Vis Comput
DOI 10.1007/s00371-012-0746-4

ORIGINAL ARTICLE

Evaluation of 3D interest point detection techniques via human-generated ground truth

Helin Dutagaci · Chun Pan Cheung · Afzal Godil

Abstract In this paper, we present an evaluation strategy based on human-generated ground truth to measure the performance of 3D interest point detection techniques. We provide quantitative evaluation measures that relate automatically detected interest points to human-marked points, which were collected through a web-based application. We give visual demonstrations and a discussion on the results of the subjective experiments. We use a voting-based method to construct ground truth for 3D models and propose three evaluation measures, namely False Positive and False Negative Errors, and Weighted Miss Error to compare interest point detection algorithms.

Keywords 3D interest points · 3D salient points · 3D shape analysis

1 Introduction

In the area of 3D model processing, detection of interest points is essential for many applications, such as registration, mesh simplification, mesh segmentation, viewpoint selection, and object matching and retrieval. Use of interest points to match 3D shapes has the advantage of providing local features that are semantically significant and also invariant to rotation, scaling, noise, deformation, and articulation. This approach is also suitable for 3D range image recognition, and partial matching.

Interest points, also referred to as feature points, salient points, or keypoints, are those points which are distinctive in their locality, and are present and stable at all instances of an object, or of its category of objects. In this work, we are particularly concerned with analyzing the subjective judgments of humans about 3D interest points, and relating this analysis to automatic interest point detection algorithms. This analysis will provide human-generated ground truth data, which in turn will make it possible to quantitatively measure how much an algorithm relates to the human perception. The ground truth data can also serve to tune the parameters of an existing algorithm, or to inspire the development of new algorithms.

A widely used evaluation criterion is "repeatability", which measures the stability of the detected points on a particular object with respect to various transformations that object undergoes, such as deformation, change in resolution, and addition of noise. Our motivation, on the other hand, is to measure the detection and localization success of the algorithms with respect to human-generated ground truth.

Most of the 3D interest point detection algorithms developed in the last decade defined functions summarizing the geometrical content of localities on a 3D model in multiple scales, and selected local extrema of those functions as interest points. This approach is in accordance with the fact that humans respond more to significant local changes on the surface. However, we aimed to investigate humans' choices empirically on a number of 3D models, typically used in 3D shape research community.

We designed experiments to measure how close the points detected by an algorithm are to those considered as in-

H. Dutagaci (✉)
Electrical-Electronics Engineering Department, Eskisehir Osmangazi University, Eskisehir, Turkey
e-mail: helindutagaci@gmail.com

C.P. Cheung · A. Godil
National Institute of Standards and Technology (NIST), Gaithersburg, USA

A. Godil
e-mail: godil@nist.gov

Published online: 29 June 2012

terest points by human subjects. We developed a web-based application where human subjects were asked to mark the interest points of a model. Using the ground truth we have constructed from the human-marked points, we compared six different interest point detection techniques based on "false negative" and "false positive" errors, and "weighted miss" error. The data, including the 3D models, interest points marked by human subjects, the ground-truth points, as well as the evaluation code are available at our benchmark site [28].

2 Related work

2.1 Interest point detection techniques

In this section, we briefly mention feature detection algorithms that detect isolated points of interest on 3D triangular meshes. More detailed discussions on current interest point detection algorithms can be found in [2] and [1].

The methods dedicated to the detection of interest points on 3D mesh models are relatively recent. Most of them rely on local surface descriptors, such as curvature, extrema of which are assumed to correspond to candidate interest points. It is common practice to employ a multi-scale approach, where the algorithm analyzes the 3D surface at successive scales to search for interest points at various levels of detail.

One of the earliest work aiming to detect interest points [6] involves the use of integral volume descriptor which is related to the surface curvature and which is invariant to rotation and translation of the 3D object. A ball is placed at a vertex of the object, and the integral volume descriptor at that vertex corresponds to the volume of the intersection of the ball with the interior of the object. A histogram of the descriptor over the model is calculated, and points that correspond to the least populated bins are selected as candidate interest points. The least populated bins indicate there are rare occurrences of certain values of the descriptor, thus point to special points on the model. By using balls of varying radii, the authors calculate the curvature-related descriptor at multiple levels of detail of the model [6].

Another early approach is based on mesh saliency, which is defined at each vertex as a function of the differences of Gaussian-weighted mean curvatures at successive scales (Lee [11]). The various scales are obtained by filtering the mean curvature with Gaussian filters of varying standard deviation. High values of mesh saliency indicate that a vertex is consistently salient across various scales of the mesh, thus vertices with high saliency points are considered as interest points.

Castellani et al. [3] define another saliency measure: They apply Gaussian filtering directly on the vertex positions rather than the curvature values. Difference-of-Gaussians (DoG) are calculated at various scales, and vertices that are highly displaced after the filtering are marked as interest point candidates. The DoG approach is also used in [26] and [25]. In [24], the mesh is filtered with a set of Laplacians of Gaussian (LoG) to construct a pyramid. Points with local minima in both spatial and scale dimensions are declared as interest points.

Walter et al. [23] extend the 2D SUSAN operator to 3D meshes to compute the saliency degree on the vertices. 2D SUSAN operator is a corner detection method for 2D intensity images [21]. USAN (Univalue Segment Assimilating Nucleus) is a measure of how similar a center pixel's intensity is to those in its neighborhood. Pixels with small USAN (SUSAN) are assumed to be on corners since the center pixel will be different from other pixels in its neighborhood. Walter et al. [23] apply this method to 3D meshes by using mean curvature instead of the intensity values.

Sipiran and Bustos [19, 20] use a 3D extension of the 2D Harris operator, which is based on the local autocorrelation of an image. They detect the local maxima of the Harris response as candidate interest points and then reduce the set either by thresholding or clustering. More details for this method is provided in Sect. 3.3.

Mian et al. [16], define a local coordinate system—which is invariant to global rotation and translation of the 3D object—around a point using the cropped surface surrounding it. They calculate the covariance matrix of the points on the cropped surface and use this covariance matrix to calculate the ratio between the first two principal axes of the surface. A high ratio indicates unsymmetrical surface around the point, a property assumed to underlie keypoints. Novatnack and Nishino [17] parameterize a 3D mesh model onto a 2D plane, and construct a dense surface normal map. They build a scale-space by convolving the normal map with a set of Gaussian filters, and detect corners on each scale separately (see Sect. 3.5).

Hu and Hua [9] operate on Laplace–Beltrami spectral domain instead of spatial domain. They define the geometry energy on the vertices as a function of eigenfunctions and eigenvalues of the spectrum. A point is selected as an interest point if it remains as a local maximum of the geometry energy function within several successive frequencies. Thus the distinctiveness of an interest point is required to be stable within a portion of the spectrum. As another approach related to Laplace–Beltrami spectral analysis, Sun et al. [22] use heat kernel signature of the mesh. A point is chosen as an interest point where this function is a local maximum (see Sect. 3.6).

In this paper, we analyze the results of six of these algorithms with respect to the human-generated ground truth: Mesh saliency [11], salient points defined by Castellani et

al. [3], 3D-Harris [19], 3D-SIFT [7], scale-dependent corners [17], and Heat Kernel Signature (HKS) [2, 22]. We give detailed descriptions of these methods in Sect. 3.

2.2 Evaluation methods

There are a number of ways authors have used to demonstrate the success of their interest point detection algorithms: (1) Visualization of detected points on sample 3D mesh models; (2) End results of the ultimate task to which the detection algorithm serves, such as recognition or retrieval performance, or accuracy of registration; (3) Repeatability rate.

Repeatability rate is defined as the percentage of the detected points that are common in two different instances of a scene or an object [2, 18]. Usually, two detected points on the two instances are considered to be common if one falls within a neighborhood of the other, and the size of the neighborhood is denoted by ϵ. The repeatability is then referred to as ϵ-repeatability.

SHREC'10 and SHREC'11 robust feature detection and description benchmarks [1, 2] evaluate and compare 3D salient point detection algorithms. The evaluation is based on the repeatability of the detected points under a variety of transformations.

The SHREC'10 dataset [2] consists of three models (human, dog, horse) and their transformed versions. Each model has gone under 45 different transformations, such as changes in topology, sampling, scale, and addition of noise and holes. Three feature detection approaches are compared at [2]: Heat kernel-based signatures [22], Salient points of Castellani et al. [3], and 3D Harris Features [19].

The SHREC'11 feature detection dataset [1], on the other hand, consists of only one shape class (human), with 55 different transformations. The evaluation measure is again "repeatability" rate, and five detection algorithms are compared, namely, 3D-Harris [19], Mesh-DoG [25], Mesh SIFT [15], Mesh-Scale Dog [1], and Shape MSER [12].

Our previous workshop paper [5] was a first attempt to complement the analysis in the SHREC feature detection contests. In contrast to measuring the robustness of the detectors with respect to transformations, we aimed to relate the detected points to some human-generated ground truth. In this current paper, we substantially increased the number of human subjects, provided more solid and generalizable results, and a more detailed discussion. In addition to False Negative Error (FNE) and False Positive Error (FPE), we introduce Weighted Miss Error, which takes into account how frequently a point is marked by human subjects.

There is a recent and very interesting work that relates the previous mesh saliency computations to human eye movements [10]. In the study, five models were presented to six subjects from 10 different views, and the subjects' eye fixations were recorded. The mesh saliency [11] was computed on the 3D models, and the amount of correlation between mesh saliency values and eye fixation points was measured. The authors concluded that mesh saliency has better correlation with human eye fixations than a random model. We believe that, in contrast to human subjects' conscious decisions on which points are important, human eye fixations on 3D models can give the salience of points on a lower level of consciousness. Hence, it can give more insight to the response of human visual system to 3D interest points. Once the locations of the eye fixations are determined, our evaluation scheme based on FNE, FPE, and WME is suitable to support such an experimental setup.

In our experiments, the subjects mark model points that they "think" are important. An alternative and useful approach could be to show human subjects interest points detected by an algorithm, and ask them if they could recognize the category of the object. However, this approach would require redoing the experiment with human subjects each time a new algorithm (or a new parameter setting) was to be evaluated. In our setting, we collect human-marked interest points in advance so that researchers can make use of the available data to assess new algorithms.

3 Interest point detection techniques evaluated in this work

3.1 Mesh saliency

Mesh saliency [11] is based on the local curvature over the surface. The mean curvature at each vertex is weighted by two Gaussian filters, one with scale twice the other. The absolute difference between the weighted curvatures at two scales corresponds to the mesh saliency at that scale pair. The procedure is repeated for a number of different scale pairs, then the total mesh saliency at a vertex is calculated as the sum of mesh saliency values at these scale pairs.

Candidate interest points are picked from the local maxima of the total mesh saliency function. A vertex is marked as a local maximum if its total mesh saliency is higher than all its neighboring vertices. Then the candidate points with a saliency measure higher than a threshold are selected as final interest points. We set the threshold as the average of the total mesh saliency over the local maxima.

3.2 Salient points

Castellani et al. [3] also adopted a multi-resolution approach and defined another measure of saliency on the 3D mesh model. In their approach, instead of filtering the curvature values, they filter the 3D locations of the vertices via Gaussians, and they base their saliency measure on the amount of displacement of the vertices from those of the original mesh.

For each scale, two Gaussian filters, one with twice the scale of the other, are applied on the mesh vertices. The difference between the two filtered models corresponds to the DoG map (F_s) at that particular scale. F_s at each vertex is a 3D vector measuring the displacement of that vertex within twice the scale s. F_s is projected onto the normal of the vertex to obtain a scalar quantity, which is then referred to as "scale-map".

The scale-map is normalized to a fixed range of values, and an inhibition process is applied to enhance the peaks of the map. Then, a non-maximum suppression step is implemented to detect interest points. A local maximum with an inhibited saliency value higher than the 30 % of the global maximum is assigned as an interest point.

We refer to this method as "Salient points" to be consistent with the terminology in [3] and [2]. In this study, we used the "Mesh Tool" program available at the authors' web site [31].

3.3 3D-Harris

The 3D-Harris operator is the 3D extension of the 2D corner detection method of Harris and Stephens [8], and is based on first order derivatives along two orthogonal directions on the 3D surface [19]. Using these derivatives, a 2×2 matrix E is constructed for each vertex of the 3D mesh. The derivatives are calculated via fitting a quadratic surface to a neighborhood of the vertex.

The authors [19], define a neighborhood around each vertex x, and calculate the centroid of the points in the neighborhood. This point set is translated so that the centroid defines the origin of a local coordinate frame. In order to achieve rotation invariance of the local coordinate frame; the following steps are performed: A plane is fit to the point set using Principal Component Analysis, and the eigenvector with the lowest eigenvalue corresponds to the normal of the fitting plane. The point set is rotated so that the normal coincides with the z-axis of the local coordinate frame. Then the points are re-translated to make the vertex of analysis coincide with the origin. After the transformations, a quadratic surface is fit to the surface patch.

The first order derivatives along x and y directions are evaluated analytically at the point $(0, 0)$, which corresponds to the vertex of analysis. Then matrix E is defined as

$$E = \begin{bmatrix} A & C \\ C & B \end{bmatrix} \tag{1}$$

where $A = f_x(x, y)^2$, $B = f_y(x, y)^2$, and $C = f_x(x, y) \times f_y(x, y)$. The Harris operator value at the vertex is calculated as $H(x) = \det(E) - k(\mathrm{tr}(E))^2$. The authors [19] have set k to 0.04.

After calculating the Harris response on all the vertices of the 3D mesh, the local maxima are selected considering 1-ring neighborhood of a vertex. If the Harris response of a vertex is higher than those of all its immediate neighbors, then the vertex is declared as a local maxima. In [19], among these local maxima, the authors select a constant fraction of the total number of vertices with the highest Harris response as final interest points.

3.4 3D-SIFT

The 3D-SIFT technique [7] described here operates on 3D voxel space; therefore it involves a voxelization step. After voxelization, in parallel to the SIFT approach in [14], a scale space is constructed by applying 3D Gaussian filters with increasingly large scales to the voxelized model. If the voxelized model is denoted by a binary function $M(x, y, z)$, then each layer of the scale space is represented by its convolution with a 3D Gaussian function. Then, the Difference of Gaussian (DoG) for each level is computed by subtracting the original model from the scaled model at the corresponding level.

The extrema points are detected by searching the DoG space in both spatial and scale dimensions. The extrema points which are located on the surface are declared as interest points. Notice that these interest points are located on a voxel grid. Their locations are mapped back to the 3D space where the original mesh was defined, and the closest vertices are marked as final interest points.

3.5 Scale-dependent corners

Novatnack and Nishino [17] also built a scale-space representation of the model; however they analyzed the scales independent of each other to detect scale-dependent corners. We will refer to their method as "SD-corners" method.

As the first step, the vertices of a mesh model are unwrapped onto a 2D plane through embedding. Out of the 2D embedding, a "distortion map" is computed. The distortion map encodes the relative change in the model edge lengths after they have been mapped from the 3D surface onto the 2D plane. Then, the surface normals at the embedded vertices are interpolated to obtain a dense and regular "normal map" over the 2D plane. This 2D vector field (normal map) is filtered with Gaussians of varying scales to obtain a scale-space representation. The Gaussian kernels are modified using the distortion map so that the distance between two points on 2D can be corrected to match the geodesic distance on 3D. Then, first and second order partial derivatives of the normal map are calculated at each scale.

The authors define two types of geometric corner: Points that have high curvature isotropically, and points that have high curvature in at least two distinct tangential directions. They compute the Gram Matrix of first order partial derivatives of the normal map at each point. If the maximum eigenvalue of the Gram Matrix is high at a point then the point is

considered to have a high corner response. Some of these candidate corners may reside on edges rather than on corners, so they are eliminated using second order derivatives of the normal map.

The corners are separately detected at each scale, and then corners at different scales are collected in a single set. Finally, these corners detected in 2D domain are mapped back onto the surface of the 3D model.

3.6 HKS-based interest point detection

Feature point detection based on Heat Kernel Signature (HKS) was proposed by Sun et al. [22], and has gained considerable interest due to its high repeatability rate with respect to various transformation of an object [2]. The computation of HKS involves the application of the Laplace–Beltrami operator on the triangular mesh model. The response of the operator is a positive semi-definite matrix of size $N \times N$, where N is the number of vertices in the 3D model. This matrix is then decomposed into its eigenvectors and eigenvalues, denoted as λ_i and Φ_i, respectively. The heat kernel signature over the mesh evolves with a time variable t, and is defined by the expression $h_t(x,x) = \sum_{i=0}^{K} e^{-\lambda_i t} \Phi_i(x)\Phi_i(x)$.

For this study, we used the FEM-based code that was developed by Chung and Taylor [4] in order to calculate the Laplace–Beltrami eigenfunctions and eigenvalues. Then, we constructed HKS using 300 eigenvectors ($K = 300$), and for time 0.05 ($t = 0.05$). The interest points of the model, especially tips of extremities and protrusions, are assumed to correspond to the local maxima of HKS. As in [2], we declared a vertex x as an interest point if its HKS value ($h_t(x,x)$) is larger than HKS values of all other vertices in the 2-neighborhood of vertex x.

4 Subjective experiments

4.1 The 3D models

The 3D model set used in our experiments consists of 43 triangular meshes. Some models are standard models that are widely used in 3D shape research, such as Armadillo, David's head, and Utah teapot. We chose some of the models from The Stanford 3D Scanning Repository [29] and some others from the SHREC'2007 watertight model database [30]. The dataset can be downloaded from our benchmark site [28].

4.2 User interface for collecting ground truth

We created a web-page where users can login using an alias and participate in the experiments [27]. Figure 1 shows a snapshot of the user interface. The user is shown the 3D models, one at a time. The user is free to rotate the object in 3D. He/she is asked to mark the interest points on the 3D model, then to click on the submit button to proceed to the next model.

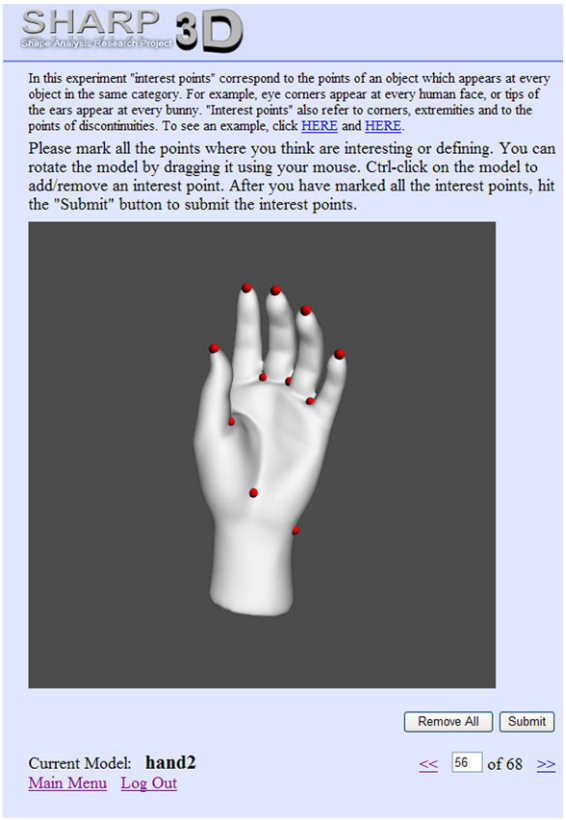


Fig. 1 User interface for marking interest points

5 Ground truth

Human judgment of interest points is subjective by nature. Figure 2 shows two models marked by three different subjects. Some people tend to elaborately mark points on the smallest details (subject a in Fig. 2), while others choose far less points (subject b in Fig. 2). There are also different choices in locating the interest points; for example, subjects do not mark interest points on the same exact location around the smooth corners of the chair in Fig. 2.

We look for some consensus among the users in order to merge all the marked points into a final set of ground truth interest points. It is also necessary to reject outliers and incorrectly marked points, and discard small variations of localization. We have two criteria while constructing the ground truth: The radius of an interest region, and the number of users n that marked a point within that interest region.

We set the radius of interest region as σd_M, where d_M stands for the model diameter; i.e. the largest Euclidean dis-

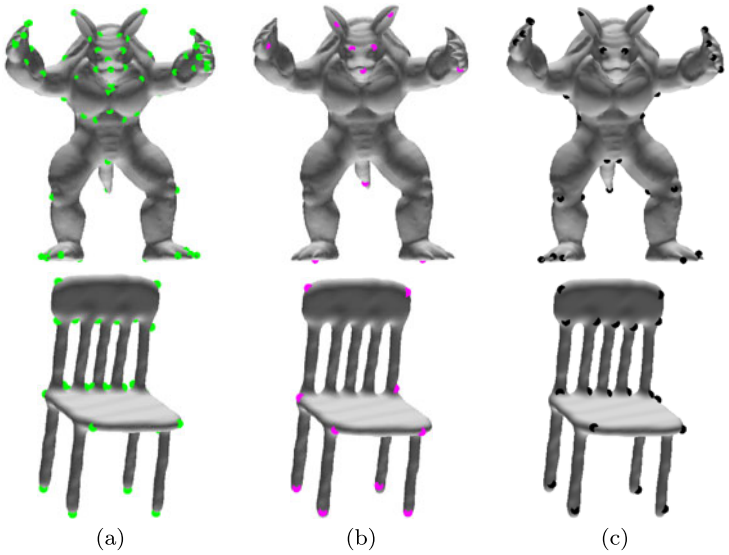


Fig. 2 Models marked by three different subjects

tance between all pairs of vertices of model M. We group all the interest points (marked by distinct subjects) whose geodesic distances to each other are less than $2\sigma d_M$. If the number of points in the group is less than n we discard that group. Otherwise, we select a representative among the group, and set it as a ground truth interest point. The point with the minimum sum of geodesic distances to the other points in the group is selected as the representative. Notice that two groups can be overlapping, i.e. can have vertices in common. If the distance between two representatives turn out to be less than $2\sigma d_M$, the representative with the smaller number of group points is discarded from the ground truth interest point set.

Figures 5 and 6 shows models with the interest points gathered from 23 subjects (first column), and the final ground truth points (second column). In addition to the locations of the ground truth points, we also keep the number of subjects who marked within the corresponding interest region. We call this number, the prominence of a ground truth interest point. As will be seen in Sect. 6, the prominence will play an important role in defining our new evaluation measure, namely Weighted Miss Error.

We will denote the set of ground truth points obtained with the parameters n and σ as $\mathcal{G}_M(n, \sigma)$ for a particular model M. These two parameters highly determine the final set of ground truth interest points. With high n, we have less number of ground truth points, since not all users choose small details as interest points (Fig. 2). As σ increases, we expect to have more ground truth points, since we accept more variation on localization of the points marked by the subjects. However as σ further increases the region it defines tend to include distinct interest regions, thus close interest points marked on distinct structures start to merge. In Sect. 7, we give the average number of ground truth points on our dataset with varying n and σ.

6 Evaluation method

Previous evaluation methods for 3D salient point detectors measured the repeatability rate according to varying factors, such as deformation of the model, scale change, different modalities, noise, and topological changes [2, 13]. We perform our evaluation on a single instance of a model with respect to human generated ground truth, and use false positive and false negative errors and weighted miss error as performance measures.

6.1 False negative and false positive errors

For simplicity, let us denote the set of ground truth points $\mathcal{G}_M(n, \sigma)$ as $\mathcal{G}$, and the set of interest points detected by an algorithm on model M as $\mathcal{A}$. For an interest point g in set $\mathcal{G}$ we define a geodesic neighborhood of radius r:

$$\mathcal{C}_r(g) = \{p \in M | d(g, p) \leq r\}$$

where $d(g, p)$ corresponds to the geodesic distance between points g and p. The parameter r controls the localization error tolerance. A point g is considered to be "correctly detected" if there exists a detected point $a \in \mathcal{A}$ in $\mathcal{C}_r(g)$, and such that a is not closer to any other points in $\mathcal{G}$. Denoting the number of correctly detected points in $\mathcal{G}$ as N_C, we define the false negative error rate at localization error tolerance r as

$$FNE(r) = 1 - \frac{N_C}{N_G}, \tag{2}$$

where N_G is the number of points in $\mathcal{G}$.

The rate of false positives of an interest point detection algorithm is another measure of its relevancy to human perception of interest points. The algorithm is not supposed to find points on regions that are not of interest to humans.

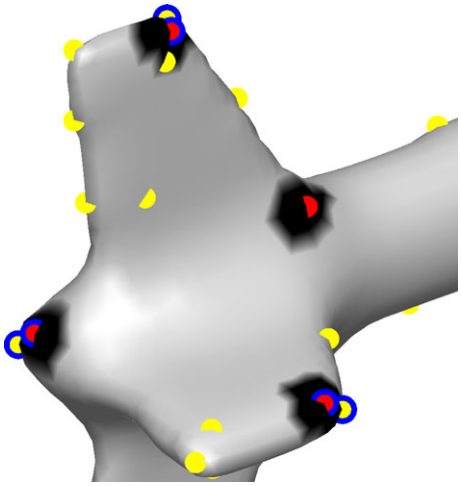


Fig. 3 The *red dots* indicate ground truth interest points. *Yellow dots* are the points marked by an interest point detection algorithm. The paired interest points are enclosed by *blue circles*. The *red dot* not enclosed by a *blue circle* is a false negative. The unenclosed *yellow dots* are false positives. The *black regions* correspond to the points with geodesic distance to a ground truth point, less than r

To calculate the false positive error rate we proceed as follows: Each correctly detected point in $g \in \mathcal{G}$ corresponds to a unique a, the closest point to g among the points in $\mathcal{A}$. All points in $\mathcal{A}$ without a correspondence in $\mathcal{G}$ are declared as false positives. Then, the number of false positives, denoted as N_F, is equal to

$$N_F = N_A - N_C, \tag{3}$$

where N_A is the number of detected interest points by the algorithm. The false positive error rate at localization error tolerance r is then

$$FPE(r) = \frac{N_F}{N_A}. \tag{4}$$

Note that our definition of false positive error rate is different than the conventional one, where the number of false positives is normalized by the number of all true negatives; i.e. the number of vertices that are not true interest points. Since this number depends on the tessellation of the mesh model, we prefer to normalize the number of false positives with the number of interest points the algorithm produces.

Figure 3 demonstrates sample false negatives and positives at the tail of an airplane model. The red dots indicate ground truth interest points. Yellow dots are the points marked by an interest point detection algorithm. Corresponding pairs of ground truth points and algorithm-detected points are enclosed by blue circles (correct detection). The red dot not enclosed by a blue circle is a false negative. The unenclosed yellow dots are false positives. The black regions correspond to the points with geodesic distance to a ground truth point, less than r.

6.2 Weighted miss error

Notice that the False Negative Error does not take into account the prominence of individual ground truth points. As long as a point has acquired enough votes from the subjects, it contributes to the detection error measure equally with all the other points. To incorporate the prominence of an interest point into the evaluation, we introduce another miss error measure, which we name as Weighted Miss Error (*WME*). Assume that within a geodesic neighborhood of radius r around the ground truth point $g_i \in \mathcal{G}$, n_i subjects have marked an interest point. Then the Weighted Miss Error is defined as

$$WME(r) = 1 - \frac{1}{\sum_{i=1}^{N_G} n_i} \sum_{i=1}^{N_G} n_i \delta_i, \tag{5}$$

where

$$\delta_i = \begin{cases} 1 & \text{if } g_i \text{ is detected by the algorithm} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The *WME* is a measure between 0 and 1, and an algorithm gets a good compensation if it manages to detect a point that is frequently voted by the human subjects. This figure will measure the ability of an algorithm to detect the most prominent, i.e. semantically significant interest points. It also makes it possible to take into account less significant interest points, which would have been discarded by hard thresholding with a high n.

7 Results

As mentioned in Sect. 4, we collected data via our web-based interface, and constructed the ground truth as described in Sect. 5. We performed the analysis on two datasets, namely Dataset A and Dataset B. Dataset A consists of 24 models which were hand-marked by 23 human subjects. Dataset B is larger with 43 models, and it contains all the models in Dataset B. However, the number of human subjects who marked all the models in this larger set is 16. The properties of the datasets are summarized in Table 1.

7.1 Subjective data

Figures 5 and 6 show some sample models from Dataset A, of man-made and natural objects, respectively. In the first column, we plot all the interest points marked by the 23 subjects; each color represents the points marked by a different subject. In the second column, ground truth interest points are plotted with red dots. The ground truth was obtained with the setting $\sigma = 0.03$ and $n = 11$, which means that we look consensus among at least half of the subjects in order for a point to qualify as ground truth. The last column shows ground truth points obtained with $\sigma = 0.03$ and

Table 1 Properties of the datasets

	# of models	# of subjects	Properties
Dataset A	24	23	3D models from SHREC'2007 Watertight Database
Dataset B	43	16	3D models as described in Sect. 4.1

$n = 2$; two subjects marking an interest point within a vicinity of $\sigma = 0.03$ are enough. However, in this case, we weight the ground truth points by their prominence. In the plots, the size of a red dot is proportional to the points' prominence. These sample models demonstrate the benefits of using the prominence of marked points instead of hard thresholding by n. We detail our observations on Figs. 5 and 6 as follows:

- Marking interest points, especially on detailed objects, is an intensive task. For example, for the *Chair* in Fig. 5a, and for the *Bird* in Fig. 6a, some subjects only marked one of two symmetrical points. Yet, some subjects fail to well-localize the interest points, for example at the corners of the *Chair*. These errors are reduced by the ground-truth building process.
- Some subjects marked points on flat or smooth regions, just to define the object. For almost all the models in Fig. 5 we observe one or two marked points on flat or smooth surfaces where there is no geometrically significant perturbation. The tendency is to mark the center of the smooth surface such as in *Table*, and *Glasses* in Fig. 5b and 5c. These marked points did not end up being ground truth points since they are far apart.
- There is a contrast between the models representing simple man-made objects in Fig. 5 and natural objects in Fig. 6. For objects with flat surfaces and sharp corners, there is little inconsistency about the location. However, when the model is detailed with many corners, some corners are not marked as often as others, as viewed in the prominence of ground truth points for *Chair* and *CAD part*. The corners defining the gross outline of the model get the most votes from the subjects.
- All subjects seem to respond to the extremities, regardless of the scale of the local perturbation. Back of the *Camel* and tips of the members of *Teddy* correspond to a larger scale, while the toes and fingers of the *Camel* and the ends of the legs of the *Octopus* are of fine scale. This necessitates an interest point algorithm to be able to operate on a wide range of scales. This result also shows that extremities of articulate parts are of interest to human viewers.
- Points marked on large scale extremities may not be reflected in the ground truth set, if n is kept high, as in *Teddy*. Notice that ground truth obtained by setting $n = 11$ consists of only three points; while the ground truth obtained with $n = 2$ reflected the interest regions. Almost all the users marked the ends of the members of *Teddy*; however, these regions are quite smooth, and there are no clear singularities. This is one of the cases where one should take into account regions of interest rather than points of interest.
- A similar situation arises along ridges. Notice that subjects have marked equally distant points along the ridges of the *Teapot* and *Cup* models. There are no clear singularities along the ridges; so no well-defined interest points. However, the same subjects marking rigorously along the ridges of the *Cup* model did not mark the ridges of the *Chair* and *Table* models. This might be due to the fact that there are not enough clear points to define the object *Cup* and subjects tend to create interest points along ridges when there are too few singularities.
- Inflections and concavities turn out to be less prominent than convexities. This situation is observable for all the models in Figs. 5 and 6, where convexities are marked with a bigger red dot in proportion to their prominence.
- Facial features (especially, eyes, nose, and ears) are very important for human and animal models. For all the models in Fig. 6b through 6f, facial features show high prominence, whether or not the face constitutes a large portion of the model. This is one of the cases, where semantic content overrides the pure geometrical content of the shapes.
- Most of the subjects marked the facial features of the *David's head* and *Girl head* in detail. Some subjects put just one or two representative points on the hair, and some others marked the details of the hair rigorously. However, the resulting ground truth point set includes very few interest points from the hair.

Figure 4a and 4b show the average number of ground truth interest points on a model with varying σ and n, for Datasets A and B, respectively. The average is taken over the number of models in the dataset. The case with $n = 1$ creates unreliable ground truth, so we exclude that case from our analysis. With increasing n, i.e. the number of subjects that vote for an interest point, we have less ground truth points. For low n, we observe an increase in the number of ground truth points, then a decrease, as σ increases. For low n and low σ, we have multiple ground truth points in close vicinity of a single interest region, especially in regions where subjects fail to well-localize a point or there is no clear singularity. With increasing σ, these multiple points tend to merge.

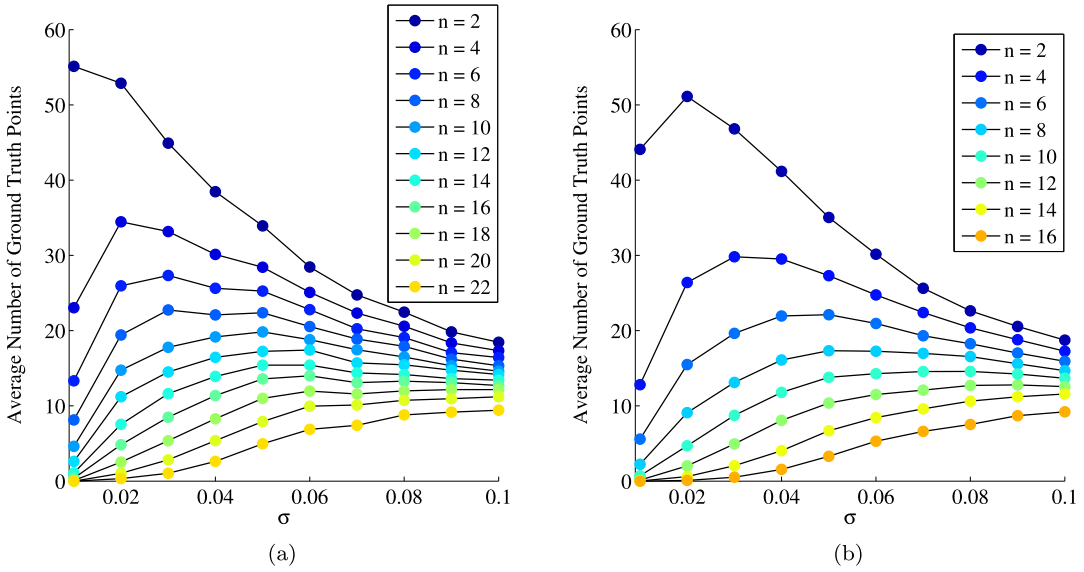


Fig. 4 Average number of ground truth interest points per model with varying σ and n. (**a**) Dataset A. (**b**) Dataset B

When n is greater than 5, the average number of points is no higher than 20 per model.

7.2 Evaluation of algorithms

Figures 7 and 8 give False Negative Error, Weighted Miss Error, and False Positive Error graphs with respect to localization error tolerance r for six interest point detection algorithms we have described in Sect. 3. Figure 7 gives the results for Dataset A and Fig. 8 for Dataset B. We provide plots for four different settings of σ and n. We observe that the error rates are consistent across Dataset A and Dataset B, for the six algorithms.

With an ideal interest point detection algorithm, we expect the errors drop very quickly with respect to r. A rapid drop in FNE means the algorithm catches the interest points with a low localization error, while a rapid drop in FPE indicates that the algorithm does not return excessive interest points. For both of the datasets, and for most of the settings of σ and n the FNE and WME drops faster with Mesh Saliency and 3D-Harris methods as compared to the other four methods (Figs. 7 and 8). With increasing localization error tolerance, Mesh Saliency method starts detecting more ground truth interest points than the 3D-Harris method; i.e. achieves lower FNE and WME. Also, the SD-corners method catches the low FNE levels of Mesh saliency method at higher values of r, which indicates a high detection rate with a low localization accuracy. We guess that the reason behind the low localization of SD-corners method is due to the unwrapping of 3D model onto a 2D plane. 3D-SIFT does not perform as well as other algorithms in terms of FNE and WME, since the coarse voxel structure does not allow good localization of interest points on the mesh model.

Mesh saliency method well localizes the interest points with a cost of a high FPE compared to the Salient points and 3D-Harris methods. The SD-corners also have a large FPE due to the large number of points it detected. The other four methods do not produce as many interest points; hence there are less "false" interest points.

HKS-based interest point detection algorithm exhibits a completely different behavior as compared to the other five algorithms. It detects very few points on the models, most of which correspond to tips of extremities of the models (see Fig. 9). HKS method gives the highest FNE among all other methods since it detects far less points than the human subjects usually mark. However, almost all points detected by HKS method correspond to a ground truth interest point, which is evident from its FPE curves. FPE is dramatically lower for HKS as compared to the other five algorithms, meaning that HKS does not return points other than those that are usually indicated as interest points by human subjects. Furthermore, the points detected by HKS method have usually high prominence (i.e. marked by many human subjects), which can be observed from Weighted Miss Error plots. WME is usually lower than FNE for HKS, which suggests that HKS makes less errors with the prominent interest points; in other words misses less of the "important" interest points.

Notice that the plots in Figs. 7 and 8 are averaged over all the models in the datasets. In order to have a closer look at the strengths and weaknesses of the algorithms, we report error rates on some individual models. Figure 9 shows five models from Dataset B. The first column shows the ground truth interest points obtained with the parameters $\sigma = 0.03$ and $n = 2$; the prominence of a ground truth point is indicated by the size of the red dot. The last six columns show

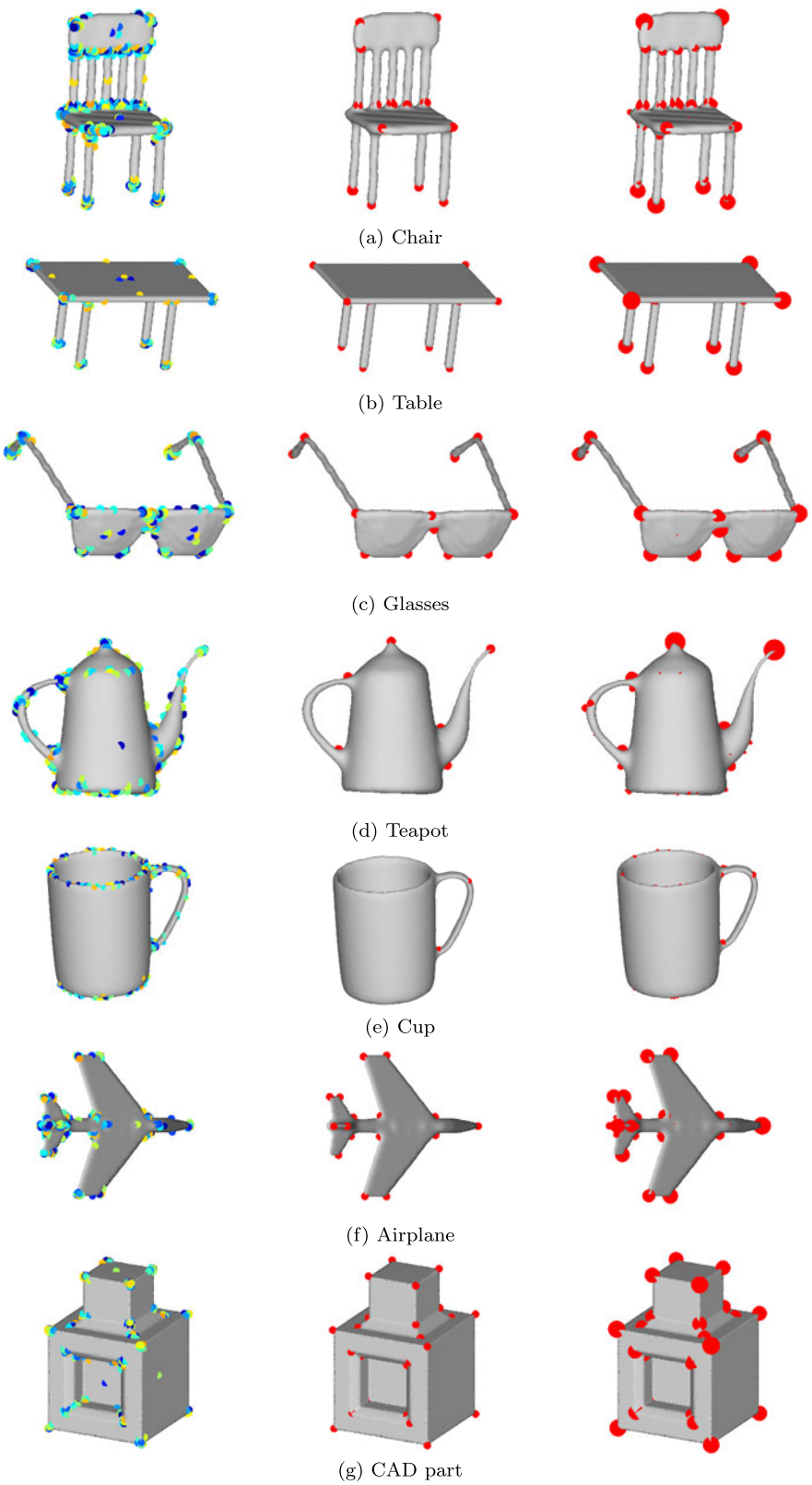


Fig. 5 *First column*: All marked points; different colors show points marked by different individuals. *Second column*: Ground truth obtained by $\sigma = 0.03$ and $n = 11$. *Third column*: Weighted ground truth obtained by $\sigma = 0.03$ and $n = 2$; the size of the point indicates the prominence of the interest point (i.e. number of subjects who marked an interest point in the vicinity)

the points from the six interest point detection algorithms. In addition to these snapshots, we also give the FNE,WME and FPE graphs of the six algorithms on these individual models (Fig. 10). The error graphs demonstrate that the performance of an interest point algorithm depends on the specific model. None of the algorithms is consistently better than the others.

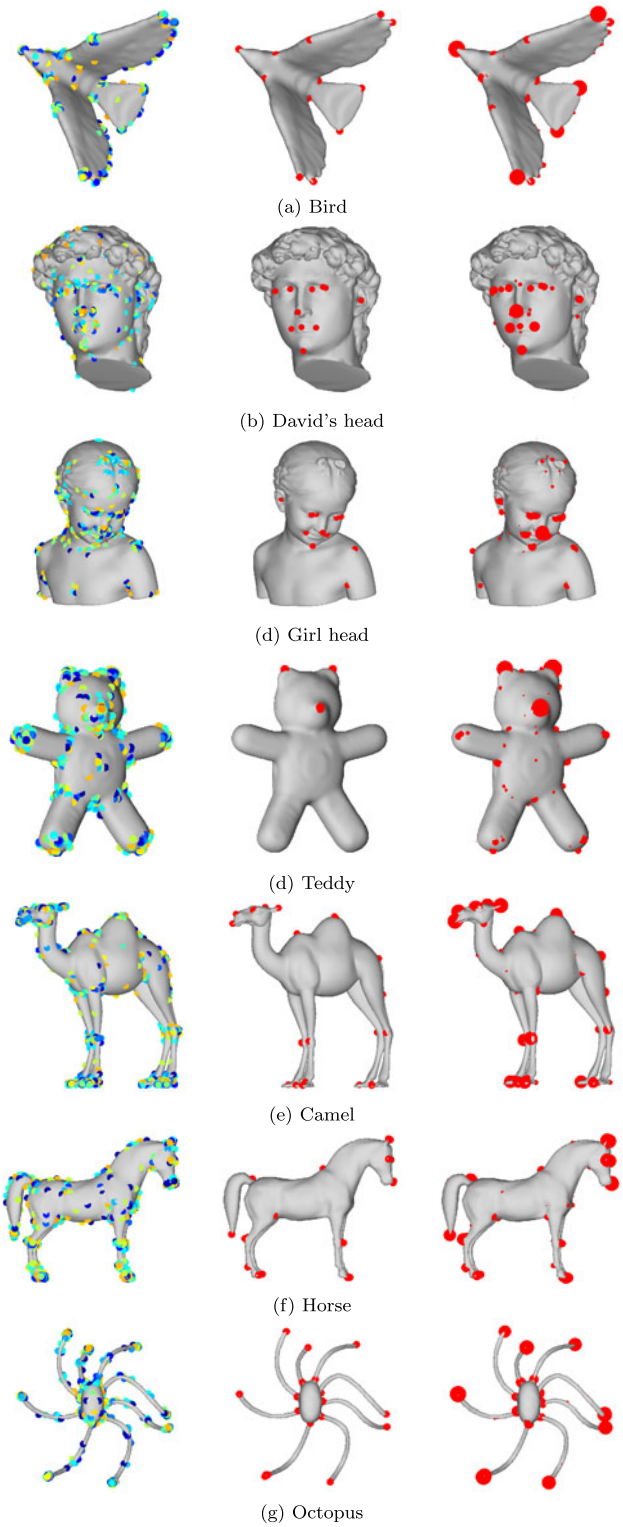


Fig. 6 *First column*: All marked points; different colors show points marked by different individuals. *Second column*: Ground truth obtained by $\sigma = 0.03$ and $n = 11$. *Third column*: Weighted ground truth obtained by $\sigma = 0.03$ and $n = 2$; the size of the point indicates the prominence of the interest point (i.e. number of subjects who marked an interest point in the vicinity)

In the following, we list our observations of the behavior of the algorithms on some individual models, in reference to Figs. 9 and 10:

- *Sensitivity to local perturbations*: Mesh saliency, Salient points, 3D-Harris, and SD-corners respond to regions where there is a local geometric perturbation; i.e. they do not locate interest points over flat or smooth regions. 3D-SIFT works on a coarse voxel structure, which is why it sometimes locates interest points onto insignificant regions. As opposed to other methods, HKS-based method works on the spectral domain, hence the interest points are detected based on the global structure of the object. Therefore, HKS can be quite insensitive to local perturbations. However, the interest points detected by the HKS method mostly correspond to high prominence ground truth points of the models shown in Fig. 9. That makes WME smaller than FNE for HKS; the difference is especially pronounced for *Armadillo* and *Chair* models (Figs. 10a, and 10d).
- *Number of detected interest points*: The algorithms, except the HKS method, tend to mark more interest points then the human users, especially the Mesh saliency and SD-corners methods. The two methods pick up interest points from almost all singularities, resulting in high FPE. 3D-Harris and Salient points methods provide less interest points; due to their parameter settings in their schemes that eliminate candidate points. The parameters can be adjusted to catch more singularities; however, that would also increase number of false positives.
- *Articulate parts and HKS*: HKS returns few interest points, and these points are usually in the vicinity of a ground truth point. Notice that FPE reaches zero for HKS method for four of the models, before the localization tolerance is 0.1. That means, HKS does not return "false" interest points for these models. As stated in Sect. 7.1, extremities of articulate parts are of significant interest to humans, and HKS seems to well detect the tips of articulate parts (see *Armadillo*, *Camel*, and *Chair* in Fig. 10).
- *WME for individual models*: The impact of the WME in comparison to FNE is more visible when we examine them on individual models. Whenever WME is lower than FNE, we conclude that the algorithm tends to miss less of the more prominent points (i.e. points marked by many human subjects). For example, for the *Armadillo* model, FNE and WME of Mesh Saliency and SD-corners are more or less the same. However, for the Salient Points, 3D-Harris and HKS methods, WME is significantly lower than FNE. For Salient Points and 3D-Harris method, WME even drops close to that of SD-corners method, which means Salient Points or 3D-Harris method can reach the same WME with a lower FPE as compared to SD-corners.

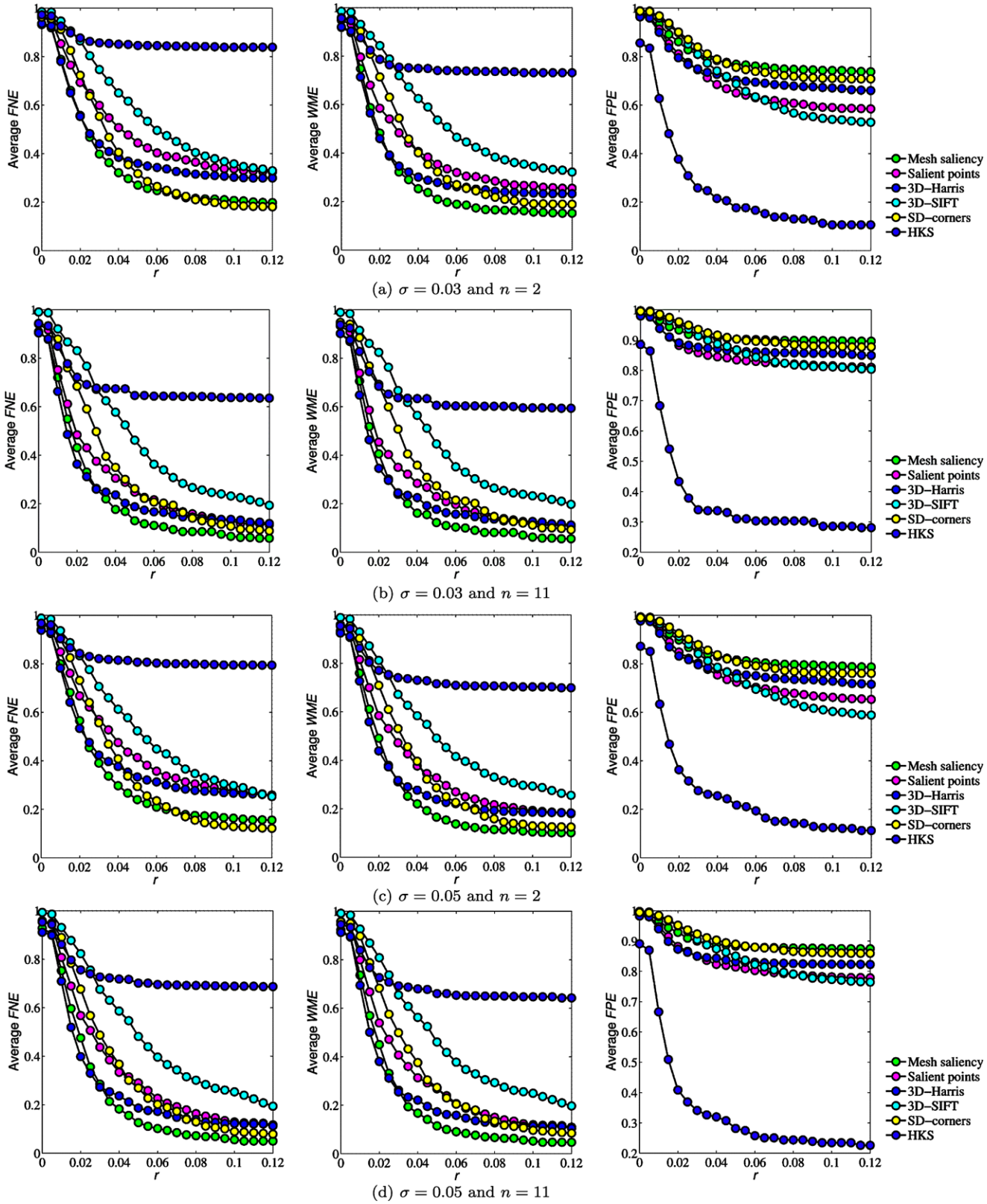


Fig. 7 Performance on Dataset A (24 Models, 23 Subjects). False Negative Error (*first column*), Weighted Miss Error (*second column*), and False Positive Error (*third column*). The settings for obtaining the ground truth are indicated under the plots

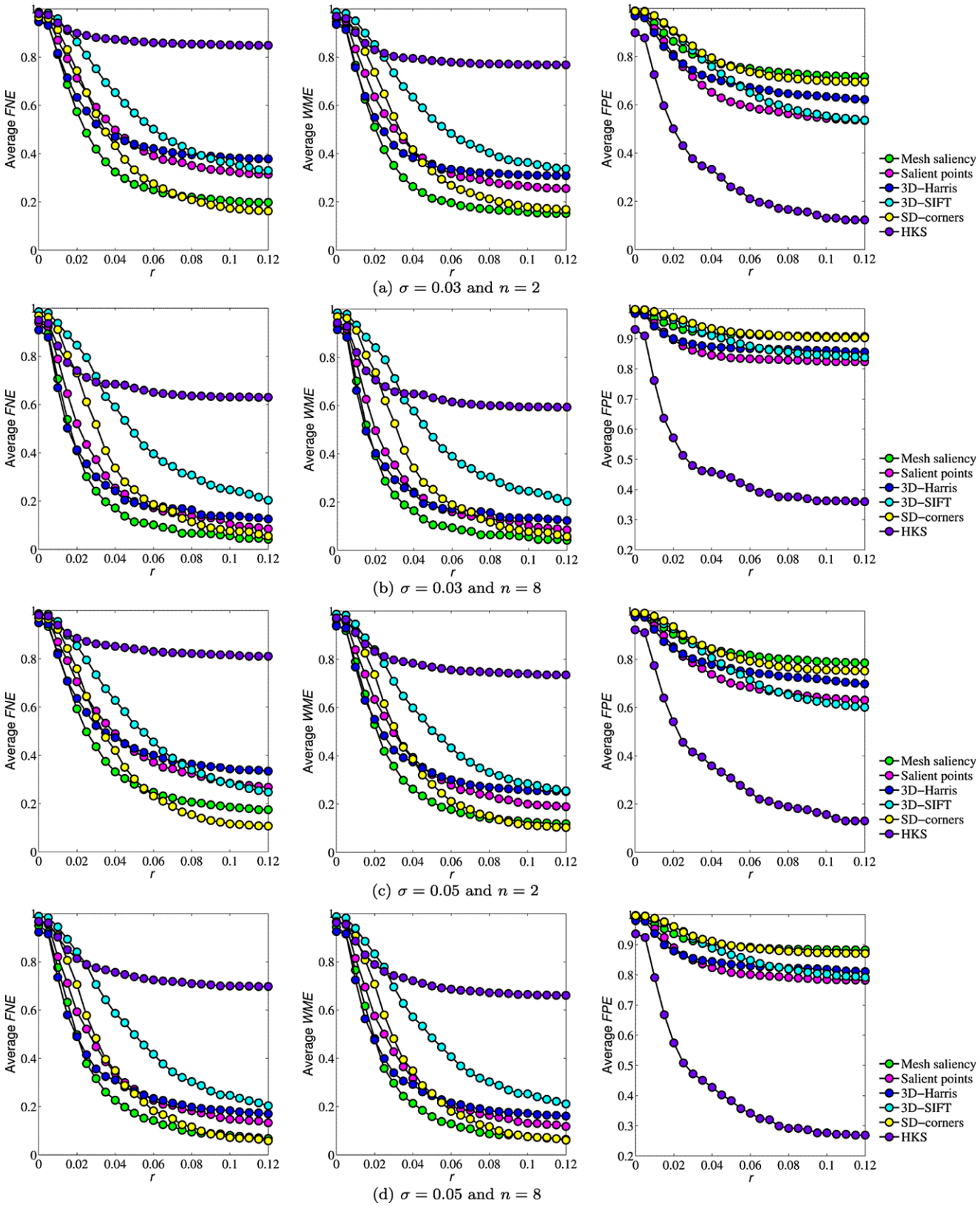


Fig. 8 Performance on Dataset B (43 Models, 16 Subjects). False Negative Error (*first column*), Weighted Miss Error (*second column*), and False Positive Error (*third column*). The settings for obtaining the ground truth are indicated under the plots

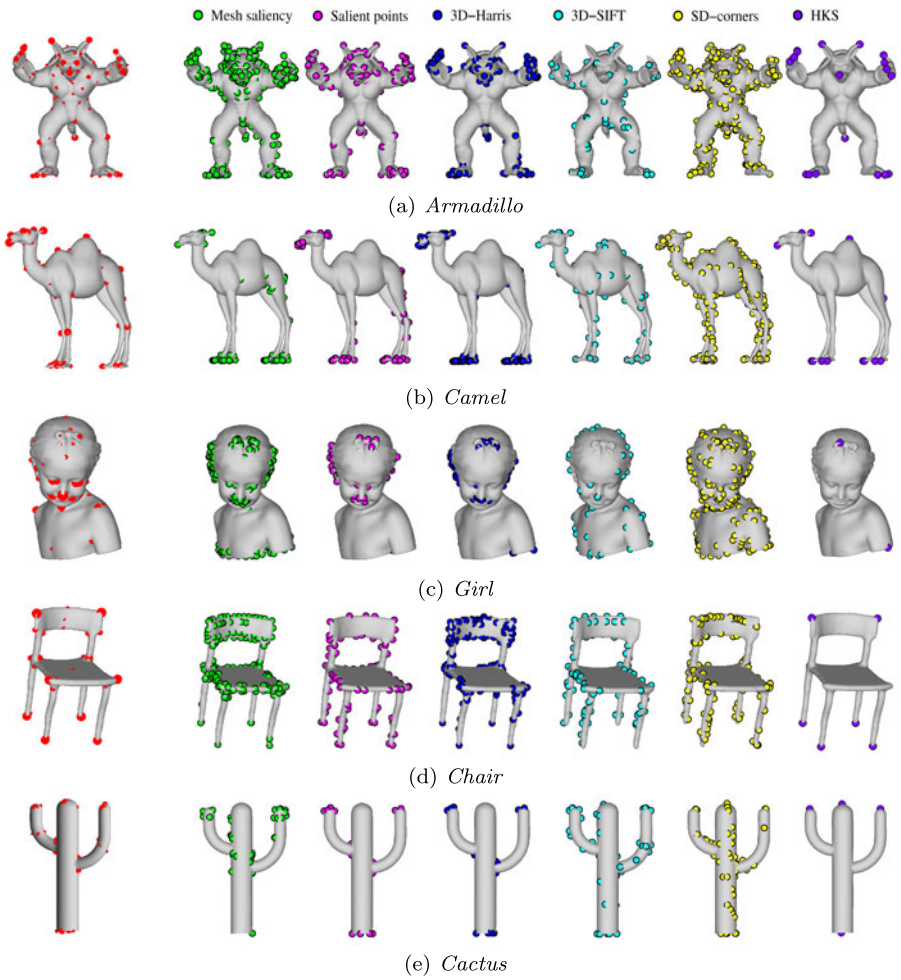


Fig. 9 Ground truth obtained by setting $\sigma = 0.03$ and $n = 2$, with Dataset B (16 subjects); the points are weighted with respect to prominence (*first column*), interest points detected by the algorithms: Mesh saliency (*second column*), Salient points (*third column*), 3D-Harris (*fourth column*), 3D-SIFT (*fifth column*), SD-corners (*sixth column*), and HKS (*seventh column*)

– *Interest points on large scales and 3D-SIFT*: The 3D-SIFT algorithm detects interest points on large scales (shoulder of the *Girl*, back of the *Camel*, tip of the *Cactus*), due to its coarse voxelization strategy. However, it did not well-localize the finer interest points as the other algorithms did (ear tips and fingers of the *Armadillo* model).
– *Interest points of different nature*: For the *Armadillo* model, Mesh Saliency performs better than the other algorithms on detecting the ground truth points, and is followed by SD-corners. Notice that *Armadillo* model has interest points of different nature: Facial features, extremities at fine and large scales, and interest points that do not correspond to extremities, such as points at chest and knees. Except with 3D-SIFT, all the four algorithms found interest points on the nose and finger tips of the *Armadillo* which are rather strong protrusions. However, Salient points, 3D-Harris, and HKS methods failed to detect interest points in the chest region of *Armadillo*, and 3D-Harris discarded the interest points in one of the legs.
– *Saliency across and within scales*: While other algorithms select points that remain salient across scales, SD-corners method gathers interest points from all the scales. The result is a large number of interest points representing both small (facial features of the *Girl*) and large details (tip of the *Cactus*). This property makes the SD-corners method to achieve the least FNE and WME for the *Camel*, *Girl*, and *Cactus* models, where both fine details and extremities of gross structures were marked as interest points by human subjects. However, returning a large number of interest points results in high FPE.
– *Edges versus corners*: Most of the algorithms find interest points along the edges of the *Chair* model due to the high curvature in one direction, while there are no ground truth points along those edges. The FNE and WME drop very fast with Mesh saliency and 3D-Harris methods for the *Chair*. They are both sensitive to corners and ridges, producing multiple interest points around the corners of the chair. The low FNE is paid with a high False Positive rate for Mesh saliency and 3D-Harris methods.
– *Facial features*: For the *Girl* model, the interest points generated by the algorithms populate on the hair, at the expense of missing some of the facial features. The result is high error rate in terms of both missing points and false

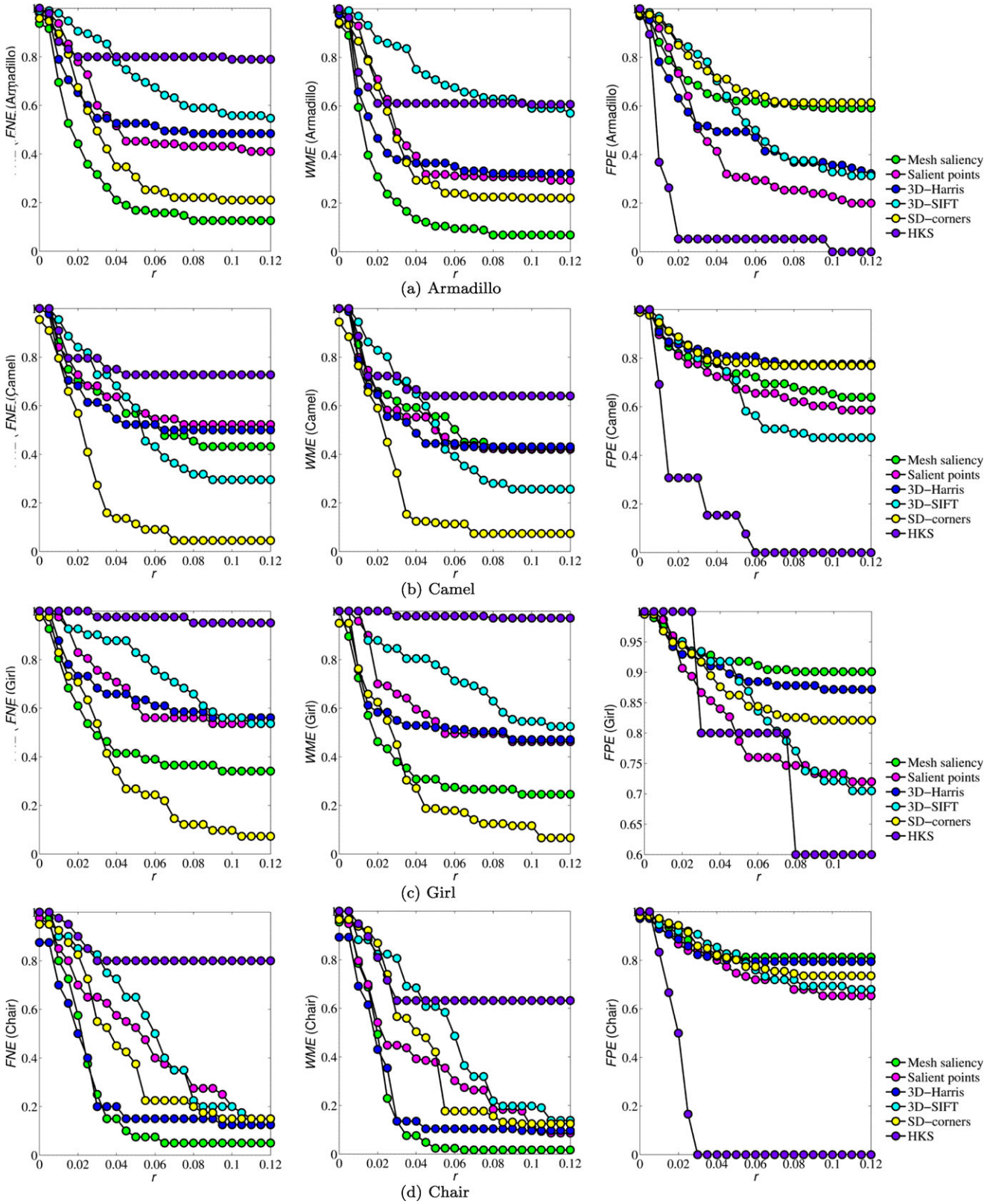


Fig. 10 False Negative Error, Weighted Miss Error, and False Positive Error graphs of all algorithms for individual models as rendered in Fig. 9. Ground truth obtained by setting $\sigma = 0.03$ and $n = 2$, with Dataset B (16 subjects)

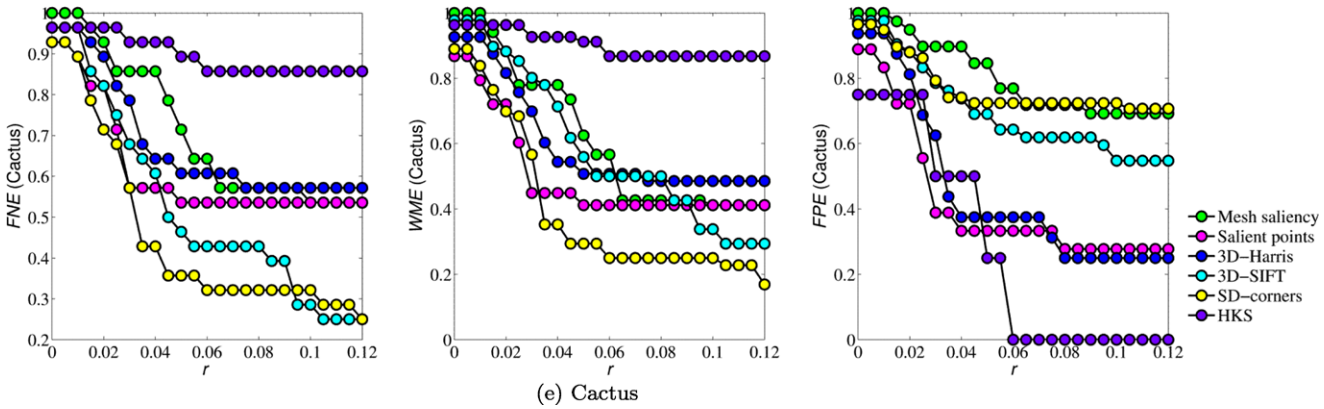


Fig. 10 (*Continued*)

positives, especially for the Salient points and 3D-Harris methods. HKS gives very high false negative error for the *Girl* model; it could not find any of the facial landmarks of the model.

8 Conclusion

We designed experiments to compare automatic salient point detection algorithms with humans' choices of interest points. We developed a web-based application where human subjects marked interest points on 3D models. We compared six different interest point detection techniques based on False Negative and False Positive errors, and Weighted Miss Error, employing the human-provided data as ground truth.

An important issue is that, in this work, the algorithms are evaluated with their fixed parameters settings, as provided by the authors of the original work. There are many parameters of these algorithms that can be optimized using the evaluation measures proposed in this work. Especially the parameters that are responsible to eliminate candidate interest points would tune the trade-off between False Negative Error (or Weighted Miss Error) and False Positive Errors, depending on the demands of the application. Furthermore, shape classification schemes can be incorporated into the interest points detection system; for instance, the system can adjust the parameters depending on whether a model is articulated, man-made or natural, complex or simple, whether it contains a face or other semantically important structures.

Detection of interest points of generic objects is one of those computer vision problems where ground-truth is fuzzy. The human subjects who contributed to the experiments were most of the time indecisive about whether a detailed point should be considered as an interest point or not. Some "perfectionist" subjects marked every single detail on complex models, while others just put a few points here and there. Their attention also degraded as they got tired of the task. It is necessary to conduct well-organized psychophysical experiments to generalize humans' reactions to interest regions; here an eye-tracking system would have been more appropriate [10]. However, our method to merge these marked points into a ground truth gives reasonable sets of points, as far as our concern is to evaluate the interest point detection algorithms on a fairly diverse set of models common to 3D shape applications. Furthermore, our scheme can be easily adapted to an experimental setup with an eye-tracking system. We believe that this study provides a starting point for selecting 3D interest point detection algorithms which are in accordance with human perception.

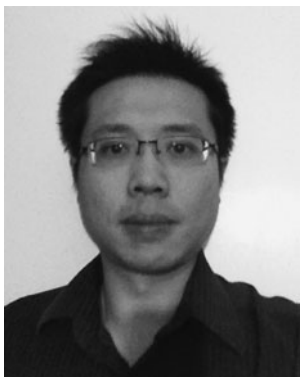
Acknowledgements We would like to thank Daniela Giorgi and AIM@SHAPE for the models from the Watertight Track of SHREC 2007, and Stanford Computer Graphics Laboratory for the models from The Stanford 3D Scanning Repository. We would like to thank Ivan Sipiran, Benjamin Bustos, Umberto Castellani, Ko Nishino, and Prabin Bariya for sharing their interest point detection codes.

References

1. Boyer, E., Bronstein, A.M., Bronstein, M.M., Bustos, B., Darom, T., Horaud, R., Hotz, I., Keller, Y., Keustermans, J., Kovnatsky, A., Litman, R., Reininghaus, J., Sipiran, I., Smeets, D., Suetens, P., Vandermeulen, D., Zaharescu, A., Zobel, V.: SHREC 2011: robust feature detection and description benchmark. In: 3DOR, pp. 71–78 (2011)
2. Bronstein, A., Bronstein, M., Bustos, B., Castellani, U., Crisani, M., Falcidieno, B., Guibas, L., Kokkinos, I., Murino, V., Sipiran, I., Ovsjanikovy, M., Patanè, G., Spagnuolo, M., Sun, J.: SHREC 2010: robust feature detection and description benchmark. In: Eurographics Workshop on 3D Object Retrieval (3DOR'10), pp. 79–86 (2010)
3. Castellani, U., Cristani, M., Fantoni, S., Murino, V.: Sparse points matching by combining 3D mesh saliency with statistical descriptors. Comput. Graph. Forum **27**(2), 643–652 (2008)
4. Chung, M., Taylor, J.: Diffusion smoothing on brain surface via finite element method. In: ISBI, pp. 432–435 (2004)
5. Dutagaci, H., Cheung, C.P., Godil, A.: Evaluation of 3D interest point detection techniques. In: 3DOR, pp. 57–64 (2011)

6. Gelfand, N., Mitra, N.J., Guibas, L.J., Pottmann, H.: Robust global registration. In: Symposium on Geometry Processing, pp. 197–206 (2005)
7. Godil, A., Wagan, A.I.: Salient local 3D features for 3D shape retrieval. In: 3D Image Processing (3DIP) and Applications II. SPIE, Bellingham (2011)
8. Harris, C., Stephens, M.: A combined corner and edge detector. In: 4th Alvey Vision Conference, pp. 147–151 (1988)
9. Hu, J., Hua, J.: Salient spectral geometric features for shape matching and retrieval. Vis. Comput. **25**(5–7), 667–675 (2009)
10. Kim, Y., Varshney, A., Jacobs, D.W., Guimbretière, F.: Mesh saliency and human eye fixations. ACM Trans. Appl. Percept. **7**(2), 12 (2010)
11. Lee, C.H., Varshney, A., Jacobs, D.W.: Mesh saliency. In: ACM SIGGRAPH 2005, pp. 659–666 (2005)
12. Litman, R., Bronstein, A.M., Bronstein, M.M.: Diffusion-geometric maximally stable component detection in deformable shapes. Comput. Graph. **35**(3), 549–560 (2011)
13. Lloyd, B.A., Szekely, G., Kikinis, R., Warfield, S.K.: Comparison of salient point detection methods for 3D medical images. The Insight Journal—2005 MICCAI Open-Source Workshop (2005)
14. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
15. Maes, C., Fabry, T., Keustermans, J., Smeets, D., Suetens, P., Vandermeulen, D.: Feature detection on 3d face surfaces for pose normalisation and recognition. In: Biometrics Theory Applications and Systems BTAS 2010 Fourth IEEE International Conference on, pp. 1–6 (2010)
16. Mian, A., Bennamoun, M., Owens, R.: On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes. Int. J. Comput. Vis. **89**(2–3), 348–361 (2010)
17. Novatnack, J., Nishino, K.: Scale-dependent 3D geometric features. In: ICCV, pp. 1–8 (2007)
18. Sebe, N., Lew, M.S.: Comparing salient point detectors. Pattern Recognit. Lett. **24**(1–3), 89–96 (2003)
19. Sipiran, I., Bustos, B.: A robust 3D interest points detector based on Harris operator. In: Eurographics 2010 Workshop on 3D Object Retrieval (3DOR'10), pp. 7–14 (2010)
20. Sipiran, I., Bustos, B.: Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes. Vis. Comput. **27**, 963–976 (2011)
21. Smith, S.M., Brady, J.M.: Susan—a new approach to low level image processing. Int. J. Comput. Vis. **23**, 45–78 (1995)
22. Sun, J., Ovsjanikov, M., Guibas, L.: A concise and provably informative multi-scale signature based on heat diffusion. In: Eurographics Symposium on Geometry Processing (SGP), pp. 1383–1392 (2009)
23. Walter, N., Aubreton, O., Laligant, O.: Salient point characterization for low resolution meshes. In: ICIP, pp. 1512–1515 (2008)
24. Wessel, R., Novotni, M., Klein, R.: Correspondences between salient points on 3D shapes. In: Vision, Modeling, and Visualization (VMV'06), pp. 365–372 (2006)
25. Zaharescu, A., Boyer, E., Varanasi, K., Horaud, R.P.: Surface feature detection and description with applications to mesh matching. In: CVPR (2009)
26. Zou, G., Hua, J., Dong, M., Qin, H.: Surface matching with salient keypoints in geodesic scale space. Comput. Animat. Virtual Worlds **19**(3–4), 399–410 (2008)
27. http://control.nist.gov/sharp/view/
28. http://www.itl.nist.gov/iad/vug/sharp/benchmark/3DInterestPoint/
29. http://www.graphics.stanford.edu/data/3Dscanrep/
30. http://watertight.ge.imati.cnr.it
31. http://profs.sci.univr.it/~castella/research.html/

Helin Dutagaci received the B.Sc., M.Sc., and Ph.D. degrees from the Bogazici University, Istanbul, Turkey, in 1999, 2002, and 2009, respectively. She worked as a Guest Researcher at National Institute of Standards and Technology between 2008 and 2011. She is currently an assistant professor at the Department of Electrical-Electronics Engineering, Eskisehir Osmangazi University, Turkey. Her major field of study is signal processing. Her research interest includes computer vision, pattern recognition, 3-D object recognition, and biometrics.



Chun Pan Cheung received his Master Degree in Software Engineering from University Maryland, University College in 2007. He worked at National Institute of Standards and Technology from 2007 to 2011 on different projects, including 3D shape searching, fingerprint and iris biometrics. His expertise is in software development. Currently he is working in the private sector as a consultant.

Afzal Godil is a project leader in the Information Technology Laboratory at National Institute of Standards and Technology (NIST) where he has been for over 15 years. Prior to that he has worked at the NASA Langley and Lewis Research Centers as a contractor. His main focus in research and development is in the area of 3D Shape Analysis and Retrieval, graphics/visualization, digital human modeling, computational methods, and pattern recognition. He was also a principle technical staff member in the initiation and development of 3D Face Recognition and Web graphics.