Eurographics Workshop on 3D Object Retrieval (2011)
H. Laga, T. Schreck, A. Ferreira, A. Godil, I. Pratikakis, and R. Veltkamp (Editors)

SHREC 2011: robust feature detection and description benchmark

E. Boyer[1], A. M. Bronstein[†2], M. M. Bronstein[†3], B. Bustos[4], T. Darom[5], R. Horaud[1], I. Hotz[6], Y. Keller[5], J. Keustermans[7], A. Kovnatsky[†8], R. Litman[†2], J. Reininghaus[6], I. Sipiran[4], D. Smeets[7], P. Suetens[7], D. Vandermeulen[7], A. Zaharescu[†9], V. Zobel[6]

[1]INRIA Grenoble Rhône-Alpes, France
[2]Department of Electrical Engineering, Tel Aviv University, Israel
[3]Institute of Computational Science, Faculty of Informatics, Università della Svizzera Italiana, Lugano, Switzerland
[4]Department of Computer Science, University of Chile
[5]School of Engineering, Bar-Ilan University, Ramat-Gan, Israel
[6]Zuse Institut Berlin, Germany
[7]Department of Electrical Engineering, K.U. Leuven, Belgium
[8]Department of Mathematics, Technion – Israel Institute of Technology, Haifa, Israel
[9]Aimetis Corp., Waterloo, Canada

Abstract

Feature-based approaches have recently become very popular in computer vision and image analysis applications, and are becoming a promising direction in shape retrieval. SHREC'11 robust feature detection and description benchmark simulates the feature detection and description stages of feature-based shape retrieval algorithms. The benchmark tests the performance of shape feature detectors and descriptors under a wide variety of transformations. The benchmark allows evaluating how algorithms cope with certain classes of transformations and strength of the transformations that can be dealt with. The present paper is a report of the SHREC'11 robust feature detection and description benchmark results.

Categories and Subject Descriptors (according to ACM CCS): H.3.2 [Information storage and retrieval]: Information Search and Retrieval—Retrieval models I.2.10 [Artificial intelligence]: Vision and Scene Understanding—Shape

1. Introduction

Feature-based approaches have recently become very popular in computer vision and image analysis applications, notably due to the works of Lowe [Low04], Sivic and Zisserman [SZ03], and Mikolajczyk and Schmid [MS05]. In these approaches, an image is described as a collection of local features ("visual words") from a given vocabulary, resulting in a representation referred to as a *bag of features*. The bag of features paradigm relies heavily on the choice of the local feature descriptor that is used to create the visual words. A common evaluation strategy of image feature detection and description algorithms is the stability of the detected features and their invariance to different transformations applied to an image. In shape analysis, feature-based approaches have been introduced more recently and are gaining popularity in shape retrieval applications.

SHREC'11 invariant feature detection and description benchmark simulates the feature detection and description stages of feature-based shape retrieval algorithms. The benchmark tests the performance of shape feature detectors and descriptors under a wide variety of different transformations. The benchmark allows evaluating how algorithms cope with certain classes of transformations and what is the strength of the transformations that can be dealt with.

† Organizer of the SHREC track. All organizers and participants are listed in alphabetical order. For any information about the benchmark, contact michael.bronstein@usi.ch. Authors listed alphabetically. Full results will be published in a technical report.

© The Eurographics Association 2011.

DOI: 10.2312/3DOR/3DOR11/071-078

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

2. Data

The dataset used in this benchmark was from the TOSCA shapes [BBK08], available in the public domain. The shapes were represented as triangular meshes with approximately 10,000–50,000 vertices. The dataset includes ones shape class (human) with simulated transformations. Compared to the SHREC 2010 benchmark, there are additional transformation classes and the transformations themselves are more challenging. For each null shape, transformations were split into 11 classes shown in Figure 1: In each class, the transformation appeared in five different versions numbered 1–5 (the higher the number, the stronger the transformation). The total number of transformations was 55. The dataset is available at http://tosca.cs.technion.ac.il/book/shrec_feat.html.

3. Evaluation methodology

The evaluation was performed separately for feature detection and feature description algorithms. Feature detectors were further divided into point and region; feature descriptors were divided into point, region, and dense. The participants were asked to provide, for each shape Y in the dataset, (i) a set of detected feature points $\mathcal{F}(Y) = \{y_k \in Y\}_k$ or regions $\mathcal{F}(Y) = \{Y_l \subset Y\}_l$; (ii) optionally, for each detected point y_k, a descriptor vector $\{\mathbf{f}(y_k)\}_{k=1}^{|\mathcal{F}(Y)|}$; or alternatively, for each detected region Y_l, a descriptor vector $\{\mathbf{f}(Y_l)\}_{l=1}^{|\mathcal{F}(Y)|}$. For dense descriptors, participants provided $\{\mathbf{f}(y_k)\}_{k=1}^{|Y|}$. The performance was measured by comparing features and feature descriptors computed for transformed shapes and the corresponding null shapes.

Feature detection. The quality of the feature detection was measured using the *repeatability* criterion. Assuming for each transformed shape Y in the dataset the groundtruth dense correspondence to the null shape X to be given in the form of pairs of points $\mathcal{C}_0(X,Y) = \{(x'_k, y_k)\}_{k=1}^{|Y|}$, a feature point $y_k \in \mathcal{F}(Y)$ is said to be *repeatable* if a geodesic ball of radius ρ around the corresponding point $x'_k : (x'_k, y_k) \in \mathcal{C}_0(X,Y)$ contains a detected feature point $x_j \in \mathcal{F}(X)$.[†] Repeatable features are

$$\mathcal{F}_\rho(Y) = \{y_k \in \mathcal{F}(Y) : \mathcal{F}(X) \cap B_\rho(x'_k) \neq \emptyset, (x'_k, y_k) \in \mathcal{C}_0(X,Y)\},$$

where $B_\rho(x'_k) = \{x \in X : d_X(x, x'_k) \leq \rho\}$ and d_X denotes the geodesic distance function in X.

Similarly, for region detectors, a region $Y_l \in \mathcal{F}(Y)$ is repeatable if the corresponding region $X'_l \subset X$ has overlap larger than ρ,

$$\mathcal{F}_\rho(Y) = \{Y_l \in \mathcal{F}(Y) : |X'_l \cap X_l| / |X'_l \cup X_l| \geq \rho\}.$$

[†] Features without groundtruth correspondence (e.g. in regions in the null shape corresponding to holes in the transformed shape) are ignored.

The *repeatability* of a feature detector is defined as the percentage $|\mathcal{F}_\rho(Y)|/|\mathcal{F}(Y)|$ of features that are repeatable, the definition being dependent of whether a point or region descriptor is used.

Feature description. Let $\{\mathbf{f}_k\}_{k=1}^{|\mathcal{F}(Y)|}$, $\{\mathbf{g}_j)\}_{j=1}^{|\mathcal{F}(X)|}$ denote descriptors computed on feature points $\mathcal{F}(X)$ and $\mathcal{F}(Y)$, respectively. For point descriptors, we consider as the point corresponding to y_k the closest point $x_j \in \mathcal{F}(X)$ to x'_k, where $(x'_k, y_k) \in \mathcal{C}_0(X,Y)$, such that $r_{kj} = d_X(x_j, x'_k) < \rho$ for some ρ.

Descriptor quality was evaluated using the normalized L_2 distance between descriptors at corresponding points,

$$d_{kj} = \frac{\|\mathbf{f}_k - \mathbf{g}_j\|_2}{\frac{1}{|\mathcal{F}(X)|^2 - |\mathcal{F}(X)|} \sum_{k,j \neq k} \|\mathbf{f}_k - \mathbf{g}_j\|_2}.$$

In addition, an evaluation using the ROC was performed as follows. The corresponding feature points x_k, y_j are considered *true positives* if $d_{kj} \leq \tau$, for some threshold τ. The *true positive rate* is defined as $TPR = |\{d_{kj} \leq \tau\}|/|\{r_{kj} \leq \rho\}|$; the *false positive rate* is defined as $FPR = |\{d_{kj} \leq \tau\}|/|\{r_{kj} > \rho\}|$. By varying the threshold τ, a set of pairs (FPR, TPR) referred to as the *receiver operation characteristic* (ROC) curve is obtained. For a fixed FPR, the higher the TPR, the better.

For a dense descriptor, the quality is measured as the average normalized L_2 distance between the descriptor vectors in corresponding points,

$$\frac{1}{|\mathcal{F}(X)|} \sum_{k=1}^{|\mathcal{F}(X)|} d_{kj}.$$

4. Feature detection methods

Four point feature detectors (Harris 3D, Mesh-DoG, Mesh SIFT, Mesh-Scale DoG) and one region feature detector (Shape MSER) were evaluated.

Harris 3D (Sipiran and Bustos [SB10]). The algorithm proposes an extension for meshes of the Harris corner detection method [HS88]. The algorithm suggests to determine a neighborhood (rings or adaptive) around a vertex. Next, this neighborhood is used to fit a quadratic patch which is considered as an image. After applying a gaussian smoothing, derivatives are calculated which are used to calculate the Harris response for each vertex. In this benchmark, three different configurations were used: adaptive neighborhoods with $\delta = 0.01$, 1-ring neighborhoods , and 2-ring neighborhoods. For details, see [SB10].

Mesh-DoG (Zaharescu *et al.* [ZBVH09]). The method considers the general setting of 2-D manifolds $\mathcal{M}$ embedded in $\mathbb{R}^3$ endowed in with a scalar function $f : \mathcal{M} \to \mathbb{R}$, such as colour or curvature. This represents a generalization of 2-D images, that can be viewed as a uniformly sampled square grid with vertices of valence 4. Operators, such

© The Eurographics Association 2011.

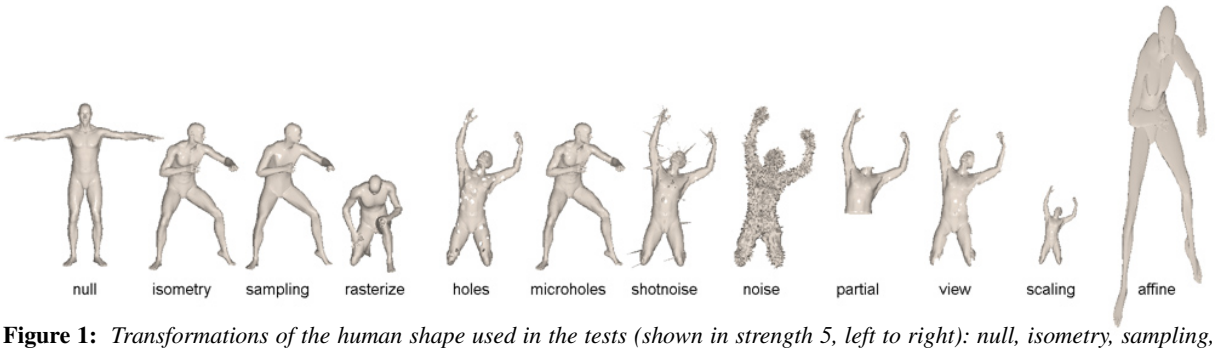

Figure 1: *Transformations of the human shape used in the tests (shown in strength 5, left to right): null, isometry, sampling, rasterize, holes, micro holes, shot noise, noise, partial, view, scaling, affine.*

as the gradient and the convolution are defined in this context. A scale-space representation of the scalar function f is build using iterative convolutions with a Guassian kernel. Feature detection consists of two steps. Firstly, the extrema of the function's Laplacian (approximated by taking the difference between adjacent scales - Difference of Gaussian) are found across scales, followed by non-maximum suppression using a 1-ring neighbourhood both spatially and across adjacent scales. Secondly, the detected extrema are thresholded (400 points). Mean and Gaussian curvature computed using [MDSB02] were the scalar functions used for current tests. For exact details and settings, see [ZBVH09].

Mesh SIFT (Smeets *et al.* [MFK*10]). The Mesh SIFT detector detects scale space extrema as local feature locations. First, a scale space is constructed containing smoothed versions of the input mesh, which are obtained by subsequent convolutions of the mesh with a binomial filter. Next, for the detection of salient points in the scale space, the mean curvature H (Mesh SIFT-H) and the principal coordinates in curvature space KK (Mesh SIFT-KK), which are minimal and maximal curvature, are computed for each vertex and at each scale in the scale space (H_i and KK_i). Note that the mesh is smoothed and not the function on the mesh (H or KK). Scale space extrema in scale spaces of differences between subsequent scales ($dH_i = H_{i+1} - H_i$ for Mesh SIFT-H and $dKK_i = KK_{i+1} - KK_i$ for Mesh SIFT-KK) are finally selected as local feature locations.

Mesh-Scale DoG (Darom and Keller [DK11]) We follow the work of Zaharescu *et al.* [ZBVH09] that presented a Difference of Gaussians based feature points detector for mesh objects. We propose to define a Gaussian filter on the mesh geometry, and compute a set of filtered meshes. Consecutive octaves are subtracted to compute the DoG function, and define the local maxima (both in location and scale) as our feature points at that point and scale. In order to make the detected features scale invariant, we suggest to set the support for each feature point to the width of the filter at that scale. For details, see [DK11].

Shape MSER (Litman *et al.* [LBB10]). The algorithm finds maximally stable components in 3D shapes, similarly to the popular MSER method for feature analysis in images [MCUP04]. The shape is represented as a component tree based on vertex- or edge-wise weighting function (VW and EW, respectively). In this benchmark, three different weights were used: edge weighting by inverse of commute time kernel (EW 1/CT) and inverse heat kernel (EW 1/HKS), and vertex weighting by heat kernel diagonal (VW HKS). For details, see [LBB10].

5. Feature description methods

Three sparse (Mesh HoG, Scale-invariant Spin Image, and Local Depth SIFT) and one dense (GHKS) feature descriptors were evaluated.

Mesh-HoG (Zaharescu *et al.* [ZBVH09]). For a given interest point, the descriptor is computed using a geodesic support region, proportional to 3% of the total surface area. For each vertex in the neighbourhood, the 3-D gradient information is computed using f at the detected scale. As a first step, a local coordinate system is chosen, in order to make the descriptor rotation invariant. Then, a histogram of gradient is computed, both spatially, at a coarse level, in order to maintain a certain high-level spatial ordering, and using orientations, at a finer level. Since the gradient vectors are 3 dimensional, the histograms are computed in 3D. The histograms are concatenated and normalized. A 96 dimensional descriptor is obtained. The gradient of the participating neighbouring vertices is computed at the scale of the detected interest point. For exact details and settings, see [ZBVH09].

Scale Invariant Spin Image (Darom and Keller [DK11]) The Spin Image local descriptor was presented by Johnson and Hebert [JH99], and has gained popularity due to its robustness and simplicity. Utilizing the local scale estimated by the Mesh-Scale DoG detector, we propose to derive a Scale Invariant Spin Image mesh descriptor, where we compute the Spin Image descriptor over the local scale estimated at the interest point. This improves feature point matching, in particular when the meshes are related significant partial matching. For details, see [DK11].

© The Eurographics Association 2011.

Local Depth SIFT (Darom and Keller [DK11]) The SIFT algorithm, presented by D. Lowe [Low04] is a state-of-the-art approach to computing scale and rotation invariant local features in images. The SIFT descriptor is based on computing a local radial-angular histogram of the pixel value derivatives. Inspired by Lowe's seminal work, we propose to compute a new local feature for 3D meshes we denote *Local Depth SIFT* (LD-SIFT). Given an interest point we estimate its tangent plane, and compute the distance from each point on the surface to that plane to create a depth map, and set the viewport size to match the feature scale, as detected by the Mesh-Scale DoG detector. This makes our construction scale invariant. We compute the PCA of the the points surrounding the interest point, and use their dominant direction as the *local dominant angle*, and rotate the depth map to a canonical angle based on the dominant angle. This makes the LD-SIFT rotation invariant. We compute a SIFT feature descriptor on the resulting depth map to create the Local Depth SIFT feature descriptor. For details, see [DK11].

Generalized HKS (Zobel *et al.* [ZRHar]). The Generalized HKS is a generalization of the HKS [SOG09] to 1-forms (where a 1-form can be regarded as vector field). It is derived from the heat kernel for 1-forms in a similar way as the HKS is derived from the heat kernel for functions. This yields a symmetric tensor field of second order with a time parameter t. For easier comparability we consider scalar tensor invariants. For details see [ZRHar] or [Zob10].

6. Results

Figures 2–3 show the repeatability of point descriptors as function of geodesic distance varying from 0 to 5. Figure 4 shows the repeatability of region feature detectors as function of overlap varying from 0 to 1. Higher values for a given distance/overlap indicate better performance.

Figure 5 shows the ROC curves of different point feature descriptors, using a fixed value of $\rho = 5$. Higher values of the vertical axis at a fixed point on the horizontal axis are indication of better performance.

Table 1 shows the performance of the GHKS dense feature description algorithm, in terms of normalized average L_2 distance between corresponding descriptors. Some results could not be computed by the participants.

	Strength				
Transform.	**1**	**≤2**	**≤3**	**≤4**	**≤5**
Isometry	0.57	0.58	0.62	0.62	0.65
Rasterization	–	–	–	–	–
Sampling	0.73	0.74	0.86	0.94	0.92
Holes	–	–	–	–	–
Micro holes	–	–	–	–	–
Scaling	0.62	0.61	0.65	0.65	0.75
Affine	1.08	1.32	1.46	1.61	1.77
Noise	3.24	3.37	3.37	3.34	3.32
Shot Noise	0.89	1.03	1.21	1.33	1.40
Partial	0.80	0.97	1.12	1.10	1.15
View	–	–	–	–	–

Table 1: *Quality of GHKS feature description algorithm (average normalized L_2 distance between descriptors at corresponding points).*

References

[BBK08] BRONSTEIN A. M., BRONSTEIN M. M., KIMMEL R.: *Numerical geometry of non-rigid shapes*. Springer, 2008. 2

[DK11] DAROM T., KELLER Y.: Scale invariant features for 3d mesh models. *http://yosikeller.web.officelive.com/publications/publications.html* (2011). 3, 4

[HS88] HARRIS C., STEPHENS M.: A combined corner and edge detection. In *Proc. of The Fourth Alvey Vision Conference* (1988), pp. 147–151. 2

[JH99] JOHNSON A. E., HEBERT M.: Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Trans. Pattern Anal. Mach. Intell. 21*, 5 (1999), 433–449. 3

[LBB10] LITMAN R., BRONSTEIN A., BRONSTEIN M.: Diffusion-geometric maximally stable component detection in deformable shapes. *Arxiv preprint arXiv:1012.3951* (2010). 3

[Low04] LOWE D.: Distinctive image features from scale-invariant keypoints. *IJCV 60*, 2 (2004), 91–110. 1, 4

[MCUP04] MATAS J., CHUM O., URBAN M., PAJDLA T.: Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing 22*, 10 (2004), 761–767. 3

[MDSB02] MEYER M., DESBRUN M., SCHRÖDER P., BARR A. H.: Discrete differential geometry operators for triangulated 2-dimensional manifolds. In *Proceedings of VisMath* (2002). 3

[MFK*10] MAES C., FABRY T., KEUSTERMANS J., SMEETS D., SUETENS P., VANDERMEULEN D.: Feature detection on 3D face surfaces for pose normalisation and recognition. In *Proc. BTAS* (2010). 3

[MS05] MIKOLAJCZYK K., SCHMID C.: A performance evaluation of local descriptors. *Trans. PAMI* (2005), 1615–1630. 1

[SB10] SIPIRAN I., BUSTOS B.: A robust 3D interest points detector based on Harris operator. In *Proc. Eurographics Workshop on 3D Object Retrieval* (2010), Eurographics Association, pp. 7–14. 2

[SOG09] 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)* (2009). 4

[SZ03] SIVIC J., ZISSERMAN A.: Video Google: A text retrieval approach to object matching in videos. In *Proc. ICCV* (2003), vol. 2, pp. 1470–1477. 1

[ZBVH09] ZAHARESCU A., BOYER E., VARANASI K., HORAUD R.: Surface feature detection and description with applications to mesh matching. 2, 3

[Zob10] ZOBEL V.: *Spectral Analysis of the Hodge Laplacian on Discrete Manifolds*. Master Thesis, 2010. 4

[ZRHar] ZOBEL V., REININGHAUS J., HOTZ I.: Generalized heat kernel signature. *Journal of WSCG, International Conference on Computer Graphics, Visualization and Computer Vision* (2011 to appear). 4

© The Eurographics Association 2011.

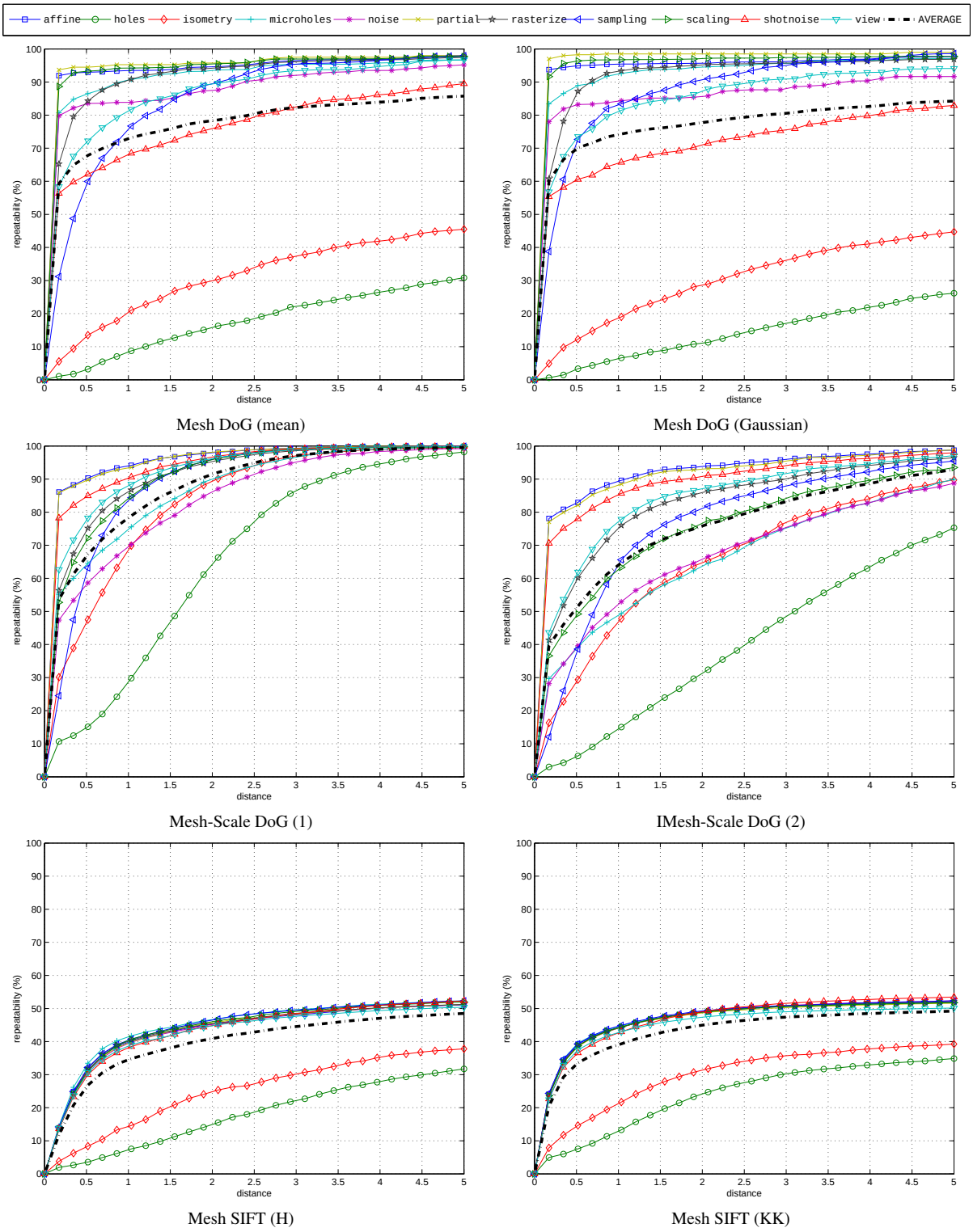

Figure 2: *Repeatability (%) vs distance of point feature detectors broken down according to different transformation classes.*

© The Eurographics Association 2011.

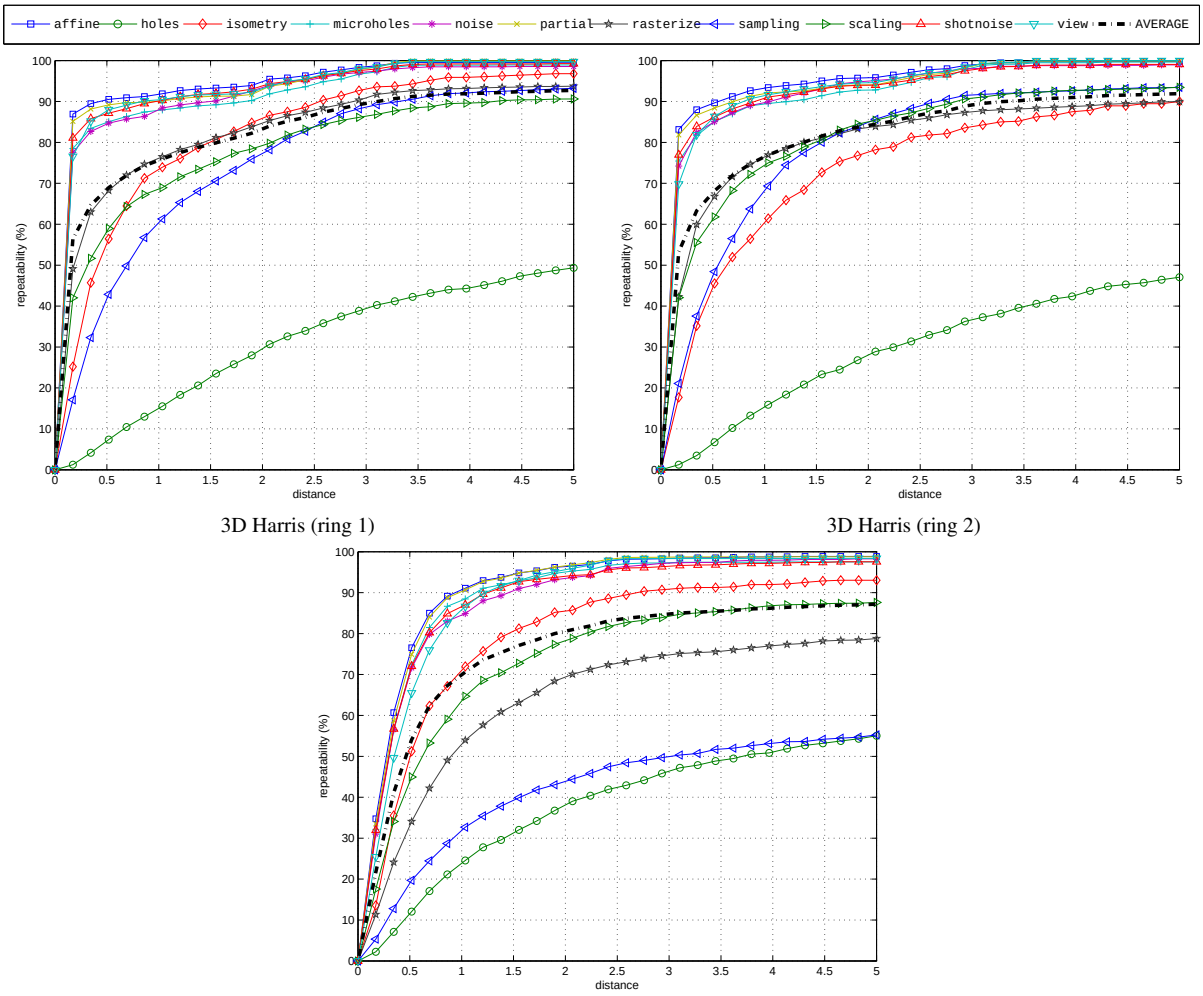

Figure 3: *Repeatability (%) vs distance of point feature detectors broken down according to different transformation classes.*

© The Eurographics Association 2011.

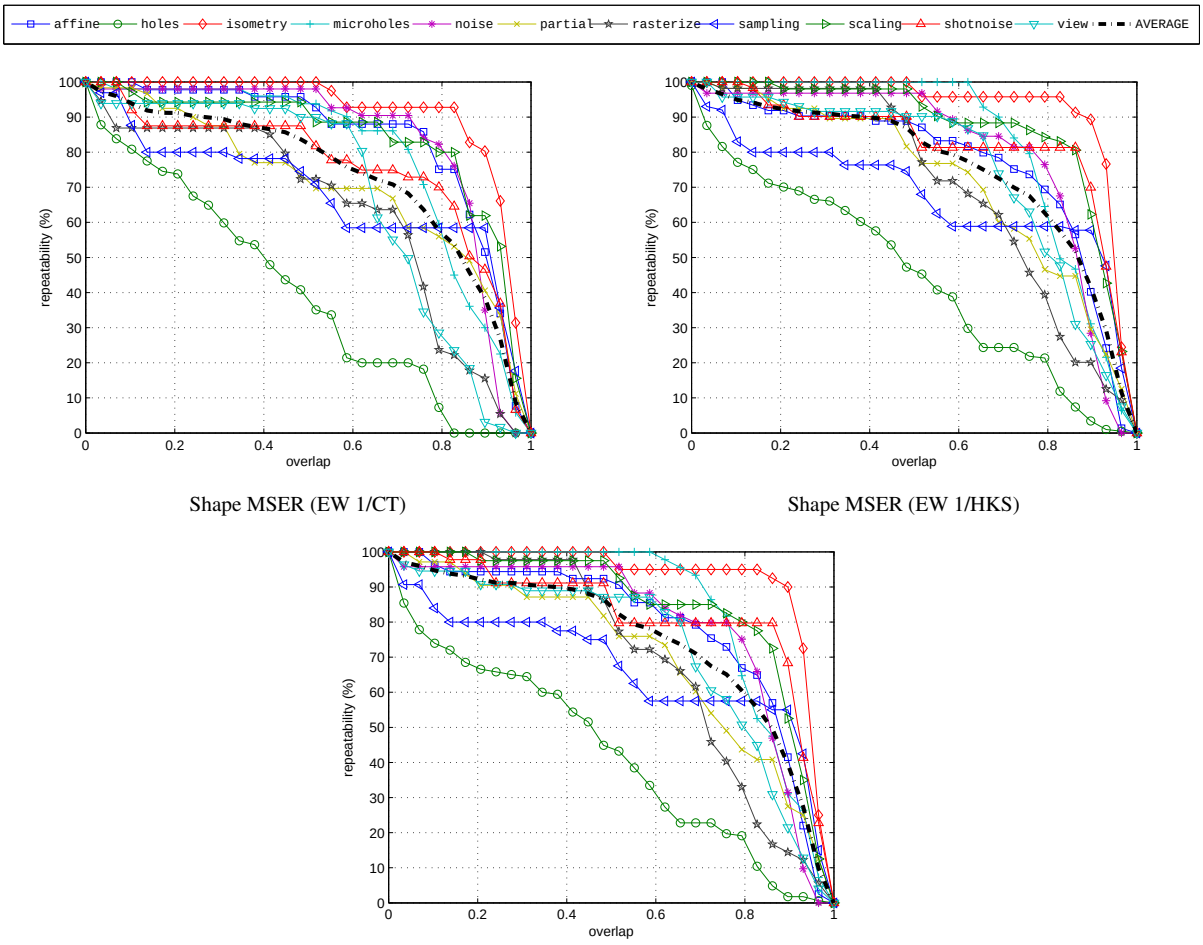

Figure 4: *Repeatability (%) vs overlap of region feature detectors broken down according to different transformation classes.*

© The Eurographics Association 2011.

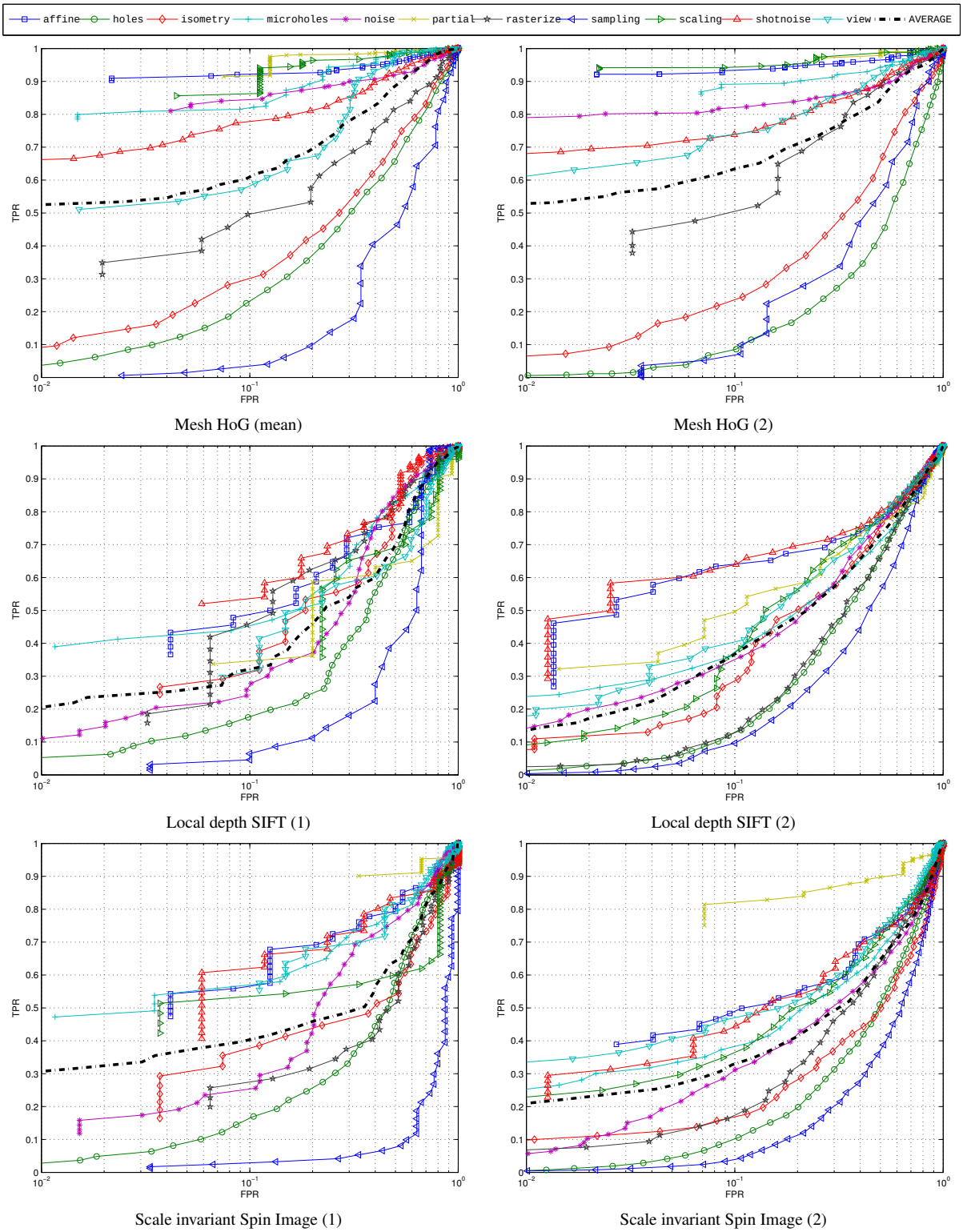

Figure 5: *ROC curves of point feature descriptors broken down according to different transformation classes.*

© The Eurographics Association 2011.