

Scale Invariant Features for 3D Mesh Models¹

Tal Darom and Yosi Keller

Abstract—In this paper we present a framework for detecting interest points in three-dimensional meshes and computing their corresponding descriptors. For that we propose an intrinsic scale detection scheme per interest point and utilize it to derive two scale invariant local features for mesh models. First, we present the *Scale Invariant Spin Image* local descriptor that is a scale-invariant formulation of the Spin Image descriptor. Second, we adapt the SIFT feature to mesh data by representing the vicinity of each interest point as a depth map and estimating its dominant angle using PCA to achieve rotation invariance. The proposed features were experimentally shown to be robust to scale changes and partial mesh matching, and compared favorably to other local mesh features on the SHREC’10 and SHREC’11 testbeds. We applied the proposed local features to mesh retrieval using the *Bag-of-Features* approach and achieved state-of-the-art retrieval accuracy. Last, we applied the proposed local features to register models to scanned depth scenes, and achieved high registration accuracy.

I. INTRODUCTION

The analysis and retrieval of three-dimensional (3D) mesh data is a contemporary research challenge. Due to the proliferation in 3D acquisition devices such data is becoming ubiquitous and can be applied to a gamut of application such as face detection [4], molecule identification [22] and medical applications [16], to name a few.

Such tasks require efficient mesh representations that are denoted as *shape descriptors*. Shapes can be represented using global or local descriptors. A *global* descriptor characterizes a mesh M by a single vector $\mathbf{x} \in \mathbb{R}^d$, where d is fixed for all shapes, and can consist of characteristics such as volume, area, statistical moments, Fourier, eigendecomposition of the Laplace-Beltrami operator [8], wavelet [31], Diffusion Embeddings [43], or skeleton based representations [1].

Local 3D shape features have been given significant attention in recent years due to their proven effectiveness in computer vision applications [30], [25]. Let $M = (V, F)$ be a mesh object with a set of N_V vertices $V = \{v_i\}$, and a set of N_F triangle faces $F = \{f_i\}$.

Faculty of Engineering, Bar Ilan University, Ramat Gan, Israel.
tal.darom@gmail.com

Faculty of Engineering, Bar Ilan University, Ramat Gan, Israel.
yosi.keller@gmail.com

Each such triangle is represented as a triplet of vertices in V . For each vertex v_i we define a set of neighboring vertices V_{n_i} that includes all the vertices in V that share a triangle with v_i . A shape is represented by a set of local descriptors $\mathbf{x} = \{\mathbf{x}_j\}_1^J$, where each local descriptor \mathbf{x}_j characterizes a certain neighborhood of vertices in M , and each mesh might be represented by a different number of local descriptors. Thus, the first step in computing local descriptors based representations is to determine the subset of vertices $V_d = \{\mathbf{v}_j\}_1^J$ for which the local descriptors \mathbf{x} will be computed. Such vertices V_d are denoted as the *interest points* of the mesh M , and a descriptors is computed over a support S_j around each $v_j \in V_d$.

Applications such as mesh alignment and retrieval require that the same interest points are detected in different (and transformed) mesh representations of a 3D object. Hence, the detection of interest points should be *repeatable* and the detection scheme is denoted as a *detector*. Once detected each interest point is characterized by a *local descriptor* that should be discriminative, e.g. given two points with similar locality, their local features should be similar compared to the features of other dissimilar localities. Local features were successfully applied to mesh retrieval [30], [25], point-to-point matching [33] and mesh classification [26], to name a few.

In this work we aim to extend contemporary approaches to local mesh detectors and descriptors, by incorporating notions inspired by state-of-the-art computer vision schemes such as Lowe’s SIFT [23]. Our **first** contribution is to derive an intrinsic local scale measure for feature vertices. This allows to define a repeatable scale parameter over which the local feature is computed. Thus, local features relating to corresponding points in different manifestation of a particular mesh will be computed over a similar support, yielding scale invariant mesh descriptors. In particular, the scale estimate allows the proposed scheme to handle meshes sampled by nonuniform densities.

In our **second** contribution we utilize the introduced local-scale estimate to derive scale-invariant local mesh features. The first being a scale-invariant Spin Image descriptor (SISI) that is shown to outperform previous results in feature matching in terms of repeatability and

robustness. **Last**, we introduce the *Local Depth SIFT* (LD-SIFT) that is a novel local mesh descriptor that extends the SIFT image feature [23] to meshes. For that we derive a dominant angle estimation scheme that allows to make the LD-SIFT rotation-invariant, contrary to Spin Images [15] descriptors, where the rotation invariance is achieved via integration over the polar attributes. This makes the LD-SIFT more distinctive than the Spin Images and SISI descriptors. We tested the repeatability and robustness of the proposed descriptors using the SHREC’10 [3] and SHREC’11 [2] challenges mesh datasets and evaluation protocols, and they were shown to compare favorably to previous approaches. Aiming for higher level applications we also implemented a mesh retrieval scheme based on the Bag-of-Words paradigm, using the proposed features and achieved state-of-the-art retrieval accuracy.

This paper is organized as follows. In section II, we survey related works on local features in meshes and the Bag-of-Words approach to mesh retrieval. Our proposed intrinsic scale detection scheme is presented in Section III, and we show how to utilize it to derive scale-invariant descriptors in Section IV, where we introduce a scale-invariant formulation of the Spin Image feature and derive a novel scale-invariant feature based on the SIFT. We verify and exemplify the use of our approach in Section V, while concluding remarks and future extensions are discussed in Section VI.

II. RELATED WORK

In this section we review previous results in the study of local descriptors in mesh objects and their applications. We also review the The Bag of Words (BoW) approach to mesh retrieval. This is a state-of-the-art statistical approach to local features based mesh retrieval that will be used as a testbed for the proposed local features.

A. Interest points detectors

Given a mesh M , we aim to detect a subset of the vertices V_d with high descriptive value of the shape. The detection of interest points in depth images and meshes was studied by Mian *et al.* [25], who proposed to use the Principle Component Analysis (PCA) of the vertices contained in a sphere with radius r , to create a local 3D coordinate system at each point. As feature points, they selected the vertices for which the ratio between the two leading eigenvalues is maximized.

Another approach to feature detection is to detect the local extremum points of scalar functions defined over the surface. Such approach was applied by Ruggeri *et*

al. [33]. They compute the eigendecomposition of the Laplace-Beltrami operator over the surface, and define the interest points as the critical points of the eigenfunctions corresponding to the smaller eigenvalues of the Laplace-Beltrami operator.

A related approach to feature detection was suggested by Sun *et al.* [37] and G  bal *et al.* [13] that utilized the *Heat Kernel Signature* (HKS) function computed over on the mesh. The heat kernel signature function of a point v at time t is the relative quantity of heat that remains at the point of origin v at time t given a point source at v at time zero. The HKS is computed by applying the heat kernel to the shape using the Laplace-Beltrami operator, and the extremum points of the HKS, at large values of t , are chosen as feature points.

Some recent works were motivated by the local features schemes derived for images. Sipiran and Bustos [35] proposed the Harris 3D detector that extends the seminal Harris corner detector [14]. A quadratic patch is fit to the vicinity of a vertex and treated as an image. The Harris corner measure is computed at each vertex, using either an adaptive or a fixed ring-shaped support around the vertex.

Unnikrishnan *et al.* [39] estimate the scale at points sampled from a 3D curve, by first computing a local coordinate system using the PCA of the neighboring points on the curve. The scale at each curve point is chosen such that it best fits the tangent of the curve. As both the scale and the tangent plane are initially unknown, Unnikrishnan *et al.* iteratively compute the tangent plane using the estimated scale, and than reestimate the scale.

Novatnack and Nishino [28] proposed a method for detecting interest points and their corresponding scale in range images, by computing a scale space based on a normal map constructed from a depth image. The scale space is computed by filtering the original normal map with a set of Gaussian filters with increasing widths. Corners are detected as the spatial maxima of the local Gram matrix of each scale of the normal map in scale space.

An adaptation of the MSER image region detector to mesh data was presented by Litman *et al.* [20]. They used the graph-based Diffusion formulation to represent a shape as tree, and compute graph-theoretic measures such as the inverse of the Markovian commute time and inverse heat kernel.

Other schemes utilize the *Difference of Gaussians* (DoG) operator as a detector. Zaharescu *et al.* [42] presented the Mesh-DoG detector by applying a set of Gaussian filters to the curvature function defined on the mesh surface, and computing the octaves of the mesh. Each octave was then subtracted from its

successor to compute the DoG function. The interest points are detected at the local maxima of the DoG function in space and scale. The scale space approach was also utilized by Smeets et al. who presented the Mesh SIFT detector [24] that detects the local feature as scale space extrema. The scale space is constructed using a binomial filter and iterative convolutions. The minimal and maximal curvatures are used to filter some of the local maxima and detect salient points.

Another DoG-based approach was introduced by Castellani *et al.* [7], who apply a set of Gaussian filters with varying width to the mesh, and compute octave differences to evaluate the DoG function. This process is repeated several times on different downsampled replica of the input mesh. Local maxima (both in position and scale) of the DoG function, that repeat at multiple downsampling ratios are detected as interest points.

Our approach to feature points detection (Section III) relates to previous works motivated by local features detectors derived for images. We suggest an improved DoG based detector for detecting interest points and their intrinsic scale.

B. Local descriptors

In a gamut of applications such as mesh alignment and retrieval, detectors are used as a preprocessing step for detecting a sparse set of interest points in a mesh, that are then represented by *descriptors*. The Spin Image mesh descriptor was introduced in the seminal work of Johnson and Hebert [15]. This descriptor consists of a two-dimensional accumulator indexed by α , the perpendicular distance to the line through the surface normal, and β , the signed perpendicular distance to the tangent plane defined by the vertex normal and position. In order to achieve rotation-invariance, the accumulator sums all of the vertices corresponding to (α, β) in a radius around the interest point. In the original paper the support of the feature was set to span most of the vertices of the mesh and a descriptor was computed for each mesh vertex to derive a global descriptor. This approach was extended by Mitra *et al.* [26] that sub-sampled the mesh surface uniformly, and then used the sampled set of vertices as the centers of Spin Image descriptors, with a Euclidean support that is wider than the average distance between interest points.

The Mesh-HoG approach by Zaharescu *et al.* [42] provides both a detector (discussed above) and a descriptor. The descriptor is computed using a geodesic support region with a fixed size support that was set as 3% of the total surface area. It consists of three dimensional

histograms and is made rotation invariant by computing a repeatable local coordinate system.

The Mesh-SIFT approach by Maes *et al.* [24] also suggests a detector (discussed above) and a descriptor. Their descriptor first detects the feature's orientation as the peak of a histogram of the directions of the normal in the vicinity of the feature point. It then describes the feature point by dividing the area surrounding the feature point into 9 subareas, and encoding each of them by a combination of the histogram of normal directions, and the histogram of a shape index derived from the local curvature function.

A different approach to local features based on Heat Kernel Signatures (HKS) was suggested by Gebal *et al.* [13] and Sun *et al.* [37], while Ovjanikov *et al.* [30] applied the HKS to mesh retrieval. A scale invariant formulation of the HKS was proposed by Bronstein and Kokkinos [6].

Some 3D datasets are represented by volumetric data using voxels, that are three dimensional volume elements. This form of representation is common in the medical imaging community, and is the output for imaging modalities such as CT and MRI. It is also common in the field of video analysis to stack consecutive frames in a video sequence to create a space-time volume of voxels

Some volume descriptors and detectors were inspired by extending the SIFT approach. Scovanner *et al.* [34] presented the 3D-SIFT method, and showed that when used within a Bag-of-Words retrieval framework, it can be used for the retrieval of video sequences. This approach was later applied by Flitton *et al.* [11] to object recognition in complex CT volumes.

In this work we suggest two scale and rotation invariant local descriptors for 3D meshes (Section IV). The *Scale Invariant Spin Images* which is a scale invariant extension of the original Spin Images descriptors [15], and the *Local Depth SIFT* which is an extension of the SIFT local image descriptors [23] to 3D meshes.

C. Bag of Words

The *Bag of Words* (BoW) method is the state-of-the-art approach to the retrieval of images [36] and meshes [30], [9] from large scale repositories using local descriptors. Given a set of objects (either images or meshes) represented by local descriptors, BoW retrieval starts with the computation of a dictionary of features, by quantizing the features into a set of representative features, used as a dictionary words. In the second stage each dataset model is represented by the histogram of its local features, where each feature is represented by the

dictionary word closest to it. In the retrieval stage the histogram representing the queried model is compared with the histograms of the models in the database, and models with similar histograms are returned as query result.

Ohbuchi *et al.* [29] render depth images of 3D shapes from multiple views and use the SIFT descriptors of these set of images to describe the 3D shape. They quantize these descriptors to create a visual words dictionary and employ a BoW framework to retrieve matching shapes from a database. This work was extended by Furuya and Ohbuchi [12], that extract significantly more local visual features by sampling each depth image densely and randomly. A decision tree is then used to encode these features into visual words that are used for BoW based object retrieval.

The downside of the Bag-of-Words approach is the loss of spatial information. Ovjanikov *et al.* [30] use the statistics of neighboring features to introduce spatial knowledge. Knopp *et al.* [18] suggest a related approach that utilizes both a visual dictionary and spatial information. They voxelize the 3D shape and adapt the SURF image descriptors to 3D voxel data. The detected SURF features are quantized to create a visual vocabulary. For each class, a set of visual words and the vectors of their offset from the shape center are used as candidates in a Hough transform-like voting scheme. Knopp *et al.* [17] extended the previous work by using the dominant scale and directions of the detected local features to make the voting process rotation and scale invariant.

A different BoW approach was suggested by Liu *et al.* [21] who proposed to use a relatively small set of probability distributions, denoted as 'shape topics', from which the 'shape words' are generated. Each local feature is then classified to the topic with the highest probability to create it. Wessel *et al.* [41] proposed to learn the feature distributions within each class given a learning set of objects per class.

In our work the BoW approach is used (Section V-D) to exemplify and quantify the usability of the proposed local features to a high-level task such as mesh retrieval.

III. INTRINSIC SCALE DETECTION VIA DENSITY INVARIANT SMOOTHING

In this section we present a novel approach to scale detection of interest points on 3D meshes. We were inspired by the seminal work of David Lowe [23] in detecting interest points and their scale in images using the Difference of Gaussians (DoG) operator. For that, we propose to define a Gaussian filter on the mesh geometry, and compute a set of filtered meshes, that we denote

as *mesh octaves* $M^s = (V^s, F)$, where $M^0 = M$. Consecutive octaves are subtracted to compute the DoG function, and we define the local maxima (in location and scale) as our feature points.

In particular, we propose to apply a *density-invariant* Gaussian filter. Namely, the core difference between applying a Gaussian to an image compared to a mesh, is that the points on a mesh are nonuniformly sampled. Thus, the repeatability of the DoG detector might be significantly hampered by changes in the sampling density. This was experimentally verified in the rasterization test of the SHREC'11 [2] challenge reported in Table II, where the proposed density invariant DoG detector achieved close to 98% repeatability rate, compared to an average of 50% of the second best scheme.

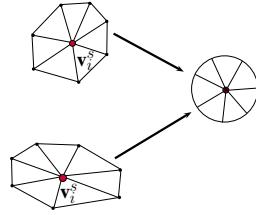


Fig. 1. Density invariant smoothing. Vn_i^s is the set of first order neighbors of v_i^s . The use of uniform weights is equivalent to mapping different sampling densities (shown on the left hand side of the figure) to the same configuration.

The mesh octaves are computed by applying a local filter with uniform weights. For each vertex v_i^s at an octave s , we compute the vertex v_i^{s+1} in the next octave by

$$v_i^{s+1} = \frac{1}{|Vn_i^s|} \sum_{v_j^s \in Vn_i^s} v_j^s, \quad (1)$$

where Vn_i^s is the set of first order neighbors of v_i^s , as shown in Fig. 1. Thus, the filter is invariant to the distance between the vertices, but not to their locations.

Let D_i be the local density at v_i^s

$$D_i = \frac{1}{|Vn_i^s|} \sum_{v_j^s \in Vn_i^s} |v_i^s - v_j^s|, \quad (2)$$

we approximate the initial variance by $\sigma_0 = D_i$.

The DoG function, scaled by the Gaussian's variance approximates the scale-normalized Laplacian of Gaussian [23]. Given two consecutive model octaves, the DoG function d_i^s at scale s is given by

$$d_i^s = \frac{1}{\sigma_s^2} |v_i^s - v_i^{s+1}|, \quad (3)$$

where σ_s^2 is the variance of the Gaussian filter at scale s . The same Gaussian filter is applied recursively to

compute the mesh octaves, resulting in a overall filtering width of $\sigma_N = \sqrt{N}\sigma$, and Eq. 3 is given by

$$d_i^s = \frac{1}{s \cdot \sigma_0^2} |v_i^s - v_i^{s+1}|. \quad (4)$$

Equivalently, in our density invariant formulation the overall filter width is given by

$$\sigma_N = \sqrt{N}D_i. \quad (5)$$

In comparison, Zaharescu *et al.* [42] apply a Gaussian operator to a scalar function defined on the mesh, while Castellani *et al.* [7] apply a Gaussian operator to the mesh geometry. Maes *et al.* [24] also compute a scale-space representation of the mesh using a DoG operator, and than use the mean curvature to detect feature points. All of these methods use Gaussian filters with a constant standard deviation, justifying their approach by assuming that the sampling of the object is nearly uniform, or that it can be made uniform by resampling.

We choose as feature points, the points v_i^s at a scale s that are local maxima both in scale and location. Figure 2 shows the interest points detected in the *Wolf* model in different scales. Each point is surrounded by a sphere depicting the support of the detected points.

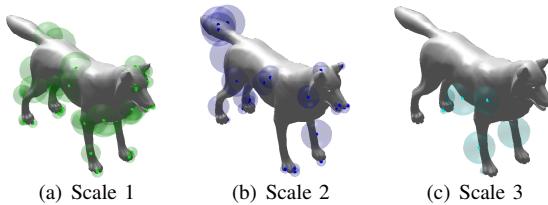


Fig. 2. The *Wolf* model with the detected feature points in different scales. Each detected point is surrounded by a sphere depicting the detected support.

In some of the previous works the same constant support was used for all of the feature points. It depended on the size of the whole model [15], or was hand tuned [26]. We follow the footsteps of works such as SIFT [23] and MeshSIFT [24] and propose that the support of each feature point is made proportional to the width of the filter, that was used to create the octave in which it was detected. It was shown in Eq. 5 that the result of recursively applying a filter s_i times with a width of D_i is equal to applying a filter with width of $S_i = \sqrt{s_i}D_i$, so we set the local scale to

$$S_i = C\sqrt{s_i}D_i, \quad (6)$$

where C is a predefined parameter used for all descriptors. The parameters C controls the trade-off of locality

vs. robustness in the computation of the local descriptors. A larger support makes the descriptor more robust to noise but less localized and distinctive. It follows that when the sampling density of the model is uniform, the local support depends only on s_i and on the constant C , and that all points detected in the same scale, will have the same support. We observed that most detected feature points were in the first ten octaves, and thus limited the number of octaves in our framework to 20. Compared to the Harris 3D corner detector [35], our method does not aim for a particular geometrical pattern (a corner), but detects any feature that is sufficiently salient.

The method of Mian *et al.* [25] computes a local 3D coordinate system using a local PCA computed at each surface point. The local features are selected at points where the ratio between the two leading eigenvalues is maximal. This method, provides a relatively small (compared to our scheme) number repeatable feature points, as it rejects locally symmetric surface patches.

IV. SCALE INVARIANT MESH DESCRIPTORS

In this section we utilize the estimate of the local scale, defined in Eq. 6, to derive scale invariant mesh descriptors. The core of our approach is to utilize the local scale to set the support over which the descriptor is computed. First we extend the Spin Image mesh descriptor and then introduce a new mesh descriptor that is based on the SIFT image descriptor [23].

A. Scale invariant Spin Images

The Spin Image local descriptor was presented by Johnson and Hebert [15], and has gained popularity due to its robustness and simplicity. A Spin images descriptor is computed at each vertex point using a constant predefined support, that spans most of the mesh. Thus, it depends on the object scale and global in nature. We propose to derive a Scale Invariant Spin Image (SISI) mesh descriptor, by computing the Spin Image descriptor over the local scale S_i . This results in an improved feature point matching. In particular, when the meshes are related by significant partial overlap as depicted in Fig. 3, where we show the optimal point-to-point correspondences in terms of the maximal ratio of the first to second matching distances. This measure of assignment validity follows the work of Lowe [23].

Another example is shown in Fig. 4 where we visually compare the SISI and Spin Image descriptors computed using a constant support based on the object's size. The descriptors are applied to a mesh models with significant partial matching.

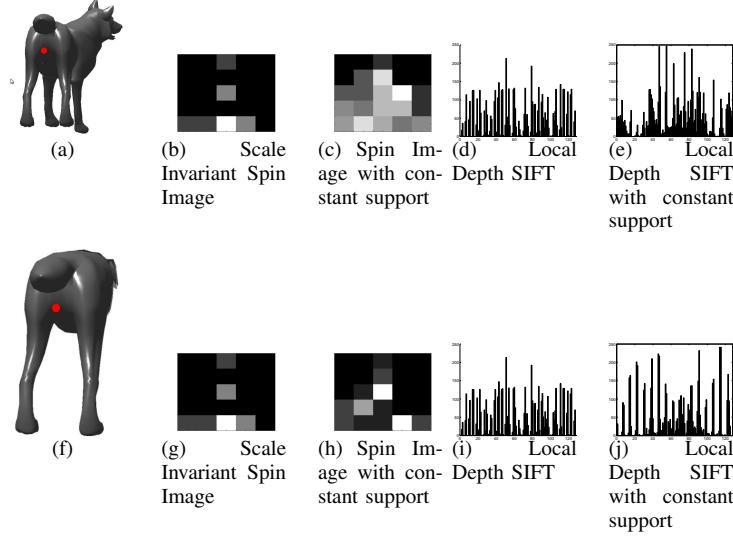


Fig. 4. Scale Invariant Spin Image and Local Depth SIFT descriptors compared to the same descriptors with a constant support based on the model's size. The Spin Image descriptors in (b) and (c) are computed with respect to the marked point in (a), and the descriptors in (g) and (h) are computed with respect to the Partial wolf model in (f). Note the similarity in (b) and (g) compared to (c) and (h). The Local Depth SIFT descriptors in (d) and (e) are computed with respect to the marked point in (a), and the descriptors in (i) and (j) are computed with respect to the Partial wolf model in (f). Note the similarity in (d) and (i) compared to (e) and (j).

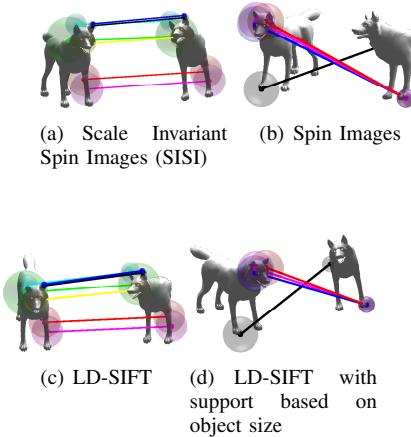


Fig. 3. Point to point matches between the *Wolf* model and its partial replica related by an isometric transform using different descriptors. We computed each of the different mesh descriptors at the same set of feature points detected by the proposed DoG-based detector. (a)-(d) shows the point-to-point correspondences having the maximal ratio between the distances of the two leading matches. The supports of the features are depicted by surrounding spheres.

B. Local Depth SIFT

The SIFT algorithm by Lowe [23] is a state-of-the-art scheme for computing scale and rotation invariant local features in images. It uses the DoG operator to detect feature points and their dominant scale, while

the dominant angle is recovered as the peak of the angular histogram of the gradients' directions. The SIFT descriptor is made invariant to scale and rotation by computing the descriptor over an adaptive support related to the local scale, and rotating this support according to the dominant angle. The descriptor consists of a radial-angular histogram of the pixel value derivatives.

We propose to compute a new local feature for 3D meshes we denote *Local Depth SIFT* (LD-SIFT). The LD-SIFT utilizes the DoG detector (Section III) to detect the interest vertices and estimate the local scale, while the descriptor is computed by representing the vicinity of the interest vertex (within a support defined by the dominant scale) as a depth map. Same as Mian *et al.* [25], we start by computing a plane at each interest point (detected by the detector), as depicted in Fig. 5.

Let $V = \{v_i^{S_i}\}$ be the set of vertices within the radius S_i around the interest vertex V_i , where S_i is the local scale of V_i . The depth map is computed by projecting the vertices V onto the *dominant plane* P , that is defined by its normal \mathbf{n} and the vertex V_i on it. The normal is computed using the two leading eigenvectors $\{\varphi_1, \varphi_2\}$ of the PCA in the vicinity of V_i within the local scale S_i

$$\mathbf{n} = \varphi_1 \times \varphi_2. \quad (7)$$

In general, P will be the tangent plane. But in practice, all we require is that its computation is repeatable.

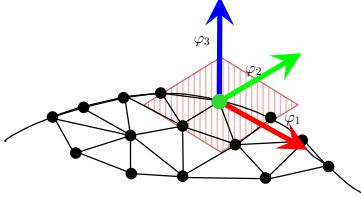


Fig. 5. The local coordinate system computed at a vertex V_i using PCA.

Equation 7 might fail if the eigenvalues $\{\lambda_2, \lambda_3\}$ corresponding to the local PCA eigenvectors are of the same order $\lambda_2 \approx \lambda_3$, as noise might cause them to swap places and corrupt the computation of P . Thus we require that

$$\lambda_2 \gg \lambda_3 \quad (8)$$

at each point where we compute the LD-SIFT. We compute the distance from each point on the surface to P where the support is given by Eq. 6. This results in a two-dimensional array for which we compute the SIFT descriptor.

In order to gain rotation invariance, we aim to define a *dominant angle* whose computation should be robust and repeatable. For that we utilize the direction of the leading eigenvector φ_1 that is guaranteed to lie in the plane P , and cyclically shift the radial-polar histograms within the SIFT descriptor, such that the dominant angle is the zero angle. In cases where

$$\lambda_1 \gg \lambda_2, \quad (9)$$

It was shown by Mian *et al.* [25] that it is guaranteed that the depth image is rotation invariant. We submit that for $\lambda_1 \approx \lambda_2$ our surface has a rotation symmetry around φ_3 , making our descriptor invariant to the selection of either φ_1 or φ_2 as a dominant angle. The Spin Image descriptor is made rotation invariant by integrating over the angular axis. Thus, the proposed LD-SIFT descriptor is more distinctive but its repeatability depends on the repeatability of the dominant angle estimate. An illustration of the resulting local coordinates is depicted in Fig. 5.

The LD-SIFT descriptor uses the PCA of the vertices detected within the feature's support to obtain the local coordinate system, while in Spin Images, the local coordinate system is determined by the surface's normal at the feature point. The normal is a *pointwise* geometrical property of the surface, while the PCA computation utilizes the covariance matrix of the patch enclosed within the local scale, and is thus an area operation that is more robust. Moreover, the computation of the leading

eigenvectors (utilized by PCA) is known to be noise-robust due to the spectral properties of p.s.d matrices [27], such as the covariance matrix.

The LD-SIFT scheme utilizes the direction of the leading local PCA eigenvector to determine the local dominant angle. The local coordinate system in which the descriptor is computed is than aligned according to that angle. In contrast, the Spin Image descriptor achieves rotation-invariance by integrating over the polar axis. Namely, aggregating the elevation in the reference coordinate system, and distance from the its origin. This results in data from all angles being accumulated into a single bin. The integration comes at the cost of reduced specificity of the descriptor.

Same as the method of Mian *et al.* [25] the LD-SIFT uses a depth image as the source of its descriptor. But, computes a SIFT-like descriptor based on derivatives, while Mian *et al.* use a down-sampled replica of the depth map as their descriptor.

V. EXPERIMENTAL RESULTS

In this section we experimentally verify and quantify the performance of the proposed DoG detector and the proposed SISI and LD-SIFT mesh features. For that we conduct extensive tests that relate to different attributes and applications. First, we directly measure the detector's repeatability and the descriptor's robustness using the SHREC'10 [3] and SHREC'11 [2] benchmark datasets and evaluation protocols. These results are reported in Sections V-A and V-B, respectively. Moreover, it allows us to compare against other state-of-the-art results in the SHREC'11 challenge. In Section V-C we measure the descriptors robustness in terms of a K Nearest Neighbors (KNN) retrieval test, that measures the probability of finding the correct assignment of a feature point within the its K most similar descriptors. This test relates to mesh alignment schemes such as RANSAC [10] that are able to recover the alignment in the presence of outlier matches. In Section V-D we applied the proposed local features to the higher-level task of mesh retrieval. Last, in section V-E we apply a RANSAC scheme to register and find the correct alignment of high quality models to a scanned depth scene in which they are partially present using the dataset of Mian *et al.* [25].

A. Interest point repeatability

In order to quantify the repeatability of the difference of Gaussians interest points detector we followed the methodology and testing protocol used in the SHREC'10 [3] and SHREC'11 [2] benchmarks. In their testing protocol we are given two mesh models M_1 and

M_2 . An interest point $\mathbf{x} \in M_1$ is considered ‘repeated’ if an interest point $\mathbf{y} \in M_2$ is detected that is no more than 5 units (about 2% of the models height) away from its location in M_1 . We then swap the models and repeat the experiment. The repeatability score is the average percentage of repeated points between the two models.

The “SHREC’10: robust feature detection and description benchmark” dataset is comprised of three mesh models, taken from the TOSCA high-resolution [5] database. Nine transformations are applied to each model at 5 different strength levels (except for the isometry transformation, for which each version is just a different isometry). Three families of feature detection methods were evaluated in the SHREC’10 benchmark: heat kernel-based features [37], 3D Harris features [35], and salient points. [7]. Three families of feature description methods were also evaluated: heat kernel signature [37]; dense heat kernel signature [30], and spin image signatures computed on the feature points detected by the Salient Points method, as used by Toldo *et al.* [38].

The repeatability results for the SHREC’10 dataset are shown in Table I. It follows that our detector proved superior in most of the deformation categories, in particular for higher strength levels, where the deformations are severe.

We also applied our DoG detector to the “SHREC’11: robust feature detection and description benchmark” [2]. In this benchmark eleven different transforms were applied to a single mesh model taken from the TOSCA high resolution database and the repeatability of the detected feature points was computed. Three other feature detection algorithms were submitted to this benchmark: Harris3D [35], Mesh-DoG [42] and Mesh-SIFT [24]. The results are reported in Table II.

As for the SHREC’ 10 and SHREC’ 11 results it follows that the proposed detector achieves almost perfect repeatability (> 97) for all transformations, other than down-sampling. Overall it compares favorably with the other result submitted to these benchmarks. In particular, our approach is proved to be fully repeatable (> 99) when applied to models with pure scale changes, which confirms that we achieved our core goal of scale-invariant detection. It can also be seen in the SHREC’ 11 results that while for some transformations the problem can be dimmed ‘solved’ (repeatability > 97), for the rasterization, sampling, and noise transformations that closely represent real life problems our method greatly improves on the current state-of-the-art, including approaches such as Mesh-DoG [42] and Mesh-Sift [24] that also employ a DoG approach. For the rasterization transformation, that simulates non-pointwise topological

artifacts due to occlusions in 3D geometry acquisition, our detection algorithm was twice as repeatable as any other scheme in the SHREC’11 benchmark, making it the preferred choice for processing raw 3D scanned data.

B. Descriptor robustness

In order to quantify our descriptors robustness we followed the methodologies and evaluation protocol used in the SHREC’10 [3] and SHREC’11 [2] benchmarks. As before, the feature robustness is quantified by considering two mesh models M_1 and M_2 , where M_2 is a transformed replica of M_1 . Let $\mathcal{F}(M_1) = \{f_l\}_{l=1}^{|\mathcal{F}(M_1)|}$ and $\mathcal{F}(M_2) = \{g_j\}_{j=1}^{|\mathcal{F}(M_2)|}$ denote the descriptors computed in M_1 and M_2 , respectively. In the SHREC’10 robustness benchmark the robustness is quantified by

$$d(l, j) = \frac{\|f_l - g_j\|_2}{\|f_l\|_2 + \|g_j\|_2}, \quad (10)$$

where $f_l \in M_1$, and $g_j \in M_2$ is its nearest descriptor in terms of location on the mesh. In the SHREC’11 robustness benchmark, the robustness is quantified using

$$d(l, j) = \frac{\|f_l - g_j\|_2}{\frac{1}{|\mathcal{F}(M_1)|^2 - |\mathcal{F}(M_1)|} \sum_{l,j \neq l} \|f_l - g_j\|_2}. \quad (11)$$

The overall descriptor robustness per mesh object (in both benchmarks) is computed by averaging the robustness over all of the feature points in M_1 . For the robustness experiments we set the local scale parameter $C = 3$.

We used the same datasets as in the interest points repeatability benchmark. The “SHREC’10: robust feature detection and description benchmark” dataset comprised of three models, with eight transformations in five transform levels each. The robustness results for the SHREC’10 dataset are shown in Tables III and IV. When compared to the robustness of the other descriptors in the SHREC10 challenge, the proposed SISI descriptor was more robust with respect to the microholes, shot noise (for transform levels 1 & 2), topology transformations. In particular, both our descriptors were shown to be invariant to scale changes. Although the LD-SIFT descriptor seems inferior to the SISI, we note that its applicability is more evident in the Point to Point matching tests in Section V-C.

We also applied the proposed SISI and LD-SIFT descriptors to the “SHREC’11: robust feature detection and description benchmark” [2] and participated in the challenge. In this benchmark eleven different transforms were applied to a single mesh model at five transformation levels, and robustness of the feature description was computed. The results for the SHREC’11 dataset

transform level	1	2	3	4	5
isometry	100.00 (100.00, HK2)	99.79 (100.00, HK2)	98.29 (100.00, HK2)	97.71 (100.00, HK2)	97.55 (100.00, HK2)
topology	100.00 (97.44, HK1)	99.95 (96.10, HK1)	99.92 (92.26, HK1)	99.87 (91.22, HK1)	99.85 (89.83, H2)
holes	99.62 (91.48, HK1)	98.78 (90.60, HK1)	98.07 (89.25, H1)	97.57 (88.82, H1)	97.16 (88.49, H1)
microholes	99.99 (100.00, HK2)	99.99 (100.00, HK2)	99.99 (98.15, HK2)	99.99 (96.58, HK2)	99.99 (95.64, HK2)
scale	100.00 (100.00, HK2)	100.00 (100.00, HK2)	100.00 (100.00, HK2)	100.00 (98.61, HK2)	100.00 (97.78, HK2)
localscale	99.97 (98.08, HK1)	99.83 (96.79, HK2)	99.66 (93.02, HK2)	98.65 (87.25, HK2)	97.58 (82.90, HK2)
sampling	98.89 (100.00, HK2)	97.11 (100.00, HK2)	94.79 (100.00, HK2)	90.09 (100.00, HK2)	74.49 (96.20, HK2)
noise	99.75 (100.00, HK2)	98.87 (95.19, HK2)	98.12 (93.16, HK2)	97.71 (89.37, HK2)	97.26 (87.62, HK1)
shotnoise	100.00 (100.00, HK2)	99.88 (96.22, HK1)	99.75 (93.39, HK1)	99.39 (90.45, HK1)	98.89 (87.32, HK1)

TABLE I
 DIFFERENCE-OF-GAUSSIANS DETECTOR REPEATABILITY TESTED ON THE “SHREC’10 FEATURE DETECTION AND DESCRIPTION BENCHMARK” DATASET. WE REPORT THE PERCENTAGE OF REPEATED FEATURES, WHERE 100% IS A PERFECT SCORE. THE TRANSFORMATIONS ARE GIVEN IN AN ASCENDING ORDER OF SEVERITY. WE SHOW IN PARENTHESES THE LEADING RESULTS IN THE SHREC’10 BENCHMARK [3]. H1 STANDS FOR THE HARRIS 3D FEATURES [35], AND HK1-2 ARE THE HEAT KERNEL FEATURES [37]. WE MARKED WITH BOLD THE CATEGORIES WHERE OUR APPROACH LED THE SHREC10 CHALLENGE.

transform level	1	2	3	4	5
Isometry	99.75 (99.81, H1)	99.83 (99.90, H1)	99.85 (99.87, H2)	99.87 (99.90, H2)	99.87 (99.77, H2)
Rasterization	98.25 (49.63, H2)	98.28 (51.63, H1)	98.23 (53.11, H1)	98.26 (50.10, H1)	98.17 (47.53, H1)
Sampling	99.62 (97.20, H2)	99.61 (98.27, H1)	99.52 (98.18, H1)	99.40 (97.13, H1)	99.52 (87.71, H1)
Holes	99.75 (100.00, H1-2)	99.76 (100.00, H2)	99.70 (99.94, H2)	99.67 (99.72, H2)	99.61 (99.63, H2)
Micro holes	99.25 (99.26, H1)	99.29 (99.09, H1)	99.26 (99.04, H2)	99.25 (98.98, H2)	99.20 (98.95, H2)
Scaling	99.92 (100.00, H1-2)	99.92 (100.00, H1-2)	99.92 (100.00, H1-2)	99.92 (99.90, H1)	99.92 (99.92, H1)
Affine	99.73 (99.43, H1-2)	99.83 (99.52, H1)	99.83 (98.16, H1)	99.83 (95.19, H1)	99.83 (93.75, H1)
Noise	99.94 (100.00, H1)	99.92 (99.24, H2)	99.92 (97.97, H2)	99.92 (95.86, H2)	99.91 (93.45, H2)
Shot Noise	99.74 (96.95, H1)	99.75 (95.71, H2)	99.74 (94.67, H2)	99.71 (93.90, H2)	99.71 (93.45, H2)
Partial	99.91 (100.00, H1-2)	99.88 (99.46, H1-2)	99.90 (99.35, H2)	99.82 (99.05, H1)	99.85 (99.24, H1)
View	99.97 (100.00, H1)	99.89 (100.00, H2)	99.88 (100.00, H2)	99.84 (100.00, H2)	99.84 (100.00, H2)
Average	99.62 (94.41, H2)	99.63 (94.39, H2)	99.61 (93.95, H1)	99.59 (93.10, H1)	99.58 (91.71, H1)

TABLE II
 DIFFERENCE-OF-GAUSSIANS DETECTOR REPEATABILITY TESTED ON THE “SHREC’11 FEATURE DETECTION AND DESCRIPTION BENCHMARK” DATASET. WE REPORT THE PERCENTAGE OF REPEATED FEATURES, WHERE 100% IS A PERFECT SCORE. THE TRANSFORMATIONS ARE GIVEN IN AN ASCENDING ORDER OF SEVERITY. WE SHOW IN PARENTHESES THE LEADING RESULTS IN THE SHREC’10 BENCHMARK [3]. H1-2 STANDS FOR THE HARRIS 3D FEATURES (RING 1-2) [35]. WE MARKED WITH BOLD THE CATEGORIES WHERE OUR APPROACH LED THE SHREC11 CHALLENGE.

are shown in Tables V and VI. It follows that the SISI descriptor outperformed the other schemes on the rasterization and view transformations. The LD-SIFT descriptor was shown to be more robust to the rasterization transformation than other algorithms, but for the SISI. It compares favorably to all the algorithms submitted to SHREC’11 in terms of robustness to noise.

From these results it follows that the combination of scale and rotation invariance with the local nature of our features makes them robust to many of the transformations applied in the these benchmarks. These transformations are the scale, local scale, topology, isometry and affine transformations. Any approach based on local feature is robust to view and partial transformations. For these transformations most other methods submitted to these benchmarks gave good results. Furthermore, our methods robustness to the noise, shot noise, holes and microholes transformations depend on the selected value of the constant C that weighs the trade-off of locality vs. robustness. The sampling and rasterization transformations presented a more challenging scenario of

changes in the sampling of points in the meshes. In this case the fact that the scale differences between octaves of our scale space is computed locally based on the local point density have given our detector an advantage over other detectors.

C. Point to point matching

A common use of mesh local features is in point-to-point matching between similar (but not identical) meshes. Matching schemes such as RANSAC [10] can integrate the matchings of a set of points and handle outlier matchings. Thus, we also measure the local features’ performance using a KNN test. For each point $x \in M_1$ we select the K points $\{y_i\}_1^K \in M_2$ closest to x in terms of the L_1 norm of the descriptors difference. If any of $\{y_i\}_1^K$ is located no more than 5 units (about 2% of the models height) away from the true corresponding point in M_2 the match is considered successful. Compared to the SHREC descriptor evaluation protocol, this test prefers sensitivity to specificity. Namely, it allows more outlier matchings (low specificity) for sake of having

transform level	1	2	3	4	5
isometry	0.0000 (0.01, DHK2)	0.0790 (0.01, DHK2)	0.0853 (0.01, DHK2)	0.0712 (0.01, DHK2)	0.0672 (0.01, DHK2)
topology	0.0012 (0.02, DHK2)	0.0025 (0.02, DHK2)	0.0039 (0.02, DHK2)	0.0059 (0.02, DHK2)	0.0083 (0.02, DHK2)
holes	0.0644 (0.02, DHK2)	0.1036 (0.02, DHK2)	0.1376 (0.02, DHK2)	0.1795 (0.03, DHK2)	0.2159 (0.03, DHK1)
microholes	0.0015 (0.01, DHK2)	0.0021 (0.01, DHK2)	0.0030 (0.01, DHK2)	0.0041 (0.01, DHK2)	0.0057 (0.02, DHK2)
scale	0.0000 (0.04, SHK2)	0.0000 (0.04, SHK2)	0.0000 (0.04, SHK2)	0.0000 (0.04, SHK2)	0.0000 (0.04, SHK2)
localscale	0.0259 (0.02, DHK2)	0.0716 (0.03, DHK2)	0.1328 (0.05, DHK2)	0.2116 (0.07, DHK2)	0.2669 (0.10, DHK2)
sampling	0.1023 (0.02, DHK2)	0.1433 (0.02, DHK2)	0.1689 (0.02, DHK2)	0.1643 (0.02, DHK2)	0.1841 (0.02, DHK2)
noise	0.1314 (0.02, DHK1)	0.1790 (0.06, DHK1)	0.2149 (0.09, DHK2)	0.2430 (0.12, SHK1)	0.2716 (0.13, SHK2)
shotnoise	0.0020 (0.01, DHK2)	0.0083 (0.01, DHK2)	0.0216 (0.02, DHK2)	0.0493 (0.02, DHK2)	0.0852 (0.02, DHK2)

TABLE III

SCALE INVARIANT SPIN IMAGE ROBUSTNESS (0 IS FULL ROBUSTNESS), TESTED ON THE “SHREC’10: ROBUST FEATURE DETECTION AND DESCRIPTION BENCHMARK” DATASET. WE SHOW IN PARENTHESES THE LEADING RESULTS OBTAINED IN THE SHREC’10 BENCHMARK [3]. DHK1 STANDS FOR THE DENSE HEAT KERNEL SIGNATURES, SHK2 IS THE SPARSE HEAT KERNEL SIGNATURE BASED ON FEATURES DETECTED BY HK2 [30]. WE MARKED WITH BOLD THE CATEGORIES WHERE OUR APPROACH LED THE CHALLENGE.

transform level	1	2	3	4	5
isometry	0.0000 (0.01, DHK2)	0.2438 (0.01, DHK2)	0.2361 (0.01, DHK2)	0.2383 (0.01, DHK2)	0.2519 (0.01, DHK2)
topology	0.0011 (0.02, DHK2)	0.0024 (0.02, DHK2)	0.0047 (0.02, DHK2)	0.0072 (0.02, DHK2)	0.0093 (0.02, DHK2)
holes	0.0599 (0.02, DHK2)	0.0976 (0.02, DHK2)	0.1316 (0.02, DHK2)	0.1648 (0.03, DHK2)	0.1874 (0.03, DHK1)
microholes	0.0134 (0.01, DHK2)	0.0197 (0.01, DHK2)	0.0259 (0.01, DHK2)	0.0311 (0.01, DHK2)	0.0358 (0.02, DHK2)
scale	0.0000 (0.04, SHK2)				
localscale	0.0488 (0.02, DHK2)	0.1078 (0.03, DHK2)	0.1616 (0.05, DHK2)	0.2123 (0.07, DHK2)	0.2407 (0.10, DHK2)
sampling	0.1149 (0.02, DHK2)	0.1677 (0.02, DHK2)	0.1996 (0.02, DHK2)	0.2357 (0.02, DHK2)	0.3297 (0.02, DHK2)
noise	0.1583 (0.02, DHK1)	0.2316 (0.06, DHK1)	0.2524 (0.09, DHK2)	0.2665 (0.12, SHK1)	0.2809 (0.13, SHK2)
shotnoise	0.0045 (0.02, DHK2)	0.015 (0.02, DHK2)	0.0367 (0.02, DHK2)	0.0707 (0.02, DHK2)	0.1101 (0.02, DHK2)

TABLE IV

LOCAL DEPTH SIFT ROBUSTNESS (0 IS FULL ROBUSTNESS), TESTED ON THE “SHREC’10: ROBUST FEATURE DETECTION AND DESCRIPTION BENCHMARK” DATASET. WE SHOW IN PARENTHESES THE LEADING RESULTS OBTAINED IN THE SHREC’10 BENCHMARK [3]. DHK1 STANDS FOR THE DENSE HEAT KERNEL SIGNATURES, SHK2 IS THE SPARSE HEAT KERNEL SIGNATURE BASED ON FEATURES DETECTED BY HK2 [30]. WE MARKED WITH BOLD THE CATEGORIES WHERE OUR APPROACH LED THE CHALLENGE.

transform level	1	2	3	4	5
Isometry	0.26 (0.15, MH2)	0.52 (0.15, MH2)	0.61 (0.17, MH2)	0.52 (0.17, MH2)	0.57 (0.19, MH2)
Rasterization	0.88 (0.98, MH2)	0.87 (0.98, MH2)	0.87 (0.99, MH2)	0.87 (0.99, MH2)	0.87 (0.99, MH2)
Sampling	0.62 (0.93, MH1)	0.66 (0.95, MH1)	0.70 (0.96, MH1)	0.75 (0.97, MH1)	0.79 (0.98, MH2)
Holes	0.36 (0.21, MH2)	0.39 (0.23, MH2)	0.42 (0.26, MH1)	0.45 (0.29, MH1)	0.47 (0.32, MH1)
Micro holes	0.84 (0.29, MH2)	0.85 (0.30, MH2)	0.86 (0.32, MH2)	0.87 (0.34, MH2)	0.87 (0.35, MH2)
Scaling	0.24 (0.19, MH2)				
Affine	0.60 (0.52, MH1)	0.74 (0.63, MH1)	0.75 (0.68, MH1)	0.76 (0.72, MH1)	0.79 (0.75, MH1)
Noise	0.96 (0.98, MH2)	0.97 (0.99, MH2)	0.98 (0.99, MH2)	0.98 (0.99, MH2)	0.99 (0.99, MH2)
Shot Noise	0.41 (0.22, MH2)	0.48 (0.23, MH2)	0.52 (0.24, MH2)	0.56 (0.25, MH2)	0.60 (0.26, MH2)
Partial	0.78 (0.30, MH1)	0.77 (0.36, MH1)	0.78 (0.37, MH1)	0.69 (0.46, MH1)	0.62 (0.49, MH1)
View	0.50 (0.64, MH1)	0.45 (0.59, MH1)	0.46 (0.61, MH1)	0.46 (0.60, MH1)	0.45 (0.60, MH1)
Average	0.58 (0.50, MH1)	0.63 (0.52, MH2)	0.65 (0.53, MH1)	0.65 (0.55, MH2)	0.66 (0.56, MH2)

TABLE V

SCALE INVARIANT SPIN IMAGE (0 IS FULL ROBUSTNESS), TESTED ON THE “SHREC’11: ROBUST FEATURE DETECTION AND DESCRIPTION BENCHMARK” DATASET. WE SHOW IN PARENTHESES THE LEADING RESULTS OBTAINED IN THE SHREC’11 BENCHMARK [3]. MH1-2 STANDS FOR THE MESH HOG DESCRIPTOR (1-MEAN 2-GAUSSIAN) [42]. WE MARKED WITH BOLD THE CATEGORIES WHERE OUR APPROACH LED THE CHALLENGE.

at least a single valid matching. For the point-to-point matching experiments we set the local scale parameter $C = 3$. Figure 6 shows the matching percentage using the SISI and the LD-SIFT descriptors for the 9 transforms, applied to the “SHREC’10: robust feature detection and description benchmark”.

It is evident that for all transformation, the LD-SIFT either outperforms, or performs on par with the SISI descriptor, making it more suitable to point to point

matching. Hence, we conclude that the difference between the two proposed descriptors (SISI and LD-SIFT) is in the trade-off between sensitivity and specificity. The SISI is better at finding the correct 1NN correspondence. In contrast, if we can allow the true matching to be one of KNN matchings with $5 \leq K \leq 10$ the LD-SIFT is preferred. The LD-SIFT descriptor is suitable when its output is used by a higher level scheme that is able to handle outliers and effectively utilizes the increased

transform level	1	2	3	4	5
Isometry	0.63 (0.15, MH2)	0.56 (0.15, MH2)	0.58 (0.17, MH2)	0.59 (0.17, MH2)	0.59 (0.19, MH2)
Rasterization	0.94 (0.98, MH2)	0.94 (0.98, MH2)	0.94 (0.99, MH2)	0.94 (0.99, MH2)	0.94 (0.99, MH2)
Sampling	0.74 (0.93, MH1)	0.77 (0.95, MH1)	0.81 (0.96, MH1)	0.84 (0.97, MH1)	0.87 (0.98, MH2)
Holes	0.71 (0.21, MH2)	0.73 (0.23, MH2)	0.75 (0.26, MH1)	0.77 (0.29, MH1)	0.78 (0.32, MH1)
Micro holes	0.70 (0.29, MH2)	0.74 (0.30, MH2)	0.76 (0.32, MH2)	0.78 (0.34, MH2)	0.80 (0.35, MH2)
Scaling	0.68 (0.19, MH2)	0.70 (0.19, MH2)	0.72 (0.19, MH2)	0.74 (0.19, MH2)	0.75 (0.19, MH2)
Affine	0.87 (0.52, MH1)	0.88 (0.63, MH1)	0.90 (0.68, MH1)	0.91 (0.72, MH1)	0.93 (0.75, MH1)
Noise	0.95 (0.98, MH2)	0.96 (0.99, MH2)	0.95 (0.99, MH2)	0.98 (0.99, MH2)	0.98 (0.99, MH2)
Shot Noise	0.70 (0.22, MH2)	0.73 (0.23, MH2)	0.76 (0.24, MH2)	0.78 (0.25, MH2)	0.80 (0.26, MH2)
Partial	0.46 (0.30, MH1)	0.47 (0.36, MH1)	0.53 (0.37, MH1)	0.57 (0.46, MH1)	0.59 (0.49, MH1)
View	0.76 (0.64, MH1)	0.74 (0.59, MH1)	0.75 (0.61, MH1)	0.74 (0.60, MH1)	0.74 (0.60, MH1)
Average	0.74 (0.50, MH1)	0.75 (0.52, MH2)	0.77 (0.53, MH1)	0.79 (0.55, MH2)	0.80 (0.56, MH2)

TABLE VI

LOCAL DEPTH SIFT ROBUSTNESS (0 IS FULL ROBUSTNESS), TESTED ON THE “SHREC’11: ROBUST FEATURE DETECTION AND DESCRIPTION BENCHMARK” DATASET. IN PARENTHESES, THE LEADING RESULTS OBTAINED IN THE SHREC’11 BENCHMARK, EXCEPT FOR THOSE SUBMITTED BY US. THE RESULTS ARE QUOTED FROM THE OFFICIAL CHALLENGE RESULTS [2]. MH1-2 = MESH HOG (1 - MEAN 2 - GAUSSIAN) [42]. CATEGORIES IN WHICH OUR PROPOSED APPROACH LED THE BENCHMARK ARE MARKED IN BOLD.

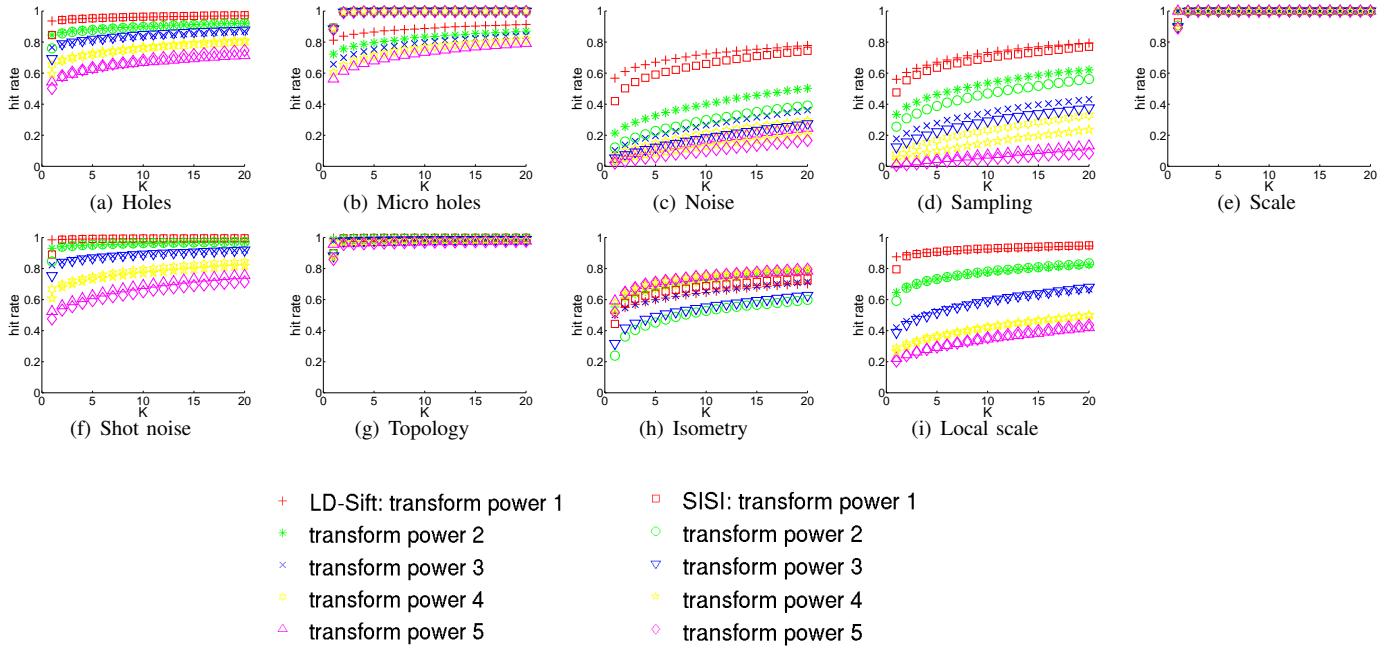


Fig. 6. KNN matching test. We show the matching hit rate vs. K , the number of NN points. The test was conducted using the “SHREC’10: robust feature detection and description benchmark” dataset.

overall number of inlier matchings.

D. Mesh retrieval experiments

We tested the proposed local features on the higher level task of mesh retrieval. For that we implemented the Bag of Words approach as described in Section II-C. In order to construct a dictionary we used the “TOSCA high resolution database” [5] that consists of 80 models that are divided into 9 classes. We extracted feature points, and computed feature descriptors from all the database models. We used this large set of features to

compute a dictionary of 2,500 words using the K-means algorithm. This process was repeated for each of the tests we conducted. In our first retrieval experiment we used the models of the TOSCA database, and the *leave one out* testing approach, where each mesh model in the dataset was used as a query while the remaining models served as the database. A query is considered successful if the best match returned is of the same class as the query model.

We conducted this experiment in two scenarios. In the first, we used the models of the TOSCA dataset in their

original form (in which all models are approximately of the same scale). In the second scenario we applied a random scaling of 0.25 - 4 to each model. In each scenario we tested the SISI, LD-SIFT and standard Spin Image descriptors. The results for this tests are presented in Fig. 7.

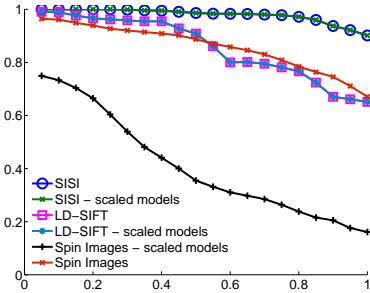


Fig. 7. Mesh retrieval: Precision-recall curve for the TOSCA high resolution database, and its random scaled replica using the SISI, LD-SIFT and Spin Image descriptors.

The SISI and LD-SIFT descriptors gained similar results during the two experiments, validating their scale invariance. The SISI descriptors outperformed the LD-SIFT and the standard Spin Images in the first scenario, while the LD-SIFT outperformed the standard Spin Images. In the second test that is a more challenging scenario, the performance of the non scale-invariant Spin Images degrades even more.

In the second retrieval experiment we used the ‘‘Feature detection and description benchmark’’ database models as queries on the TOSCA database. This database comprises of three transformed models taken from the TOSCA database. Figures 8 and 9 show the results for the SISI and LD-SIFT features with local scale parameter (Eq. 6) of $C = 3$ and $C = 0.3$ respectively. It follows that the proposed schemes are robust to most transforms, showing better results with relatively small feature support ($C = 0.3$). Increasing the feature support by using $C = 3$, makes both our features more robust to noise and sampling changes, as the support over which the descriptor is computed is larger. For low-noise setups the SISI was superior, but when noise was added, the LD-SIFT prevailed.

E. Object matching in cluttered scenes

One of the most challenging applications of local features is object matching in cluttered scenes. Given a high resolution object, and a depth image obtained using a depth sensor, we aim to register an object embedded in the scene. In order to test our method in such a scenario we used the data set provided by Mian *et al.* [25],

which contains five high resolution models and depth scans of cluttered scenes. In each scene part of the five objects are partially shown and some are partially occluded. These scans exhibit realistic attributes such as noise, occlusion and clutter, that differ from the SHREC datasets used before, making them significantly more difficult to process.

Each feature point detected on the high resolution model was matched to the feature point on the scene that had the most similar descriptor, and the RANSAC algorithm [10] was used to register the object to the scene, pruning outliers. The experiment was conducted using both SISI and LD-SIFT descriptors, both resulting in similar high quality registration. The registrations of the *chicken* and *chef* models to a depth scene are shown in Figure 10.

F. Implementation

The proposed scheme was implemented in Matlab with some critical parts coded in C. We used the ‘‘Toolbox Graph’’ of Peyre [32], for general mesh processing and display, and the ‘‘Patch Software Render’’ by Kroon [19] to render the depth images used for the implementation of the LD-SIFT descriptors. Last, we used the VLFeat package [40] by Vedaldi *et al.* to compute the SIFT descriptors. In Table VII we report the timing results of the proposed scheme, when applied to models from the TOSCA database. Our experiments were carried out on a computer using an 2.6GHz Intel i5 processor with 4GBytes of memory. It follows that the proposed detector and descriptor run in reasonable time, where critical components can be accelerated by using a GPU-based implementation.

VI. CONCLUSIONS

In this work we presented a framework for applying the notion of local features to mesh objects. Our work was inspired by state-of-the-art approaches to local features in images. First, we showed how to estimate the local scale of feature points on a mesh using a DoG based detector. Thus, paving the way for deriving scale-invariant mesh features. In our second contribution we formulated two scale invariant descriptors, the first based on an extension of the Spin Images descriptor and the second based on a variation of the SIFT descriptor. For that we derived repeatable local geometrical attributes based on the PCA of the vicinity of the vertex of interest. We applied the proposed framework to the SHREC 10 and SHREC 11 datasets and achieved promising results, compared to contemporary state-of-the-art schemes.

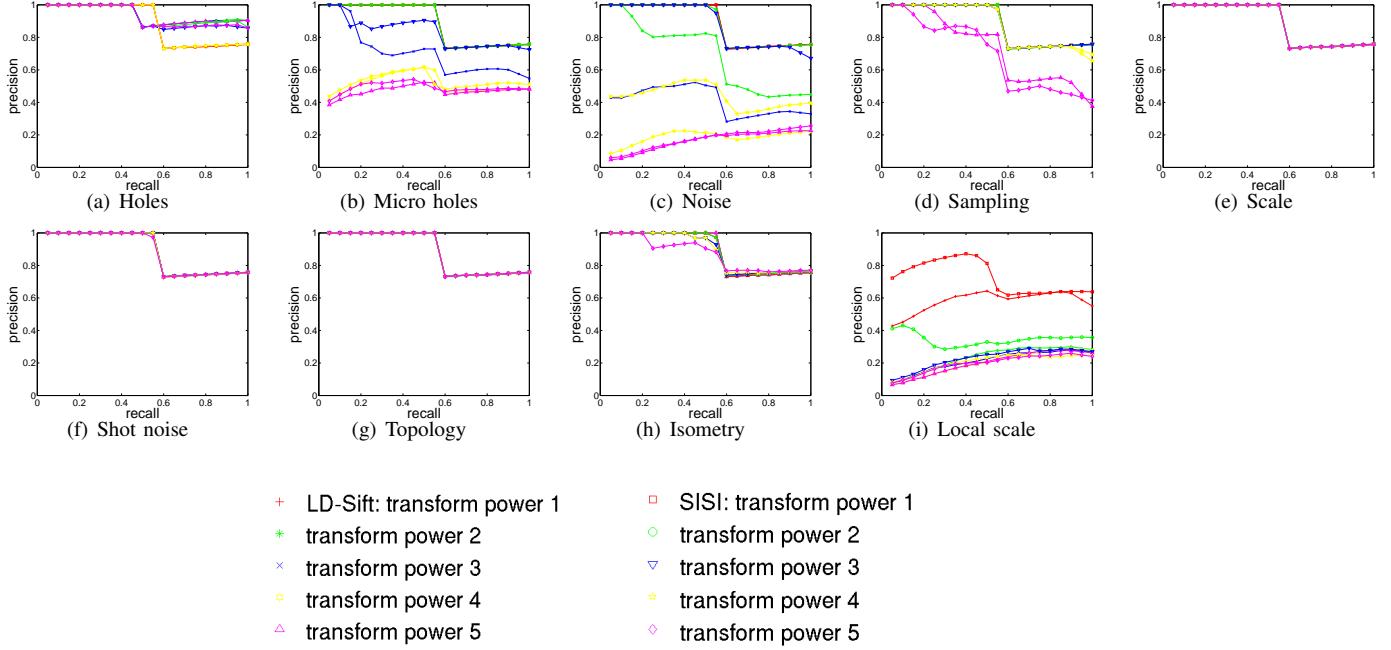


Fig. 8. Precision-recall curves for retrieval of objects under various transforms, using Scale Invariant Spin Image and Local Depth SIFT features with $C = 3$.

Model	Number of vertices	Detector	Detected points	SISI descriptors	LD-SIFT descriptors
Centaur0	15,768	0.25 sec	1,403	0.93 sec	10.81 sec
Cat0	27,894	0.35 sec	2,167	3.23 sec	23.97 sec
david0	52,565	0.71 sec	3,975	9.42 sec	71.21 sec

TABLE VII
RUNNING TIMES OF THE PROPOSED METHODS.

REFERENCES

- [1] O. K.-C. Au, C.-L. Tai, D. Cohen-Or, Y. Zheng, and H. Fu. Electors voting for fast automatic shape correspondence. In *Computer Graphics Forum (In Proc. of Eurographics 2010)*, volume 29, 2010.
- [2] E. Boyer, A. M. Bronstein, M. M. Bronstein, B. Bustos, T. Darom, R. Horaud, I. Hotz, Y. Keller, J. Keustermans, A. Kovnatsky, R. Litman, J. Reininghaus, I. Sipiran, D. Smeets, P. Suetens, D. Vandermeulen, A. Zaharescu, and V. Zobel. Shrec 2011: robust feature detection and description benchmark. In *Eurographics Workshop on 3D Object Retrieval*, 2011.
- [3] A. M. Bronstein, M. M. Bronstein, B. Bustos, U. Castellani, M. Crisani, B. Falcidieno, L. J. Guibas, I. Kokkinos, V. Murino, M. Ovsjanikov, G. Patanè, I. Sipiran, M. Spagnuolo, and J. Sun. Shrec 2010: robust feature detection and description benchmark. In *EUROGRAPHICS Workshop on 3D Object Retrieval (3DOR)*, 2010.
- [4] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Expression-invariant 3D face recognition. *Audio- and Video-based Biometric Person Authentication (AVBPA), Lecture Notes in Comp. Science No. 2688*, 2688:62–70, 2003.
- [5] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. *Numerical geometry of non-rigid shapes*. Springer, 2008.
- [6] M. M. Bronstein and I. Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *CVPR*, pages 1704–1711, 2010.
- [7] U. Castellani, M. Cristani, S. Fantoni, and V. Murino. Sparse points matching by combining 3D mesh saliency with statistical descriptors. *Comput. Graph. Forum*, 27(2):643–652, 2008.
- [8] A. Dubrovina and R. Kimmel. Matching shapes by eigendecomposition of the laplace-beltrami operator. In *3DPVT*, 2010.
- [9] J. Fehr, A. Streicher, and H. Burkhardt. A bag of features approach for 3D shape retrieval. In *Proceedings of the 5th International Symposium on Advances in Visual Computing: Part I, ISVC ’09*, pages 34–43, Berlin, Heidelberg, 2009. Springer-Verlag.
- [10] M. A. Fischler and R. C. Bolles. Readings in computer vision: issues, problems, principles, and paradigms. chapter Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, pages 726–740. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987.
- [11] G. Flitton, T. Breckon, and N. Megherbi Bouallagu. Object recognition using 3D sift in complex 3D volumes. *Proceedings of the British Machine Vision Conference 2010*, pages 11.1–11.12, 2010.
- [12] T. Furuya and R. Ohbuchi. Dense sampling and fast encoding for 3D model retrieval using bag-of-visual features. In *Proceedings of the ACM International Conference on Image and Video Retrieval, CIVR ’09*, pages 26:1–26:8, New York, NY, USA, 2009. ACM.
- [13] K. Gébal, J. A. Bærentzen, H. Aanæs, and R. Larsen. Shape

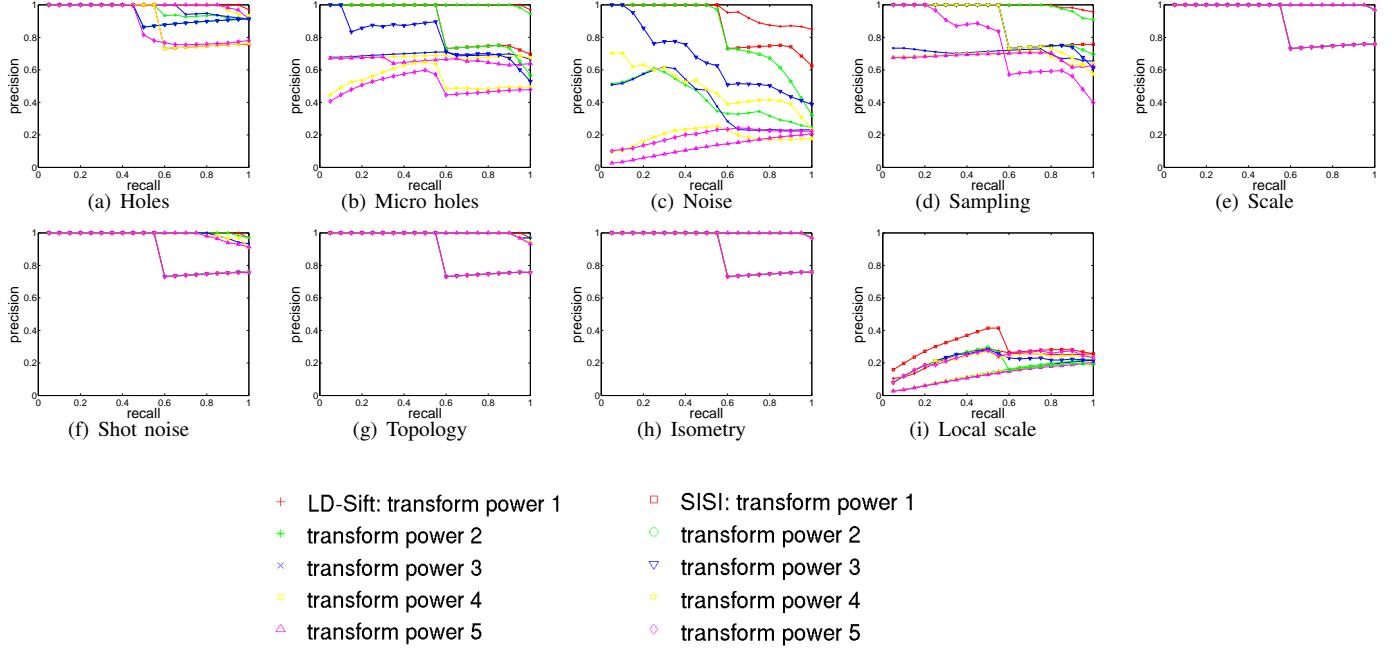


Fig. 9. Precision-recall curves for retrieval of objects undergoing various transforms, using Scale Invariant Spin Image and Local Depth SIFT features with $C = 0.3$.

- analysis using the auto diffusion function. *Comput. Graph. Forum*, 28(5):1405–1413, 2009.
- [14] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988.
 - [15] A. E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(5):433–449, 1999.
 - [16] M. M. Kazhdan, P. D. Simari, T. McNutt, B. Wu, R. Jacques, M. Chuang, and R. Taylor. A shape relationship descriptor for radiation therapy planning. In *MICCAI(1)*, pages 100–108, 2009.
 - [17] J. Knopp, M. Prasad, and L. Van Gool. Orientation invariant 3D object classification using hough transform based methods. In *ACM Workshop on 3D Object Retrieval*, pages 15–20, 2010.
 - [18] J. Knopp, M. Prasad, G. Willem, R. Timofte, and L. J. V. Gool. Hough transform and 3D surf for robust three dimensional classification. In *ECCV (6)*, pages 589–602, 2010.
 - [19] D.-J. Kroon. Patch software render. <http://www.mathworks.com/matlabcentral/fileexchange/27084-patch-software-render>.
 - [20] R. Litman, A. M. Bronstein, and M. M. Bronstein. Diffusion-geometric maximally stable component detection in deformable shapes. *Comput. Graph.*, 35:549–560, June 2011.
 - [21] Y. Liu, H. Zha, and H. Qin. Shape topics: A compact representation and new algorithms for 3D partial shape retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006.
 - [22] Y.-S. Liu, Y. Fang, and K. Ramani. Idss: deformation invariant signatures for molecular shape comparison. *BMC Bioinformatics*, 10, 2009.
 - [23] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
 - [24] C. Maes, T. Fabry, J. Keustermans, D. Smeets, P. Suetens, and D. Vandermeulen. Feature detection on 3D face surfaces for pose normalisation and recognition. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1–6, sept. 2010.
 - [25] A. Mian, M. Bennamoun, and R. Owens. On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes. *International Journal of Computer Vision*, 89(2-3):348–361, 9 2010.
 - [26] N. J. Mitra, L. J. Guibas, J. Giesen, and M. Pauly. Probabilistic fingerprints for shapes. In *Symposium on Geometry Processing*, pages 121–130, 2006.
 - [27] B. Nadler. Finite sample approximation results for principal component analysis: A matrix perturbation approach. *Annals of Statistics*, 36(6):2791–2817, 2008.
 - [28] J. Novatnack and K. Nishino. Scale-dependent / invariant local 3D shape descriptors for fully automatic registration of multiple sets of range images. In *European Conference on Computer Vision*, volume 5304, pages 440–453, 2008.
 - [29] R. Ohbuchii, K. Osada, T. Furuya, and T. Banno. Salient local visual features for shape-based 3D model retrieval. In *Shape Modeling International*, pages 93–102, 2008.
 - [30] M. Ovjanikov, A. M. Bronstein, M. M. Bronstein, and L. J. Guibas. Shape google: geometric words and expressions for invariant shape retrieval. *ACM Transactions on Graphics*, 28(4), 8 2009.
 - [31] E. Paquet, A. Murching, T. Naveen, A. Tabatabai, , and M. Rioux. Description of shape information for 2-d and 3-d objects. *Signal Processing: Image Communication*, (16):103–122, 2000.
 - [32] G. Peyre. Toolbox graph. <http://www.mathworks.com/matlabcentral/fileexchange/5355-toolbox-graph>.
 - [33] M. R. Ruggeri, G. Patanè, M. Spagnuolo, and D. Saupe. Spectral-driven isometry-invariant matching of 3D shapes. *International Journal of Computer Vision*, 89(2-3):248–265, 9 2010.
 - [34] P. Scovanner, S. Ali, and M. Shah. A 3-dimensional sift descriptor and its application to action recognition. *Proceedings of the 15th international conference on Multimedia MULTIMEDIA 07*, pages 357–360, 2007.

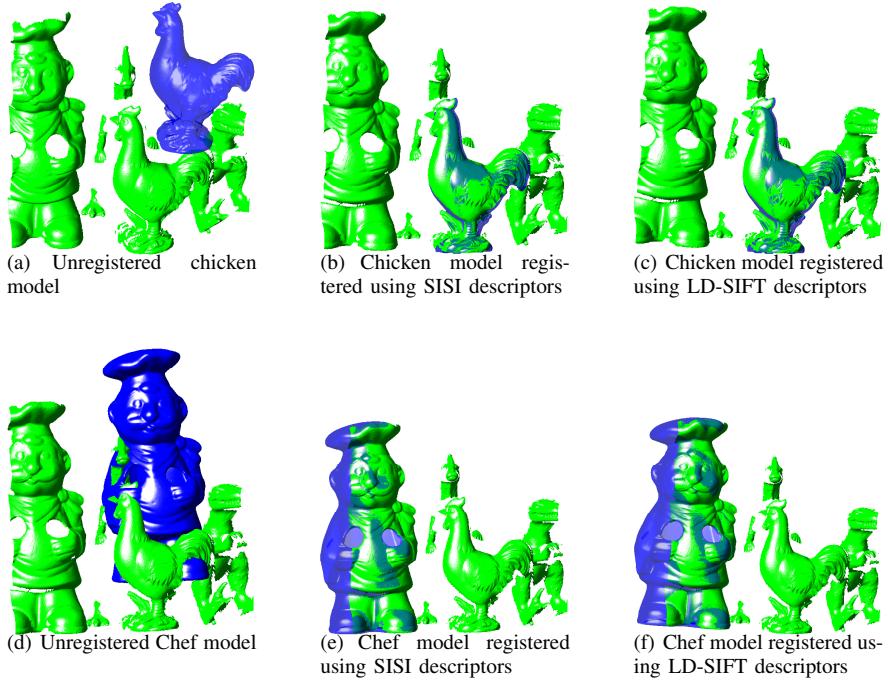


Fig. 10. Registration of models to noisy range images. (a) and (d) show an overlay of both models (in blue) and the cluttered scene. (b) and (e) show the registration using SISI descriptors and RANSAC. Matched vertices are overlayed in blue over the green range image. (c) and (f) depict the registration using LD-SIFT and RANSAC.

- [35] I. Sipiran and B. Bustos. A robust 3D interest points detector based on harris operator. In *In Proc. Eurographics Workshop on 3D Object Retrieval (2010)*, Eurographics Association, pages 7–14, 2010.
- [36] J. Sivic and A. Zisserman. Video Google: Efficient visual search of videos. In J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, editors, *Toward Category-Level Object Recognition*, volume 4170 of *LNCS*, pages 127–144. Springer, 2006.
- [37] J. Sun, M. Ovsjanikov, and L. J. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *Comput. Graph. Forum*, 28(5):1383–1392, 2009.
- [38] R. Toldo, U. Castellani, and A. Fusello. Visual vocabulary signature for 3D object retrieval and partial matching. In *Eurographics Workshop on 3D Object Retrieval*, 2009.
- [39] R. Unnikrishnan, J.-F. Lalonde, N. Vandapel, and M. Hebert. Scale selection for the analysis of point-sampled curves. In *Third International Symposium on 3D Data Processing Visualization and Transmission 3DPVT06*, pages 1026–1033, 2006.
- [40] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. <http://www.vlfeat.org/>, 2008.
- [41] R. Wessel, R. Baranowski, and R. Klein. Learning distinctive local object characteristics for 3d shape retrieval. In O. Deussen, D. Keim, and D. Saupe, editors, *Vision, Modeling, and Visualization 2008 (VMV 2008)*, pages 167–178. Akademische Verlagsgesellschaft Aka GmbH, Heidelberg, Oct. 2008.
- [42] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud. Surface feature detection and description with applications to mesh matching. In *CVPR*, pages 373–380, 2009.
- [43] C. Zhang and T. Chen. Efficient feature extraction for 2d/3d objects in mesh representation. In *ICIP*, pages 935–938, 2001.



Tal Darom received the B.Sc. degree in electrical engineering in 1998 from The Technion-Israel Institute of Technology, Haifa, Israel, and the M.Sc degree in electrical engineering from Tel-Aviv University, Tel-Aviv, Israel, in 2006. From 1998 to 2010 he held several algorithm and software development positions in the fields of computer vision, computer graphics and computational geometry. He is currently a Ph.D student in electrical engineering at the Faculty of Engineering, Bar-Ilan University, Ramat-Gan, Israel.



Yosi Keller received the BSc degree in Electrical Engineering in 1994 from the Technion-Israel Institute of Technology, Haifa. He received the MSc and PhD degrees in electrical engineering from Tel-Aviv University, Tel-Aviv, in 1998 and 2003, respectively. From 2003 to 2006 he was a Gibbs assistant professor with the Department of Mathematics, Yale University. He is a senior lecturer at the Faculty of Engineering in Bar Ilan University, Israel.