions, and other saliency metrics can help mitigate these limitations.

Computational models that are the closest to ground-truth have high AUC and NSS scores and high absolute value of LCC. We use the Wilcoxon rank-sum test [Wil45], a non-parametric alternative to the two-samples t-test, at a 0.05 significance level to report statistically significant differences between saliency performances of competing methods.

4.3. Selected saliency methods

We selected four computational models from the literature discussed in Section 2, based on an informal assessment of their likely quality and on the availability of data or source code. We either obtained source code from the authors or implemented the method proposed in their papers:

Saliency of large point sets (LS) [SLT13]: We use our implementation of Shtrom et al.’s technique [SLT13] to generate the saliency maps used in the evaluation.

Mesh saliency via spectral processing (MS) [SLMR14]: Source code provided by the authors. Note that the Laplacian on a range scan with isolated points is not guaranteed to be semi positive definite, and thus may not have real eigenvectors. In these cases, we assign a default saliency value of 0 to all points.

Cluster-based point set saliency (CS) [TKD15]: Source code provided by the authors.

PCA-based saliency (PS) We implemented this saliency model as described in Section 3.

Table 1 gives the parameter values used for the four saliency methods above, and introduced in Section 2.

In addition to the above computational saliency models, we evaluate the following two baseline models:

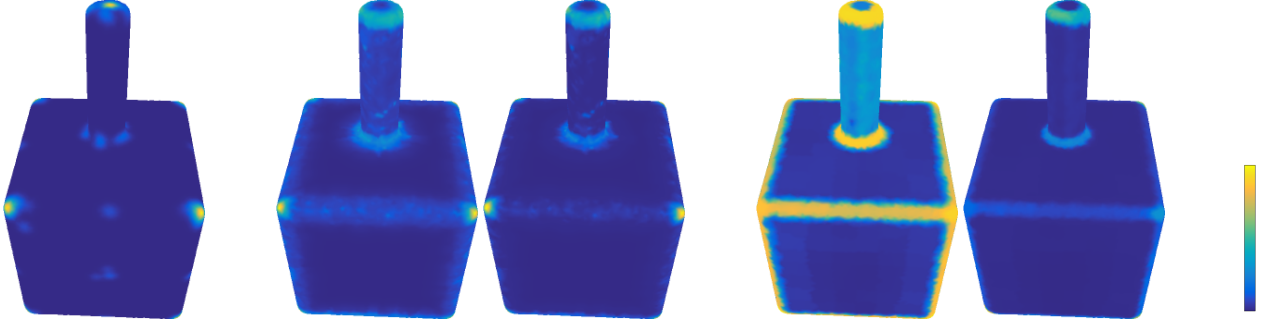


Figure 3: Saliency maps after their histograms were matched to ground truth. From left to right: Ground-truth, LS before matching, LS after matching, CS before matching, CS after matching. This matching allows a fair comparison of various saliency models.

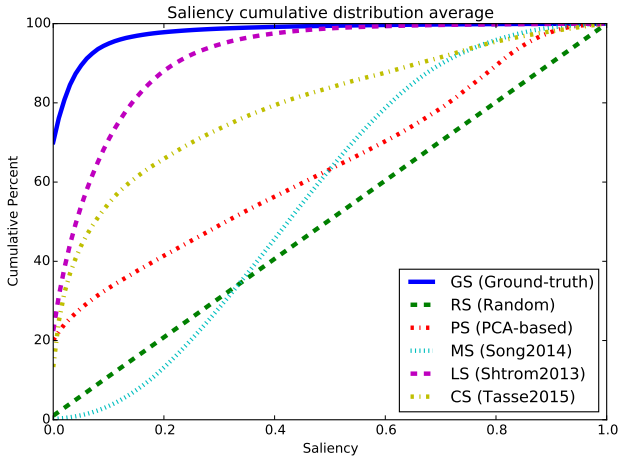


Figure 2: Cumulative distribution functions (CDF) of maps generated by the evaluated saliency models. Each curve is the average of CDFs of SHREC07 saliency maps generated by the corresponding method. Random saliency has a CDF similar to that of uniform CDF, whereas maps generated by spectral mesh saliency have CDFs similar to the normal CDF. The graph shows that different saliency models produce different saliency distributions, hence the need to adjust each shape saliency CDF to the average ground-truth CDF (histogram matching) before quantitative evaluation so that fair comparisons are made.

Chance (RS): We test saliency performance when saliency values are randomly assigned. Computational saliency models should have a better performance than a random model. For each point on a mesh, we choose at random a value between 0.0 and 1.0 to set its saliency value.

Human performance (HS): Here, we are interested in how well one or more human participants’ predictions differ from the consensus of all the participants. We investigate how well saliency data collected from n_p participants predict ground-truth saliency from other participants. By default, we use $n_p = 1$ but we vary n_p to produce the results in Figure 5. We can say that a computational

Method	Parameters	
LS [SLT13]	$r_{low} = 0.01R$	$r_{high} = 0.1R$
MS [SLMR14]	$\epsilon = 0.004R$	$n = 9$
CS [TKD15]	$r = 0.02R$	$K = 100$
PS	$r = 0.01R$	

Table 1: Parameters per evaluated saliency method. We used parameter values recommended in the corresponding papers and presented in Section 2. R is the bounding sphere radius.

saliency model predicts saliency as well as a human, if its performance is similar to the performance of the average human.

The six saliency models we are evaluating vary significantly in how saliency is distributed over the mesh, with some models having more salient regions than others. This difference in the distributions of saliency maps is illustrated in Figure 2. The distribution of HS is equivalent to that of GS when summed up over all participants. For a fair comparison, the histogram of each saliency map is matched to that of the ground-truth, similarly to the previous work on 2D saliency evaluation [JDT12]. We refer to the average histogram of a ground-truth saliency map as the reference histogram. Given a saliency map, we find the mapping that optimally transforms its cumulative distribution function towards that of the reference histogram [GW01, Section 5.6.4]. This ensures that all saliency maps have the same distribution of saliency values. Examples of saliency maps with histograms matched with ground-truth are displayed in Figure 3.

5. Experimental results

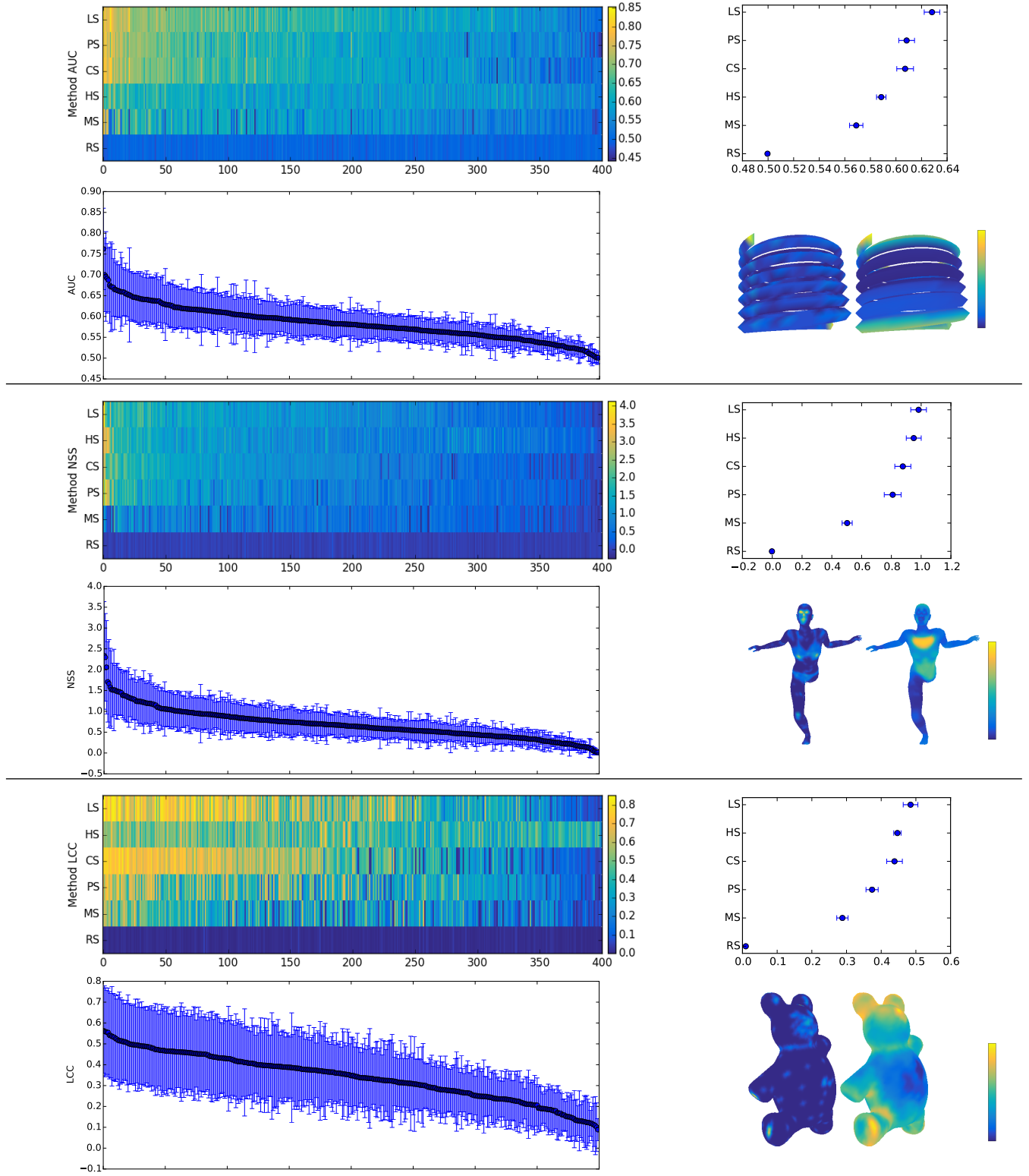
We present the performance of the six selected saliency models, under the AUC, NSS and LCC metrics. We also present results on how human performance varies with the number of participants, and human consistency.

5.1. Saliency on watertight meshes

First, we report results on the SHREC07 dataset.

5.1.1. Model performances

We discuss the performance results illustrated in Figure 4.



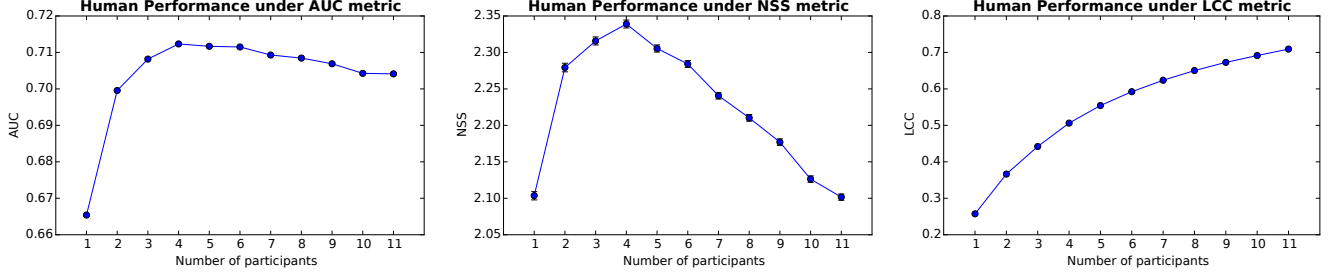


Figure 5: Performance of n_p participants to predict saliency from n_p participants, as n_p varies from 1 to 11.

Table 2: AUC performance per shape class in SHREC07. For each class, we average the saliency AUC scores of all shapes belonging to that class. The last column reports the average of all seven AUC scores. The table shows, for instance, that it is easier to detect saliency on a *mechanic* shape than a *spring*. It also shows that no tested saliency model can yet detect saliency on a *teddy* shape better than a human.

Class	LS (Shtrom2013)	PS (PCA-based)	CS (Tasse2015)	HS (1 vs all)	MS (Song2014)	RS (Random)	Avg-AUC
<i>mechanic</i>	0.69 ± 0.03	0.71 ± 0.03	0.70 ± 0.03	0.64 ± 0.02	0.53 ± 0.03	0.50 ± 0.00	0.63 ± 0.05
<i>fish</i>	0.67 ± 0.02	0.67 ± 0.02	0.67 ± 0.01	0.61 ± 0.01	0.63 ± 0.01	0.50 ± 0.00	0.62 ± 0.03
<i>armadillo</i>	0.67 ± 0.01	0.66 ± 0.01	0.66 ± 0.01	0.62 ± 0.01	0.60 ± 0.01	0.50 ± 0.00	0.62 ± 0.03
<i>airplane</i>	0.67 ± 0.02	0.64 ± 0.02	0.63 ± 0.02	0.60 ± 0.01	0.62 ± 0.02	0.50 ± 0.00	0.61 ± 0.03
<i>table</i>	0.67 ± 0.03	0.62 ± 0.02	0.63 ± 0.03	0.61 ± 0.02	0.58 ± 0.02	0.50 ± 0.00	0.60 ± 0.03
<i>vase</i>	0.62 ± 0.02	0.63 ± 0.02	0.62 ± 0.02	0.58 ± 0.01	0.61 ± 0.02	0.50 ± 0.00	0.59 ± 0.02
<i>buste</i>	0.63 ± 0.03	0.64 ± 0.03	0.62 ± 0.03	0.59 ± 0.01	0.56 ± 0.02	0.50 ± 0.00	0.59 ± 0.03
<i>four-legged</i>	0.62 ± 0.03	0.61 ± 0.03	0.60 ± 0.03	0.59 ± 0.01	0.60 ± 0.03	0.50 ± 0.00	0.59 ± 0.03
<i>cup</i>	0.62 ± 0.02	0.61 ± 0.02	0.61 ± 0.02	0.57 ± 0.01	0.59 ± 0.03	0.50 ± 0.00	0.58 ± 0.03
<i>hand</i>	0.63 ± 0.02	0.60 ± 0.02	0.61 ± 0.02	0.58 ± 0.01	0.58 ± 0.01	0.50 ± 0.00	0.58 ± 0.02
<i>chair</i>	0.66 ± 0.02	0.58 ± 0.03	0.59 ± 0.03	0.62 ± 0.02	0.55 ± 0.02	0.50 ± 0.00	0.58 ± 0.03
<i>ant</i>	0.63 ± 0.01	0.58 ± 0.01	0.61 ± 0.01	0.60 ± 0.01	0.57 ± 0.01	0.50 ± 0.00	0.58 ± 0.02
<i>bearing</i>	0.64 ± 0.03	0.63 ± 0.03	0.65 ± 0.03	0.55 ± 0.01	0.50 ± 0.02	0.50 ± 0.00	0.58 ± 0.04
<i>bird</i>	0.62 ± 0.02	0.60 ± 0.02	0.60 ± 0.02	0.57 ± 0.01	0.56 ± 0.02	0.50 ± 0.00	0.58 ± 0.02
<i>plier</i>	0.62 ± 0.01	0.58 ± 0.01	0.60 ± 0.01	0.56 ± 0.01	0.56 ± 0.01	0.50 ± 0.00	0.57 ± 0.02
<i>human</i>	0.59 ± 0.02	0.59 ± 0.02	0.57 ± 0.02	0.58 ± 0.01	0.57 ± 0.02	0.50 ± 0.00	0.57 ± 0.02
<i>octopus</i>	0.62 ± 0.02	0.57 ± 0.02	0.55 ± 0.02	0.60 ± 0.01	0.53 ± 0.01	0.50 ± 0.00	0.56 ± 0.02
<i>teddy</i>	0.56 ± 0.01	0.57 ± 0.01	0.57 ± 0.01	0.60 ± 0.01	0.55 ± 0.01	0.50 ± 0.00	0.56 ± 0.02
<i>glasses</i>	0.57 ± 0.01	0.55 ± 0.01	0.52 ± 0.02	0.56 ± 0.01	0.53 ± 0.01	0.50 ± 0.00	0.54 ± 0.02
<i>spring</i>	0.55 ± 0.01	0.53 ± 0.02	0.55 ± 0.02	0.54 ± 0.01	0.55 ± 0.01	0.50 ± 0.00	0.54 ± 0.02

AUC is the area under the ROC curve generated by plotting true positive rate against false positive rate. Figure 4 (top) shows the performance of the selected saliency models, under the AUC metric. All models are significantly better, on average, than chance. Saliency models based on the PPFH descriptor all perform better than both human performance HS and spectral mesh saliency MS. Among these descriptor-based methods, LS has the highest AUC performance. There is no statistically significant difference between the other two descriptor-based techniques PS and CS.

NSS is the average of saliency values at human-selected keypoints. Figure 4 (middle) presents models' performance under this metric. Similarly to the AUC metric, all models perform better than chance under NSS, with spectral mesh saliency MS having a statistically lower mean score than others. LS performs as well as human performance, with no statistically significant difference between their two mean scores. CS is not statistically different from PS. Note that under the NSS metric, HS is one of the top performing mod-

els, in contrast with its low AUC score. This is because AUC is influenced by the ordering of saliency values within a ground-truth saliency map. NSS uses the user-selected keypoints directly in its formulation with no consideration of how many times a point was selected by participants.

LCC estimates the strength of the relationship between the distributions of a shape saliency map and its ground-truth saliency. Results of models' performance under LCC are presented in Figure 4 (bottom). Models' rankings under this metric are similar to rankings by NSS scores. The key difference is that there is no significant difference between CS and human performance HS. Thus CS performs better under LCC.

In summary, LS [SLT13] achieves the best performance under all metrics. There is little difference between the performance of the other two PPFH-based saliency models, namely CS [TKD15] and PCA-based saliency. It is important to note that the PCA approach

to 3D saliency, first introduced in this paper, uses dramatically less data in its comparisons and yet produces competitive results compared to the state-of-the-art. Of course, while PS uses only a single value in its comparison (the first principal axis of the PCA result), it comes at the expense of a preprocessing step: running PCA on an $n \times 33$ matrix to get the single saliency value, where n is the number of points. This is reasonably fast and the preprocessing is amortized over many thousands of queries.

5.1.2. Performance per shape classes

We analyse how the selected saliency models perform for each of the 20 classes in SHREC07. This is interesting because it gives us insight into which classes have the worst saliency detection and thus could benefit from future work in the field. Table 2 shows AUC scores for each class and saliency model. For the class *teddy*, human performance (HS) outperforms all other saliency models. In this particular case, human participants know that real-life shapes from this class have facial features and thus consider the face to be salient even if the 3D shape presented to them has no discriminating features. This is an example of humans using their prior experience in saliency detection, which is not yet possible for unsupervised saliency models. Man-made shapes such as *mechanic*, *airplane* and *chair* are easier for saliency detection due to their sharp features and simple structure.

5.1.3. Human performance and consistency

Figure 5 illustrates how well n_p participants predict the ground-truth averaged over n_p participants, under AUC, NSS and LCC. AUC performance stops improving significantly when n_p reaches 3 participants. NSS only increases significantly when n_p increases to 2 and starts to drop off for $n_p > 4$. The drop is not statistically significant. In contrast with these metrics LCC keeps improving with increasing number of participants. LCC computes the correlation between the distributions of a saliency map and ground-truth saliency, which explains why this value increases with the number of participants. The negligible difference in AUC and NSS performance with $n_p \geq 3$ may suggest that on average 3 participants are enough to obtain ground-truth saliency.

5.2. Saliency on simulated range scans

We present saliency performance on range scans generated from the SHREC07 dataset used above. Figure 6 presents average performances per metric and Figure 7 illustrates saliency maps on a few scans. We did not include HS in the evaluated methods for this simulated dataset since the human participants did not directly work with range scans.

AUC Similarly to results on SHREC07, LS has the top AUC performance. All other performance differences are significant: CS is better than PS, which performs better than MS. Thus, CS performs similarly to PS on the set of watertight meshes, but outperforms the latter on range scans.

NSS LS is the top performing method, followed by CS and PS. Again, the difference between CS and PS is only significant on

range scans. MS and RS have the lowest NSS scores. This is because MS is penalized for its spectral-based approach that is designed for 2D manifolds and is not able to support any range scan.

CC All performance differences between methods were significant, with the exception of CS and LS. This reinforces the indication that CS compares better with other methods when computed over range scans.

In summary, LS remains the top performing method on 2D manifolds and range scans. It is followed by CS, another method designed for point sets. MS performs poorly (although better than RS) on range scans since it requires a 2D manifold to obtain real solutions of the Laplacian.

6. Discussion

This paper reported an analysis of saliency models on watertight meshes and range scans. We did not report timings since implementations were built in different programming environments, including C++, Matlab and Python.

Results show, in Section 5.1.1, that the two-scale approach of LS appears to detect saliency peaks similar to ground-truth irrespective of the data type. The other multi-scale approach MS has the lowest performance, after chance, possibly due to the fact that it was designed to capture both large and small salient regions, while the ground-truth collection process is focused on local salient patches. An advantage of MS is that it is invariant to isometric transformations, since the Laplacian is an intrinsic operator. We expect that on a dataset of non-rigid shapes MS will have better performance compared to the alternatives. This would be an interesting investigation for future work.

We also introduce a simple saliency method based on PCA analysis of FPFH descriptors. Results on watertight meshes show that it outperforms HS, and performs similarly to CS on some metrics. On range scans, PS only performed better than MS and RS although the differences with LS and CS were not large. Given the simplicity of the PCA-based saliency and the small performance loss compared to better saliency models, it may be a good candidate for applications where accuracy loss in the saliency map can be tolerated. Examples of such applications are shape simplifications [SLMR14] and viewpoint selection [SLT13].

Our evaluation against ground-truth is based on saliency maps that were extracted from multiple human-selected keypoints. Such saliency maps do not accurately reflect the result of low-level visual attention. Running a similar quantitative analysis on eye fixations data will help determine which methods are closer to emulating low-level saliency.

7. Conclusion

We introduce the first saliency evaluation framework for 3D shapes, based on three performance metrics. We compare six saliency models including chance, human performance and a PCA-based approach. The performance results showed that all models performed better than chance, and Shtrom et al.'s method [SLT13] has, on average, the best scores on both watertight meshes and simulated

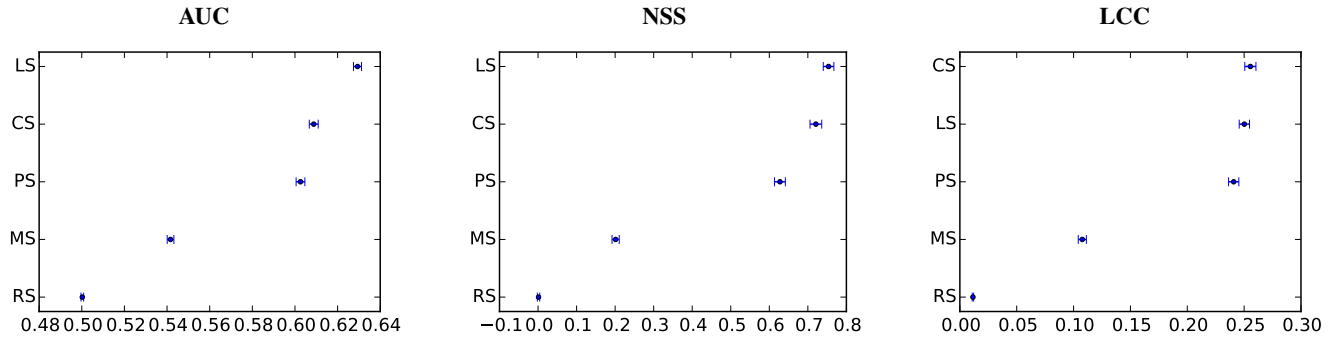


Figure 6: Saliency performance on simulated range scans, ordered by the average metric score per method.

range scans. PCA-based saliency performed as well as cluster-based point set saliency [TKD15], and significantly better than spectral mesh saliency [SLMR14]. We are releasing our benchmark online, so that previous methods not tested here and future saliency techniques can be evaluated objectively.

References

- [BSI13] BORJI A., SIHITE D. N., ITTI L.: Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Trans. Im. Proc.* 22, 1 (2013), 55–69. [2, 4](#)
- [CCFM08] CASTELLANI U., CRISTANI M., FANTONI S., MURINO V.: Sparse points matching by combining 3D mesh saliency with statistical descriptors. *Computer Graphics Forum* 27, 2 (2008), 643–652. [2, 3](#)
- [CSPF12] CHEN X., SAPAROV A., PANG B., FUNKHOUSER T.: Schelling points on 3D surface meshes. *ACM Trans. Graph.* 31, 4 (July 2012), 29:1–29:12. [1, 2, 3, 4, 10](#)
- [DCG12] DUTAGACI H., CHEUNG C. P., GODIL A.: Evaluation of 3D interest point detection techniques via human-generated ground truth. *Vis. Comput.* 28, 9 (Sept. 2012), 901–917. [2](#)
- [GBP07] GIORGI D., BIASOTTI S., PARABOSCHI L.: Shape retrieval contest 2007: Watertight models track, 2007. [2](#)
- [GCO06] GAL R., COHEN-OR D.: Salient geometric features for partial shape matching and similarity. *ACM Trans. Graph.* 25, 1 (Jan. 2006), 130–150. [3](#)
- [GMGP05] GELFAND N., MITRA N. J., GUIBAS L. J., POTTMANN H.: Robust global registration. In *SGP 2005* (2005). [3](#)
- [GW01] GONZALEZ R. C., WOODS R. E.: *Digital Image Processing*, 2nd ed. Addison-Wesley, 2001. [5](#)
- [HDD*92] HOPPE H., DEROSE T., DUCHAMP T., McDONALD J., STUETZLE W.: Surface reconstruction from unorganized points. In *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques* (1992), SIGGRAPH '92, pp. 71–78. [3](#)
- [HHO05] HOWLETT S., HAMILL J., O'SULLIVAN C.: Predicting and evaluating saliency for simplified polygonal models. *ACM Trans. Appl. Percept.* 2, 3 (July 2005), 286–308. [2](#)
- [JDT12] JUDD T., DURAND F., TORRALBA A.: A benchmark of computational models of saliency to predict human fixations. In *MIT Technical Report* (2012). [2, 4, 5](#)
- [KVJG10] KIM Y., VARSHNEY A., JACOBS D. W., GUIMBRETIERE F.: Mesh saliency and human eye fixations. *ACM Trans. Appl. Percept.* 7, 2 (Feb. 2010), 12:1–12:13. [2](#)
- [LB13] LE MEUR O., BACCINO T.: Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior Research Methods* 45, 1 (2013), 251–266. [10.3758/s13428-012-0226-9. 4](#)
- [LST12] LEIFMAN G., SHTROM E., TAL A.: Surface regions of interest for viewpoint selection. In *CVPR 2012* (2012), pp. 414–421. [2, 3](#)
- [LVJ05] LEE C. H., VARSHNEY A., JACOBS D. W.: Mesh saliency. *ACM Trans. Graph.* 24, 3 (July 2005), 659–666. [2, 3](#)
- [PKG03] PAULY M., KEISER R., GROSS M. H.: Multi-scale feature extraction on point-sampled surfaces. *Comput. Graph. Forum* 22, 3 (2003), 281–290. [2, 3](#)
- [RBB09] RUSU R. B., BLODOW N., BEETZ M.: Fast point feature histograms (FPFH) for 3D registration. In *Proceedings of the 2009 IEEE International Conference on Robotics and Automation* (Piscataway, NJ, USA, 2009), ICRA'09, IEEE Press, pp. 1848–1853. [2](#)
- [RC11] RUSU R., COUSINS S.: 3D is here: Point Cloud Library (PCL). In *ICRA* (May 2011), pp. 1–4. [3](#)
- [SF07] SHILANE P., FUNKHOUSER T.: Distinctive regions of 3D surfaces. *ACM Trans. Graph.* 26, 2 (June 2007). [3](#)
- [SLMR14] SONG R., LIU Y., MARTIN R. R., ROSIN P. L.: Mesh saliency via spectral processing. *ACM Trans. Graph.* 33, 1 (Feb. 2014), 6:1–6:17. [1, 2, 4, 5, 8, 9, 10](#)
- [SLT13] SHTROM E., LEIFMAN G., TAL A.: Saliency detection in large point sets. In *ICCV 2013* (2013), pp. 3591–3598. [1, 2, 3, 4, 5, 7, 8, 10](#)
- [SMB*13] SIPIRAN I., MERUANE R., BUSTOS B., SCHRECK T., JOHAN H., LI B., LU Y.: SHREC'13 Track: Large-Scale Partial Shape Retrieval Using Simulated Range Images. In *Eurographics Workshop on 3D Object Retrieval* (2013). [4](#)
- [TCL*14] TAO P., CAO J., LI S., LIU L., LIU X.: Mesh saliency via ranking unsalient patches in a descriptor space. *Computers and Graphics* (2014). [3](#)
- [TKD15] TASSE F. P., KOSINKA J., DODGSON N.: Cluster-based point set saliency. In *ICCV* (Dec 2015), pp. 163–171. [1, 2, 3, 4, 5, 7, 9, 10](#)
- [UH08] UNNIKRIKSHNAN R., HEBERT M.: Multi-scale interest regions from unorganized point clouds. In *Workshop on Search in 3D (S3D)* (June 2008). [2, 3](#)
- [Wil45] WILCOXON F.: Individual comparisons by ranking methods. *Biometrics Bulletin* 1, 6 (1945), 80–83. [4](#)
- [WSZL13] WU J., SHEN X., ZHU W., LIU L.: Mesh saliency with global rarity. *Graphical Models* 75, 5 (2013), 255–264. [3](#)

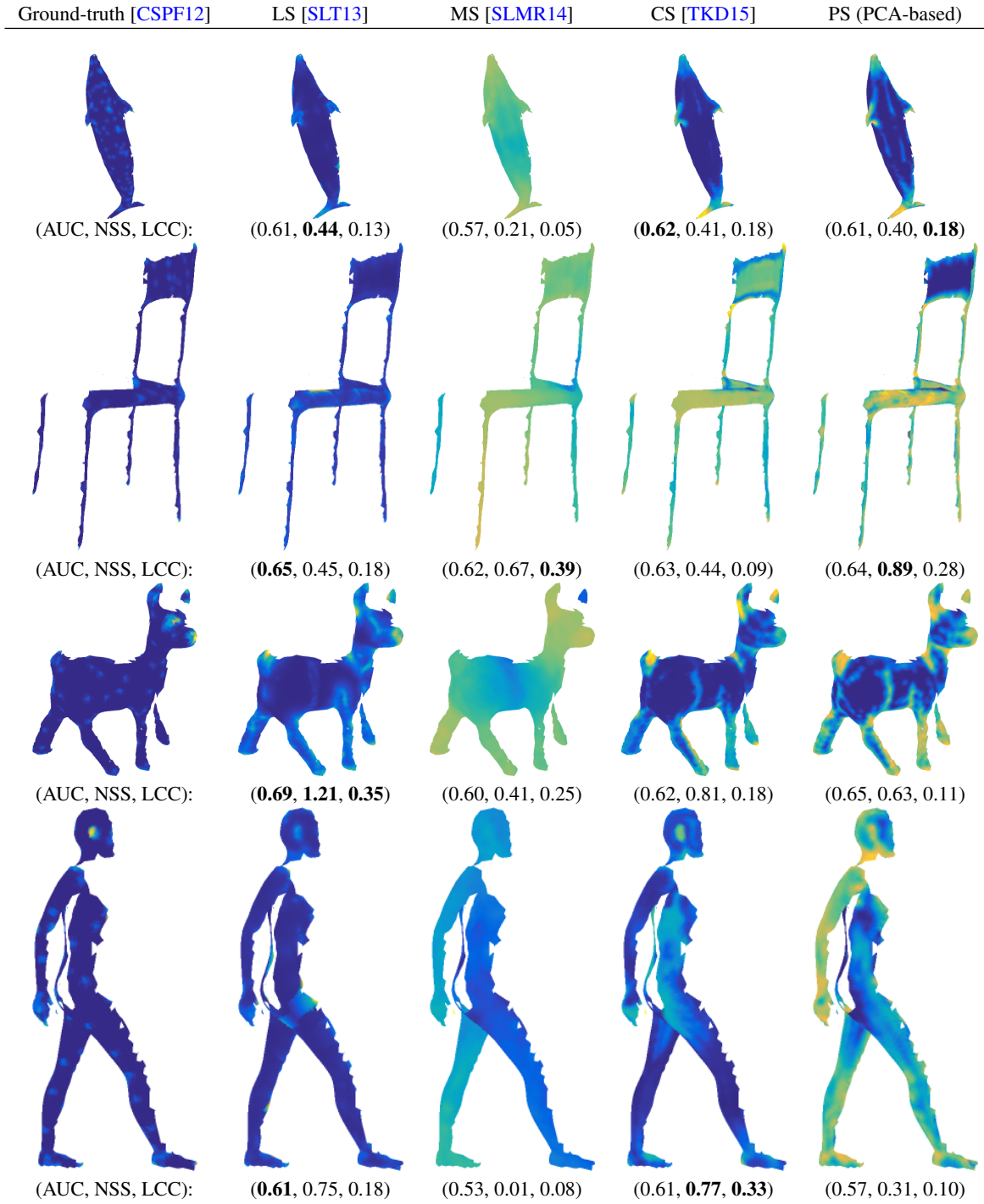


Figure 7: Saliency maps (before histogram matching) on 4 simulated range scans, obtained by virtually scanning watertight meshes in the SHREC07 dataset.