A Performance Evaluation of 3D Keypoint Detectors

Samuele Salti, Federico Tombari, Luigi Di Stefano
DEIS, University of Bologna
Bologna, Italy
{*samuele.salti,federico.tombari,luigi.distefano*}@*unibo.it*

***Abstract*—Intense research activity on 3D data analysis tasks, such as object recognition and shape retrieval, has recently fostered the proposal of many techniques to perform detection of repeatable and distinctive keypoints in 3D surfaces. This high number of proposals has not been accompanied yet by a comprehensive comparative evaluation of the methods. Motivated by this, our work proposes a performance evaluation of the state-of-the-art in 3D keypoint detection, mainly addressing the task of 3D object recognition. The evaluation is carried out by analyzing the performance of several prominent methods in terms of robustness to noise (real and synthetic), presence of clutter, occlusions and point-of-view variations.**

***Keywords*-3D detectors; performance evaluation; 3D object recognition**

I. INTRODUCTION

Automatic recognition of shapes in 3D data is an extremely active research field, due to the significant number of applications for which it represents a key stage, such as object manipulation and grasping, robot localization and navigation, scene understanding. Usually this task is tackled by either a global or a local approach. According to the former, a surface is described entirely by means of global features whereas the latter relies on local keypoints and regional feature descriptions to determine point-to-point correspondences between surfaces. Borrowing a denomination typical of the face recognition community [1] we refer here to these two approaches as, respectively, *holistic* and *feature-based*. While the holistic approach is popular in the context of 3D *object retrieval* [2]–[4], feature-based methods are inherently more effective for 3D *object recognition* in presence of cluttered backgrounds and occlusions [5]–[12].

As aforementioned, feature-based methods rely on 3D keypoints that are extracted from a 3D surface. This task is accomplished by 3D detectors, whose aim is to determine points which are distinctive, to allow for effective description and matching, and repeatable with respect to point-of-view variations and noise [5]–[7]. Sometimes, a characteristic scale is also associated to each keypoint, so as to provide a local neighborhood to the following description stage [5], [8], [13]–[15]. Then, a description of the local neighborhood of each keypoint is computed by means of a 3D descriptor [5]–[12], [15]. Descriptors are finally matched across different views to attain point-to-point correspondences.

This work is motivated by the belief that, given the wealth of recent literature proposals concerning 3D detectors, there is now the need to sum up the state-of-the-art and compare quantitatively the different approaches within a common and well defined experimental framework. Hence, inspired by the work concerning 2D features [16], [17], we propose a comparison of state-of-the-art 3D detectors. We mainly address the object recognition scenario, characterized by occlusion and clutter. In such a framework, we evaluate robustness to noise (real and synthetic) and point-of-view variations as well as computational efficiency.

A recent work, similar in motivation and spirit to ours, proposed an experimental evaluation of 3D detectors and descriptors focused on a shape retrieval scenario [18]. However, unlike object recognition, shape retrieval does not require to deal with occlusion, clutter and changes of viewpoint, the large intraclass shape variations being instead the main issue to be dealt with. In comparison presented in this work, we focus on object recognition and, in addition, propose some basic retrieval experiments to highlight how, interestingly, the absolute performance of the detectors as well as the ranking of their performance are influenced by the application scenario. Therefore, this paper and [18] provide complementary perspectives within the topic of quantitative evaluation of 3D local features.

II. 3D DETECTORS

This section briefly reviews state-of-the-art methods for detection of 3D keypoints. They are divided into two categories: fixed-scale detectors and scale-invariant detectors.

A. Fixed-scale Detectors

Fixed-scale detectors find distinctive keypoints at a specific, constant scale which is preset as a parameter of the algorithm. These approaches compute a distinctiveness, or *quality*, measurement associated with each point, that can be either point-wise (i.e. a property of a vertex of the mesh) or region-wise (i.e. a property of a region around each vertex, hereinafter referred to as *support*). Then, keypoints are selected by maximizing the quality measurement in a spatial neighborhood defined by the scale.

One example of the approaches relying on point-wise quality measurements is *Local Surface Patches* (LSP) [6]. It defines the quality of a vertex as its Shape Index (SI) [19], which in turn is based on the maximum and minimum principal curvatures at the vertex. A vertex is considered a keypoint if it is a global extremum of the SIs in the considered neighborhood and is significantly greater or smaller than the mean SI in the neighborhood, i.e. $SI_i \geq (1+\alpha)\mu_{SI}$ or $SI_i \leq (1-\beta)\mu_{SI}$.

As for the methods relying on region-wise quality measurements, both *Intrinsic Shape Signatures* (ISS) [7] and the proposal in [5], referred to hereinafter as *KeyPoint Quality* (KPQ), compute the Eigen Value Decomposition (EVD) of the scatter matrix of the points belonging to the support. ISS uses as distinctiveness the magnitude of the smallest eigenvalue (to include only points with large variations along each principal direction) and the ratio between two successive eigenvalues (to exclude points having similar spread along principal directions). In KPQ, instead, the support is aligned with its principal axes and a first pruning of non-distinctive points is performed by thresholding the ratio between the maximum lengths along the first two principal axes. The quality measurement is then determined by means of an empirical combination of the curvatures computed over a smoothed and re-sampled surface fitted to the aligned data.

B. Scale-invariant Detectors

Scale-invariant detectors perform a search for distinctive keypoints in a scale-space of the mesh, that extends the well-known concept defined for images [20]. This allows for detecting keypoints at different scales and for associating to them a characteristic scale used to define the support for the subsequent description stage. Similarly to fixed-scale methods, these approaches compute a quality measurement, which is however associated with each spatial position and *scale*. Then, keypoints are selected by maximizing the quality measurement spatially and across scales.

The proposals in [14], [21], [8] lay somewhere in the middle between 2D and 3D scale-spaces. They require a parametrization that maps the 3D mesh to a 2D plane, so as to exploit the lattice structure of the 2D image and apply conventional scale-spaces techniques. In [21], the parametrization is computed by mapping the border of the mesh (that must be already present or manually created by cutting a watertight mesh) to the border of a 2D image and then using the parametrization algorithm proposed in [22]. In [8] and [14] the parametrization is already available in the input data, since these methods work on range images. Given the parametrization, [21] and [8] create a scale-spaces representation of the *normal map* of the mesh, i.e. a color image where color channels represent normal components. The quality measure is the cornerness defined by the eigenvalues of the Gram matrix of the support. The algorithm flow is similar in [14] but the quality measure is represented by the mean (H) and Gaussian (K) curvatures (HK maps) and, instead of corners, connected regions of similar curvature are sought for.

The proposals in [13], [23] and MeshDoG [15] build scale-spaces directly out of the 3D mesh. The method in [23] uses as quality measure the displacement of each vertex from its original position after the application of the Difference-of-Gaussians (DoG) filter [24]. MeshDoG and [13], instead, explicitly avoid to modify the mesh geometry while creating the scale-space, by smoothing the value of an operator defined at each vertex instead of smoothing directly the 3D coordinates of the vertices. These operators are: an invariant computed on the Laplace-Beltrami operator, which corresponds to the displacement of a point along its normal of a quantity proportional to the mean curvature (H), in [13] (therefore this method is hereinafter referred to as Laplace-Beltrami Scale-Space (LBSS)); the approximation of the Laplacian operator as DoG applied to the mean curvature, the Gaussian curvature or the photometric appearance of a vertex, in MeshDoG. In MeshDoG additional filtering steps are introduced after detection: a maximum number of keypoints is detected, corresponding to a percentage value of the number of vertices of the mesh; as done in [24], non-corner responses are eliminated.

Heat Kernel Signature [25] employs as quality measurement the heat kernel [25] computed over the mesh: solving the heat equation over space and time allows for building an equivalent of the scale-space. The maxima of the kernel are then chosen as keypoints.

Differently from previous proposals, 3D SURF [26] builds a scale-spaces out of a voxelized version of the original mesh. The quality measurement, computed for each grid bin and at different octaves, is the Hessian of Gaussian second-order derivatives, that, given the nature of the data, can be computed efficiently by means of box-filtering.

Finally, in [5], scale-invariant keypoints are obtained from sets of fixed-scale keypoints extracted at different scales by the fixed-scale detector introduced in the previous paragraph as KPQ. The characteristic scale of a point is defined as that corresponding to the global maximum along the scale axis of the ratio between the maximum lengths along the principal directions (which, conversely to the fixed-scale case, is no longer thresholded). We denote the scale-invariant flavor of KPQ as KPQ-SI.

III. Methodology

A. Datasets

In our experiments we use four datasets. Two of them are synthetic, in the sense that they have been created applying known artificial deformation to 3D meshes in order to simulate two different application scenarios. The synthetic datasets have been created using models taken from the Stanford Repository[1]. The other two datasets are: the dataset[2] used in the experimental validation in [5], acquired with a laser scanner; the dataset[3] used in the experimental

[1] www.graphics.stanford.edu/data/3Dscanrep
[2] www.csse.uwa.edu.au/~ajmal
[3] vision.deis.unibo.it/SHOT

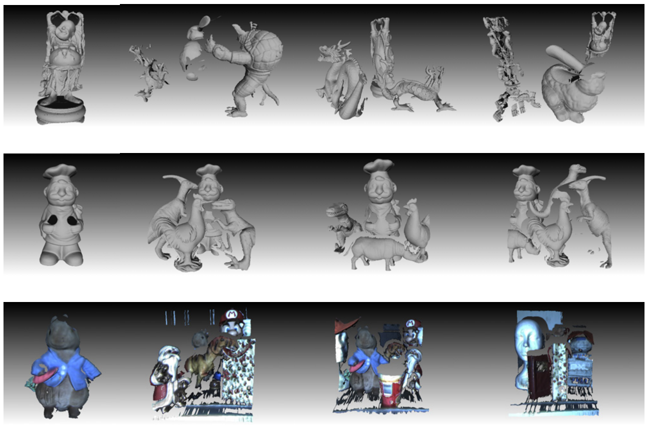


Figure 1. One model and three scenes from the datasets. From top to bottom row: *Random Views*, *Laser Scanner*, *Space Time*

validation in [12], obtained with the SpaceTime Stereo acquisition technique. In the following we will refer to the synthetic datasets as *Retrieval* and *Random Views*, to the dataset of [5] as *Laser Scanner* and to that of [12] as *Space Time*.

Each dataset comprises a set of models, $\mathcal{M} = \{\mathbf{M}_h\}_{h=1}^{N}$ and a set of scenes, $\mathcal{S} = \{\mathbf{S}_l\}_{l=1}^{M}$. Each scene contains a subset of the models. Only in the *Space Time* dataset objects not present in the model library has been additionally used to create the scenes. The ground-truth rotations and translations to align each model with its instance in the scene are known. In the case of the synthetic datasets, ground-truth is known by construction. For details on the way it was estimated in the other datasets the reader is referred to [12] and [5]. Fig. 1 shows examples of models and scenes taken from three of the datasets.

Datasets can also be categorized according to the application scenario they address. In one of the synthetic dataset, *Random Views*, as well as in *Laser Scanner*, each scene is a 2.5D mesh, i.e. a view of the spatial arrangement of the models from a specific vantage point, whereas the models are full 3D meshes. Therefore, these datasets are suitable for comparing the performance of the detectors in an object recognition scenario wherein a full 3D model is matched against a 2.5D view of the scene to detect its presence. The *Space Time* dataset represents a simpler scenario. In this dataset 2.5D models are retrieved in cluttered 2.5D views. Although simpler and not fully representative of all the challenges of an object recognition scenario, we have included it to test the performance of the detectors on a dataset acquired with a less accurate technique than laser scanning, that produces smoother, significantly less detailed meshes.

The second synthetic dataset, 'Retrieval', deal with a shape retrieval context, and is similar in spirit to the dataset used in [18]: only one full 3D model is used to create each scene and there are no occlusions and clutter. On the other hand, this dataset is much simpler than that used in [18], since the only difficulties represented in the scenes are rigid transformations and synthetic noise. Then main purpose of this dataset is to address a retrieval scenario using the same data as the *Random Views* dataset to highlight the impact of the application context on the performance of the detectors.

The synthetic 2.5D views were created by using from a random point of view the algorithm described in [27] on the 3D scene built by randomly rotating and translating the selected 3D models. Both synthetic datasets will be made publicly available.

B. Repeatability measures

The most important characteristic of a keypoint detector is its *repeatability*. This characteristic accounts for the ability of the detector to find the same set of keypoints on different instances of a given model, where the differences may be due to noise corruption, view point change, occlusion by other models or a combination of the previous nuisances.

Similarly to what was done in [16] for 2D keypoints, a keypoint extracted from the model $\mathbf{M}_h$, k_h^i and transformed according to the ground-truth rotation and translation, (R_{hl}, t_{hl}), is said to be repeatable if the distance from its nearest neighbor, k_l^j, in the set of keypoints extracted from the scene $\mathbf{S}_l$ is less than a threshold ϵ:

$$\left|\left|R_{hl}k_h^i + t_{hl} - k_l^j\right|\right| < \epsilon \ . \quad (1)$$

We evaluate the overall repeatability of a detector both in relative and absolute terms. Given the set RK_{hl} of repeatable keypoints for an experiment involving the model-scene pair $(\mathbf{M}_h, \mathbf{S}_l)$, the absolute repeatability is defined as

$$r_{abs} = |RK_{hl}| \quad (2)$$

whereas the relative repeatability is given by

$$r = \frac{|RK_{hl}|}{|K_{hl}|} \ . \quad (3)$$

The set K_{hl} is the set of all the keypoints extracted on the model $\mathbf{M}_h$ that are not occluded in the scene $\mathbf{S}_l$. This set is estimated by aligning the keypoints extracted on $\mathbf{M}_h$ according to the ground-truth rotation and translation and then checking for the presence of vertices in $\mathbf{S}_l$ in a small neighborhood (1 ring in our implementation) of the transformed keypoints. If at least a vertex is present in the scene in such a neighborhood, the keypoint is added to K_{hl}.

We consider the absolute repeatability, as in [17], because another important characteristic of a detector is the amount of repeatable keypoints it can provide to the subsequent modules of an applications. Too few keypoints can not be enough to apply geometrical verification or outlier removal schemes, whereas too many waste computational resources and make the task of pruning spurious higher-level hypothesis, such as object presence, considerably more challenging.

The various detectors generate different numbers of keypoints. This differences are due to intrinsic factors, such as the design of the algorithm, the filtering steps applied after the detection of salient structure, a predefined limit on the number of keypoints, etc... as well as extrinsic factors, such as the abundance of the regions considered salient by a detector in the test data. As discussed in [17], these differences may have an undesired impact on the repeatability scores: if the number of keypoints is large, many of them may be considered repeatable by accident and not because of the design of the detector. As done in [17], we choose to use the default parameter supplied by the authors rather than tuning them to make the detectors output the same number of keypoints, mainly because this is not possible for all the considered detectors and, in any case, the influence of the data can not be eliminated.

As discussed in the previous section, two classes of detectors are considered. In case of scale-invariant detectors, an additional repeatability score is introduced, the scale repeatability. Given the scales σ_h^i, σ_l^j of a pair of repeatable keypoints, (k_h^i, k_l^j), the scale repeatability for one pair is defined as:

$$r_{scale}^{ij} = \frac{V\left(\text{Sphere}(\sigma_h^i) \cap \text{Sphere}(\sigma_l^j)\right)}{V\left(\text{Sphere}(\sigma_h^i) \cup \text{Sphere}(\sigma_l^j)\right)} \tag{4}$$

with Sphere(σ) indicating the sphere of radius σ and V(Sp) the volume of the 3D region Sp. The overall scale repeatability for one model versus one scene is given by

$$r_{scale} = \frac{\sum\limits_{(k_h^i, k_l^j) \in RK_{hl}} r_{scale}^{ij}}{|RK_{hl}|} \ . \tag{5}$$

As noted in [13], the difference in dimensionality with 2D images makes this overlapping measure drop faster than for 2D detectors. Hence, care should be taken when interpreting results of this measure.

Finally, to give aggregates results we plot the average of the repeatability measures on the number of model-scene pairs of each dataset.

C. Selected Methods

The set of 3D detectors evaluated in our experiments includes: all the fixed-scale proposals introduced in Sec. II, namely LSP, ISS and KPQ; the MeshDoG, Laplace-Beltrami Scale-Space and KPQ-SI methods among scale-invariant detectors. As for MeshDoG, we used the publicly available original C++ implementation[4]. All other methods have been implemented in C++, as well. For the KPQs detector, however, it was necessary to use a surface smoothing and fitting routine available as a MATLAB script [28], that has been interfaced with the C++ implementation of the rest of the detector. It is important to keep in mind this difference when analyzing the performance reported in Tab. I.

Some of the methods presented in Sec. II have specific requirements on the input data that made their inclusion in this comparison unfeasible. Specifically, [21] requires that one and exactly one border is present in the input mesh. While this may be reasonable for a partial views registration, it is not in the case of retrieval of full 3D meshes nor of an object recognition scenario.

As for [14] and [8], these methods have been designed to work with range images. They both exploit the lattice structure that the range image provides in order to build a scale-space representation of the input. Although the meshes of the scenes we use are obtained from range images (laser scans or disparity maps), and in principle the transformation is invertible, there is no way to obtain a single range image for the 3D models. Because of this, they are not suitable for an object recognition scenario in which full 3D models are sought for in 2.5 views, as defined in this comparison.[5]

D. Parameters

All parameters have been fixed for the experiments on all datasets. Metric parameters, such as radius, distances, noise standard deviation, etc.. are expressed throughout the paper in units of *mesh resolution* (*mr*) [9], i.e. the mean length of the edges in the mesh. Default parameters proposed in the original publications have been used. The only tuned parameter is the Non Maxima Suppression radius in ISS and in the two variants of KPQ, because it was specified in metric units by the authors. It has been fixed as $4mr$ after running the detectors on a tuning scene with different values. MeshDoG results are reported using the mean curvature as quality measure, for we found that it yields better results than the Gaussian curvature.

Scale-invariant detectors have been run on the set of scales $\Sigma = \{2mr, 6mr, 10mr, 14mr, 18mr, 22mr\}$. This allows the detector to look for discriminative and repeatable structures ranging from point-wise scales to local and object sub-part scales. Since the first and last scale are used only to assess the presence of a local extremum in the immediately subsequent or antecedent scale, detections can happen only at scales $\tilde{\Sigma} = \{6mr, 10mr, 14mr, 18mr\}$. To compare results on the same set of structure sizes, we ran the fixed-scale detectors for each scale in $\tilde{\Sigma}$. The distance threshold ϵ is $2mr$. To simulate sensor noise, on synthetic datasets we

[4]svn://scm.gforge.inria.fr/svn/mvviewer

[5]Recently, the detector proposed in [8] has been used [29] for object recognition on the *Laser Scanner* dataset, by synthesizing range images from a number of uniformly distributed overlapping views of the 3D model of the object. This technique is not suitable for our experimental comparison because the performance of detectors working on range images will be influenced by external factors such as the synthetic views position and distribution.

added three levels of Gaussian noise with standard deviation equal to $0.1mr$,$0.2mr$ and $0.3mr$, respectively.

IV. Results and Discussion

A. Retrieval *and* Random Views *datasets*

Comparing the performance yielded by fixed-scale detectors, it is clear that on the *Retrieval* dataset the best results in terms of relative repeatability (Fig. 2a,2b,2c) are yielded by ISS, although KPQ shows to be overall more robust to noise. LSP, instead, performs poorly in presence of noise. This is probably due to the quality measure it employs, which is based on second-order derivatives. Also, the choice of selecting the maximum SI within the local support appears to be particularly error-prone since spurious peaks in the distribution of SI can easily occur because of noise. Since in terms of absolute repeatability (Fig. 2e,2e,2f) ISS yields a good number of points (≈ 100) and it is dramatically more efficient than LSP and KPQ (respectively 1 and 2 orders of magnitude, see Table I), this approach appears as the best choice for the object retrieval scenario.

By comparing these results with those obtained on the *Random Views* dataset (Fig. 3) we can see that algorithms performances change significantly. Overall, EVD-based fixed-scale detectors (i.e. ISS and KPQ) perform worse than in the retrieval scenario in presence of partially occluded shapes, since the absence of parts of the geometric structure modifies the scatter matrix, thus reducing the repeatability of the detector. Furthermore, and conversely to the case of retrieval, due to the presence of clutter it is not anymore beneficial to use large supports to increase repeatability (Fig. 3a,3b, 3c). On this dataset KPQ clearly outperforms ISS both in terms of relative and absolute repeatability, although it is still notably less efficient. LSP still performs poorly compared to both approaches, mainly due to the same reasons outlined for the *Retrieval* dataset.

For what concerns scale-invariant detectors, on the *Retrieval* dataset KPQ-SI reports overall the best repeatability results (Fig. 2g, 2h). Although with low noise-levels MeshDoG yields a similar relative repeatability and a higher number of repeatable keypoints, KPQ-SI shows superior robustness towards noise in terms of both absolute and relative repeatability, and a better scale invariance. This superior robustness can be motivated by the fact that the quality measure of KPQ averages curvatures computed at all the vertices in the support, while MeshDoG relies on DoGs of point-wise curvatures. Its efficiency is also comparable to that of MehsDoG (it runs 1.5 times slower partially using MATLAB code). As for LBSS, the local maxima of its invariant are extremely effective in determining the characteristic scale of the 3D structures even in presence of noise (Fig. 2i), which proves experimentally its theoretical characteristics. On the other hand, though, its performance are unsatisfactory in terms of spatial localization. Another important drawback of this technique is that it runs 1 order of magnitude slower than KPQ-SI and MeshDoG.

	Retrieval	Random Views	Laser Scanner	Space Time
LSP	$56 \sim 65$	$31 \sim 100$	$65 \sim 76$	$74 \sim 92$
ISS	$2 \sim 10$	$2 \sim 7$	$5 \sim 13$	$6 \sim 18$
KPQ*	$266 \sim 493$	$413 \sim 662$	$799 \sim 1109$	$544 \sim 1222$
LBSS	1585	461	1148	1397
MeshDoG	198	185	425	469
KPQ-SI*	303	364	634	767

Table I

Mean detection times on scenes for each dataset (in seconds). For fixed-scale detectors the minimum and maximum detection time, that varies with the scale, are reported. The asterisk indicates that part of the detector runs in MATLAB, hence the results indicate only the order of magnitude of the method efficiency.

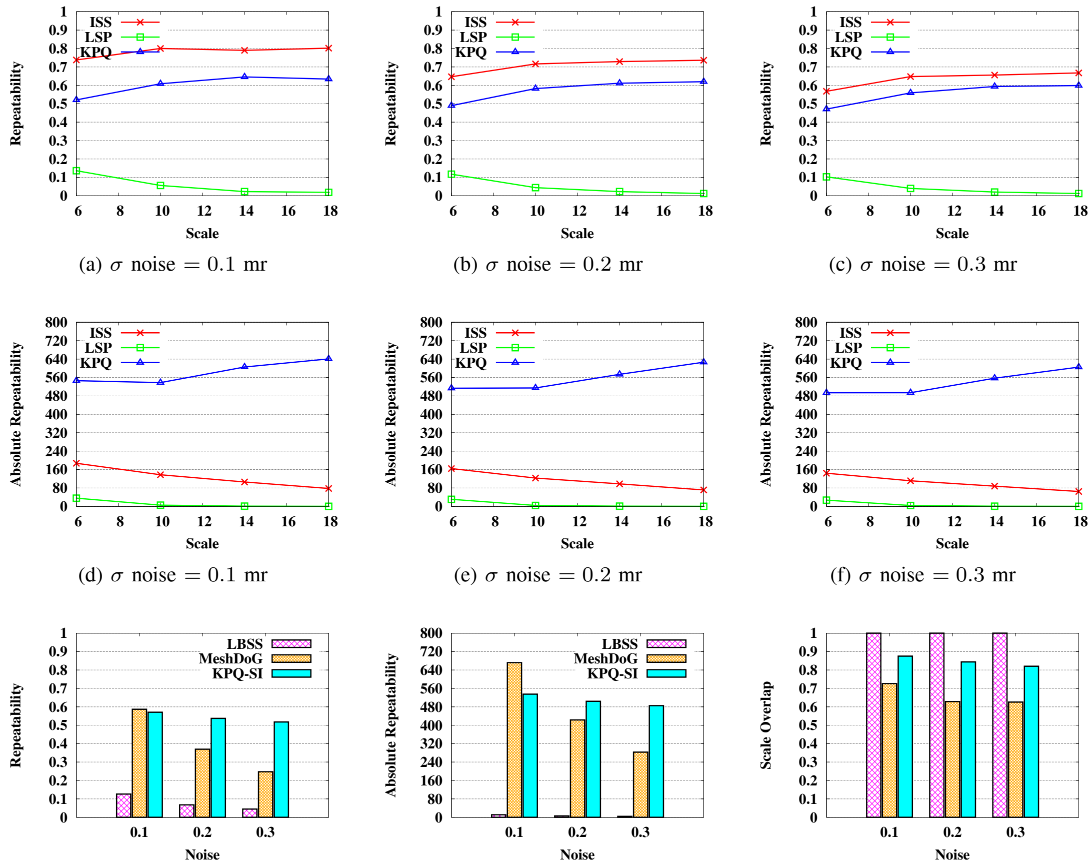
Analogously to case of fixed-scale detectors, the object recognition scenario appears notably more challenging than the retrieval one also for the scale-invariant detectors, with reduced performance reported by all algorithms on the *Random Views* dataset. More specifically, both MeshDoG and KPQ-SI report lower relative and absolute repeatabilities due to missing parts of the mesh (Fig. 3g, 3h). Still, KPQ-SI demonstrates being significantly more robust to noise, thus resulting as the best technique also on this dataset. It is interesting to note that, conversely to the previous scenario, MeshDoG performs better than KPQ-SI in terms of scale invariance (Fig. 3i), due to the fact that the characteristic scale in KPQ-SI in determined only by principal directions and, as aforementioned, EVD-based methods have problems in dealing with partial shapes. As for LBSS, its scale invariance is still the best one, and its efficiency is notably improved: nevertheless, the relative and absolute repeatability are yet not comparable to those of the other approaches.

Finally, by comparing the best approaches between both fixed-scale and scale-invariant detectors, ISS obtains a higher relative repeatability compared to MeshDoG and KPQ-SI on the *Retrieval* dataset, while KPQ and KPQ-SI have equivalent performances on the *RandomViews* dataset. ISS is notably the most efficient detector among all.

B. Laser Scanner *and* Space Time *datasets*

The main differences between the *Laser Scanner* and *Space Time* datasets are the point density variation between models and scenes and the dimensionality of the models. In the *Space Time* dataset, models and scenes have the same dimensionality (2.5D) and the same point density. In the *Laser Scanner* dataset models are full 3D meshes and their point density is one order of magnitude higher than in scenes.

The results obtained on real datasets are mainly consistent with the observations done for the 'Stanford Views' dataset. In particular, among fixed-scale detectors, KPQ is the top performer, given that it obtains higher or comparable relative repeatability than ISS while yielding an one order of magnitude greater absolute repeatability. The LSP



(a) σ noise $= 0.1$ mr (b) σ noise $= 0.2$ mr (c) σ noise $= 0.3$ mr

(d) σ noise $= 0.1$ mr (e) σ noise $= 0.2$ mr (f) σ noise $= 0.3$ mr

(g) (h) (i)

Figure 2. Results on the *Retrieval* dataset. Fixed-scale detectors: relative (a, b, c) and absolute repeatability (d,e,f) at different noise levels; scale-invariant detectors: relative (g) and absolute (h) repeatability, scale repeatability (i).

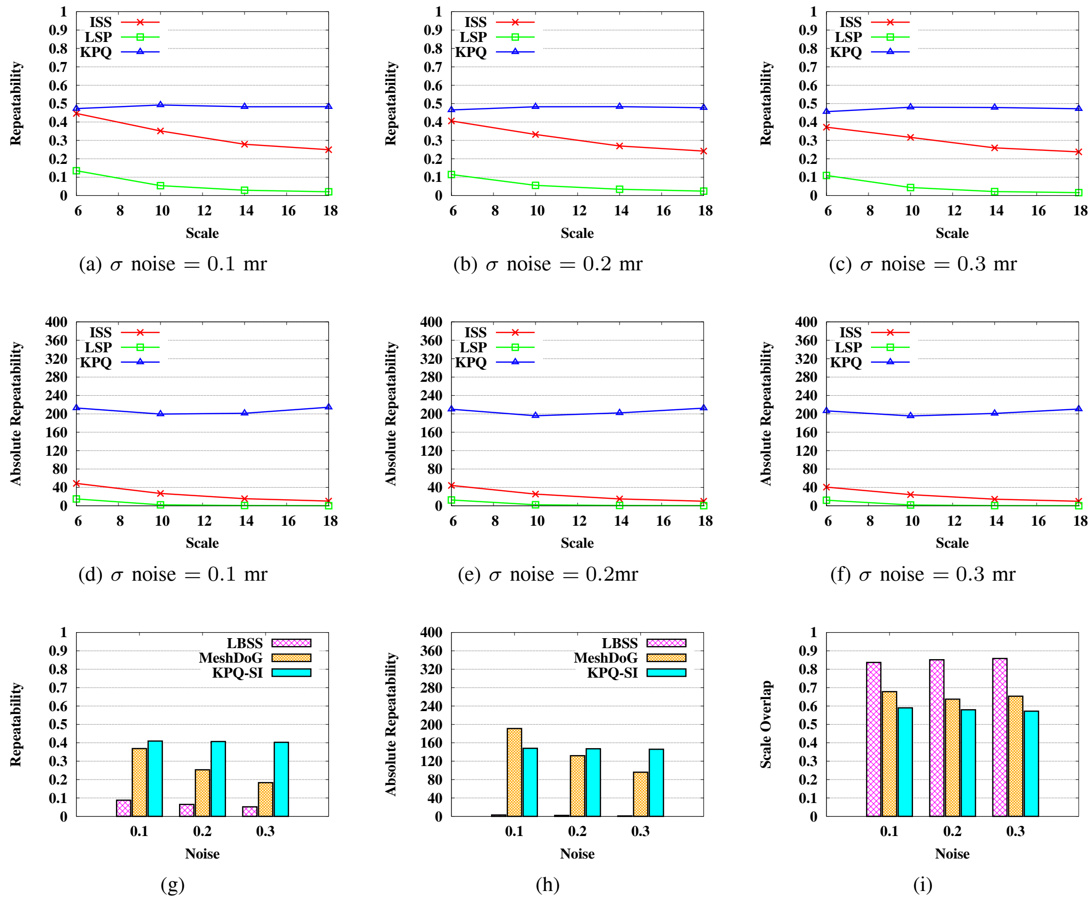
detector is not robust to the sensor noise present in these datasets. The detection times are as well consistent with the observations in the previous section, with ISS providing the most efficient solution. Even as far as scale-invariant detectors are concerned, many findings are consistent: LBSS provides outstanding scale overlaps between detections bat lacks spatial repeatability; the scale repeatability of both MeshDoG and KPQ are satisfactory, with KPQ suffering the difference in model and scene dimensionality, as the drop of performance between the *Space Time* and the *Laser Scanner* indicates.

A main difference is that, contrary to the synthetic dataset, MeshDoG has higher repeatability than KPQ-SI, both in absolute and relative terms, and it is the top performer among scale-invariant detectors. By comparing Fig. 4a with Fig. 5a we can also notice that, while MeshDoG is also the overall best detector on the *Space Time* dataset, its performance deteriorates on the *Laser Scanner* dataset. This fact combined with the good results MeshDoG yields on the 'Stanford Views' dataset (Fig. 3a), where a nuisance is the difference between model and scene dimensionality, indicates that MeshDoG suffers point density variations.

An interesting observation steams from the comparison of Fig. 3a, 3b, 3c with Fig. 5a. In both tests there is no difference in point density between the model and the scene. The only difference is the model dimensionality, that makes a part of the mesh included in the support be present only at detection time on the model in the 'Stanford Views' dataset. The fact that the performance of ISS deteriorates at greater scales on this dataset whereas they are constant on the *Space Time* dataset confirms that the alteration of the scatter matrix induced by the occlusion of part of the support is a severe challenge for this detector.

V. Conclusions

The experimental comparison proposed in this work has outlined many interesting aspects of state-of-the-art methods for 3D detection. First of all, it allowed assessing the best performing fixed-scale and scale-invariant methods over different datasets. Overall, KPQ-SI, MeshDoG and ISS yielded the best scores in terms of repeatability and ISS demonstrated to be the most efficient. Furthermore, it highlighted



(a) σ noise $= 0.1$ mr (b) σ noise $= 0.2$ mr (c) σ noise $= 0.3$ mr

(d) σ noise $= 0.1$ mr (e) σ noise $= 0.2$mr (f) σ noise $= 0.3$ mr

(g) (h) (i)

Figure 3. Results on the *Random Views* dataset. Fixed-scale detectors: relative (a, b, c) and absolute (d,e,f) repeatability at different noise levels; scale-invariant detectors: relative (g) and absolute (h) repeatability, scale repeatability (i).

different behaviors of the detectors on the tested datasets, which have been justified in light of the design of each method. Future work includes testing recent proposals [25], [26] and extending the proposed methodology to detectors working only with range maps [8], [14].

REFERENCES

[1] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Survey*, pp. 399–458, 2003.

[2] M. Iyer, S. Jayanti, K. Lou, Y. Kalyanaraman, and K. Ramani, "Three dimensional shape searching: state-of-the-art review and future trends," *Computer Aided Design*, vol. 5, no. 15, pp. 509–530, 2005.

[3] M. Ovsjanikov, J. Sun, and L. Guibas, "Global intrinsic symmetries of shapes," *Comput. Graph. Forum*, vol. 5, pp. 1341–1348, 2008.

[4] G. Somanath and C. Kambhamettu, "Abstraction and generalization of 3d structure," in *ACCV 2010*, 2010.

[5] A. S. Mian, M. Bennamoun, and R. A. Owens, "On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes," *IJCV*, vol. 89, no. 2-3, pp. 348–361, 2010.

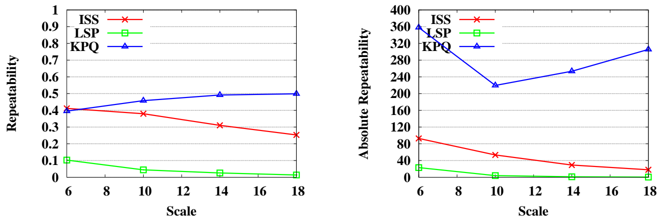
[6] H. Chen and B. Bhanu, "3d free-form object recognition in range images using local surface patches," *Pattern Recognition Letters*, vol. 28, no. 10, pp. 1252–1262, 2007.

[7] Y. Zhong, "Intrinsic shape signatures: A shape descriptor for 3d object recognition," in *ICCV-WS:3DRR*, 2009.

[8] J. Novatnack and K. Nishino, "Scale-dependent/invariant local 3d shape descriptors for fully automatic registration of multiple sets of range images," in *ECCV*, 2008, pp. 440–453.

[9] A. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3d scenes," *PAMI*, vol. 21, no. 5, pp. 433–449, 1999.

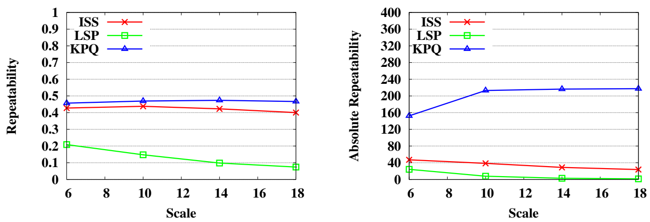
[10] C. S. Chua and R. Jarvis, "Point signatures: A new representation for 3d object recognition," *IJCV*, vol. 25, no. 1, pp. 63–85, 1997.



(a) (b)

	Rel. rep.	Abs. rep.	Scale rep.
LBSS	0.07	4	0.98
MeshDoG	0.36	351	0.63
KPQ-SI	0.30	233	0.51

Figure 4. Results on the *Laser Scanner* dataset. Fixed-scale detectors: relative (a) and absolute (b) repeatability; scale-invariant detectors: relative, absolute and scale repeatability (table).



(a) (b)

	Rel. rep.	Abs. rep.	Scale rep.
LBSS	0.06	2	1.00
MeshDoG	0.54	230	0.69
KPQ-SI	0.39	188	0.69

Figure 5. Results on the *Space Time* dataset. Fixed-scale detectors: relative (a) and absolute (b) repeatability; scale-invariant detectors: relative, absolute and scale repeatability (table).

[11] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik, "Recognizing objects in range data using regional point descriptors," in *ECCV*, vol. 3, 2004, pp. 224–237.

[12] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," in *ECCV*, 2010.

[13] R. Unnikrishnan and M. Hebert, "Multi-scale interest regions from unorganized point clouds," in *CVPR-WS: S3D*, 2008.

[14] E. Akagunduz and I. Ulusoy, "3D object representation using transform and scale invariant 3D features," in *ICCV*. IEEE, 2007, pp. 1–8.

[15] A. Zaharescu, E. Boyer, and K. Varanasi, "Surface feature detection and description with applications to mesh matching," *CVPR*, 2009.

[16] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of Interest Point Detectors," *IJCV*, vol. 37, no. 2, pp. 151–172, 2000.

[17] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A Comparison of Affine Region Detectors," *IJCV*, vol. 65, no. 1-2, pp. 43–72, 2005.

[18] A. M. Bronstein, M. M. Bronstein, B. Bustos, U. Castellani, M. Crisani, B. Falcidieno, L. J. Guibas, I. Kokkinos, V. Murino, M. Ovsjanikov, G. Patan, I. Sipiran, M. Spagnuolo, and J. Sun, "Shrec 2010: robust feature detection and description benchmark," in *EUROGRAPHICS Workshop on 3D Object Retrieval (3DOR)*, 2010.

[19] J. J. Koenderink and A. J. van Doorn, "Surface shape and curvature scales," *Image and Vision Computing*, vol. 10, no. 8, pp. 557 – 564, 1992.

[20] T. Lindeberg, "Feature detection with automatic scale selection," *IJCV*, 1998.

[21] J. Novatnack and K. Nishino, "Scale-dependent 3d geometric features," in *ICCV*, 2007, pp. 1–8.

[22] S. Yoshizawa, A. Belyaev, and H.-P. Seidel, "A fast and simple stretch-minimizing mesh parameterization," in *Shape Modeling Applications*, 2004, pp. 200–208.

[23] U. Castellani, M. Cristani, and S. Fantoni, "Sparse points matching by combining 3d mesh saliency with statistical descriptors," *Comp. Graphics Forum*, pp. 643–652, 2008.

[24] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, pp. 91–110, 2004.

[25] J. Sun, M. Ovsjanikov, and L. Guibas, "A concise and provably informative multi-scale signature based on heat diffusion," in *Proc. Symp. Geom. Proc.*, 2009, pp. 1383–1392.

[26] J. Knopp, M. Prasad, G. Willems, R. Timofte, and L. Van Gool, "Hough transform and 3D SURF for robust three dimensional classification," in *ECCV*, 2010.

[27] S. Katz, A. Tal, and R. Basri, "Direct visibility of point sets," *ACM Trans. Graph.*, vol. 26, July 2007.

[28] D. John, "Surface fitting using gridfit," MATLAB Central File Exchange, July 2010.

[29] P. Bariya and K. Nishino, "Scale-Hierarchical 3D Object Recognition in Cluttered Scenes," in *CVPR*, 2010, pp. 1657–1664.