Keypoints and Local Descriptors of Scalar Functions on 2D Manifolds

Andrei Zaharescu · Edmond Boyer · Radu Horaud

Received: 18 June 2011 / Accepted: 15 April 2012 / Published online: 4 May 2012

Abstract This paper addresses the problem of describing surfaces using local features and descriptors. While methods for the detection of interest points in images and their description based on local image features are very well understood, their extension to discrete manifolds has not been well investigated. We provide a methodological framework for analyzing real-valued functions defined over a 2D manifold, embedded in the 3D Euclidean space, e.g., photometric information, local curvature, etc. Our work is motivated by recent advancements in multiple-camera reconstruction and image-based rendering of 3D objects: there is a growing need for describing object surfaces, matching two surfaces, or tracking them over time. Considering polygonal meshes, we propose a new methodological framework for the scale-space representations of scalar functions defined over such meshes. We propose a local feature detector (MeshDOG) and region descriptor (MeshHOG). Unlike the standard image features, the proposed surface features capture both the local geometry of the underlying manifold and the scale-space differential properties of the real-valued function itself. We provide a thorough experimental evaluation. The repeatability of the feature detector and the robustness of feature descriptor are tested, by applying a large number of deformations to the manifold or to the scalar function.

A. Zaharescu (✉)
Aimetis Corporation, Waterloo, Ontario, Canada
e-mail: Andrei.Zaharescu@aimetis.com

E. Boyer · R. Horaud
INRIA Grenoble Rhône-Alpes, Montbonnot Saint-Martin, France

E. Boyer
e-mail: Edmond.Boyer@inria.fr

R. Horaud
e-mail: Radu.Horaud@inria.fr

Keywords 3D shape · Meshed surfaces · Riemannian manifold · Scale space · 3D keypoint detection · Local shape descriptors · Shape matching

1 Introduction

The representation of visual information in terms of a structured collection of local features has been an active research topic for the last decade and it is of great importance for a variety of tasks, such as tracking, registration, recognition, retrieval, etc. Feature-based approaches were introduced in the computer vision literature three decades ago (Bolles and Cain 1982; Bolles and Horaud 1986) for the purpose of recognizing and localizing partially occluded objects. Initially, features represented local geometric information. More recently, feature-based image analysis has become very popular (Lowe 2004; Mikolajczyk and Schmid 2005). The vast majority of existing methods detect and describe features using photometric information from a single image. Recently, image features were extended to 2D$+t$ features, used for characterizing short image sequences for video analysis (Laptev 2005).

Recent progress in multiple-view stereo and image-based modelling and rendering allows the recovery of geometric *and* photometric information directly from images (Seitz et al. 2006). This means that one can characterize 3D shapes based on both geometric and photometric features. However, if taken separately, geometric or photometric information have limited utility, as the whole richness of the data is not fully exploited. Consider, for example, 3D deformable or articulated objects. Their 2D appearances, i.e., images, are not full invariant to motions and to viewing conditions and hence, image-based features do not yield robust 3D descriptors. Similarly, geometric features are not robust to complex object motions that can considerably change the topology. Therefore, we believe that photometric and geometric information need to be handled in a consistent and simultaneous manner. To this purpose, we observe that both photometric and geometric information available with a 3D object can both be viewed as scalar (real-valued) functions defined over a 2D manifold, e.g., the surface of an object; This may well be considered as a generalization of scalar functions defined over an Euclidean domain, e.g., light intensity over an image, to non-Euclidean domains. One can thus build on existing feature-based image description paradigms, in order to investigate and propose extensions to 3D shapes.

The main contribution of this paper is a novel family of keypoint-based local surface descriptors that takes into account both the geometric properties of the surface and any scalar field defined over the surface, e.g. the photometric information available from multiple-camera setups. We develop an interest point detector and associated local scale-space descriptors that can be applied to various function defined over the surface, e.g, texture, colour, Gaussian curvature, mean curvature, geodesic integral, etc. To this end, we use discrete operators, e.g., the gradient defined for scalar functions on discrete surface domains (meshes), thus taking into account both the functions' differential properties as well as the underlying intrinsic geometry of the surface. Based on these operators, both an interest point detector and a local descriptor are introduced, namely MeshDOG and MeshHOG. MeshDOG is a generalization of the Difference of Gaussian (DOG) operator (Marr and Hildreth 1980; Lowe 2004) and it is used to build a discrete Laplacian operator on a mesh. This allows us to represent scalar functions over multiple scales, i.e., by convolution with a discrete Laplacian operator, and to detect points of interest as local extrema. MeshHOG is a generalization of the Histogram of Oriented Gradients (HOG) descriptor that was proposed for describing 2D images (Dalal and Triggs 2005). The newly descriptor, MeshHOG, is defined with respect to the measurements available at each vertex of the discrete surface and it can be implemented with any scalar function.

The newly proposed surface-based interest point detector and region descriptor exhibit a number of interesting properties with respect to the corresponding image-based counterparts:

1. A 3D detector takes more information into account than an image based one and it does not suffer from false detections due to occluding contours;
2. There are no affine (nor perspective) distortions, since the computations are performed in a local metric space. Hence, there is no need for affine-invariance;
3. Image descriptors are sensitive to 2D occlusions. This is not an issue with surface descriptors, provided that a complete reconstruction of the underlying 3D object is available, which is typically the case with reconstructions from multiple camera setups;
4. The descriptor captures both the (local) 3D geometry and the gradient information associated with the scalar function;
5. In a multiple-camera setting, photometric information can be elegantly fused together from several images. This provides a photometric descriptor that is both image-invariant and more robust to image noise.

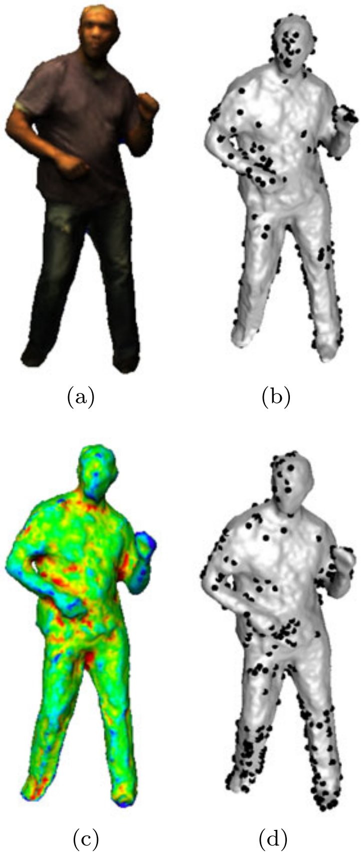


Fig. 1 The feature detection method described in this paper can be applied to any scalar function defined over a 2D manifold such as the meshed surface shown here: photometric data (**a**) and associated points of interest (**b**); mean surface curvature (**c**) and the detected features (**d**)

The remainder of the paper is organized as follows. Section 2 discusses related work and it emphasizes our own contribution to the problem of extracting local surface features. Section 3 introduces the mathematical formulation used to define the gradient and hessian operators on discrete manifolds that are needed to build local features. Sections 4 and 5 introduce the local feature detector and region descriptor, respectively. Section 6 presents a detailed performance evaluation and comparison with other methods, before concluding in Sect. 7.

2 Related Work

In the recent past, there has been a lot of work aimed at visually characterizing 3D objects for the purposes of modelling and recognition.

2.1 Surface Features Based on Image Processing

One possible approach is to rely on image keypoints with their associated local descriptors and to use them in order to build 3D surface features.

Detectors Keypoints are often associated with the detection of interest points in images, such as the extrema of the Laplacian of the image intensity function. They can be detected at various scales using the difference of Gaussian (DOG) approximation of the Laplacian (see Mikolajczyk and Schmid 2005 for a review and the references therein).

Descriptors 2D feature descriptors are generally designed to be robust to changes in illumination and invariant to image transformations such as translation, rotation, or scale (Matas et al. 2004; Lowe 2004; Dufournaud et al. 2004; Dalal and Triggs 2005; Bay et al. 2008) and, more generally, to 2D affine transformations (Mikolajczyk and Schmid 2004). This type of image-based descriptors have been successfully used to characterize 3D objects (Rothganger et al. 2006). We note, however, that there are inherent limitations with these approaches. First, it is required that the image descriptors are back projected onto the 3D surface of the object which may lead to a redundant and ambiguous representation, since a 3D surface point corresponds to several points belonging to different images. Second, these descriptors are well suited only for objects which are locally planar. Third, the image-based descriptors are limited to photometric information and hence, one cannot build 3D descriptors based on the geometric properties of the underlying surface, e.g., curvature. Our method is quite different because it directly exploits the local geometric properties of the object's underlying surface.

Image keypoints and associated local descriptors were extended to image sequences (Laptev 2005; Wong and Cipolla 2007; Kläser et al. 2008), by considering the $2D + t$ (spatio-temporal) volume defined by a short image sequence. Such *image* + *time* features can be seen as local detectors/descriptors defined over a volumetric representation, i.e., a regular 3D grid. We consider a different problem, namely the extension from features defined on regular domains to features defined over irregular non-Euclidean domains.

2.2 Surface Features Based on Local Geometric Properties

Another category of methods attempts to extract local geometric information from range images or from point-cloud data.

Detectors Novatnack and Nishino (2007) defines the scale space in a planar parameterization of the surface using the normal map; 3D keypoints are detected as the extrema of this representations, based on a gradient operator defined over a planar vector field. An analysis of the scale-variability of geometric structures captured in range images is proposed in Novatnack and Nishino (2008), while Bariya and Nishino (2010) extends this method to deal with cluttered 3D scenes in an object recognition task. The automatic identification of interest regions on surfaces, taking into account geometric features such as scale-space extrema based on the average mean curvature flow, is proposed in Schlattmann et al. (2008). In Mian et al. (2010), it is proposed to detect keypoints that characterize 3D and 2.5D surfaces. Mesh saliency methods are proposed in Lee et al. (2005), Castellani et al. (2008), based on the centre-surround operator, adapted from the visual attention literature.

Descriptors A number of 3D descriptors were proposed. *3D spin images* (Johnson and Hebert 1999), one of the first proposed methods, build 2D histograms by accumulating points that fall on a rotating plane along the normal. *3D shape contexts* (Körtgen et al. 2003; Frome et al. 2004) extend the idea of spin images, by accumulating 3D histograms within a spherical support region. For a detailed survey, see Tangelder and Veltkamp (2004), Bustos et al. (2005). *Intrinsic shape signatures* are proposed in Zhong (2009), thus improving on shape contexts by using a different histogram partitioning scheme. In Novatnack and Nishino (2008), an image-based descriptor is proposed using the local $\mathbb{R}^2$ embedding of the surface normal information. Similarly, Hua et al. (2008) builds 2D conformal maps of a 3D shape by mapping an irregular domain, such as a mesh, to a regular grid. In the resulting *shape vector image*, standard 2D scale-space keypoint detection and description (Lowe 2004) can be applied to build 3D features. Such features are robust to isomorphic deformations. However, due to conformal mapping limitations, they are sensitive to even small topological changes. More recently, Ruggeri et al. (2010) proposed to build local 3D descriptors of a meshed surface at the locations of the critical points, the maxima, minima and saddle points of the eigenfunctions of the Beltrami-Laplace operator. Similarly, Bronstein et al. (2011) proposed to use the critical points of the autodiffusion function, the diagonal elements of the heat-kernel matrix of a mesh, in order to obtain an intrinsic scale-space representation of the mesh's geometry. These methods allow to describe an object in terms of its intrinsic local geometry, but they do not allow to characterize other scalar functions. Again, these methods cannot exploit photometric information.

2.3 Combining Photometric Information with 3D Features

Closer to our own methodology, a category of methods attempts to characterize the photometric information that is available when considering 3D objects obtained from multiple-view reconstructions. Starck and Hilton (2007) proposes a concatenated surface descriptor, encompassing local geometry (a region descriptor based on geodesic-intensity histograms) and photometric information, edge and corner descriptors that take into account the local isometric mapping to $\mathbb{R}^2$; In Wu et al. (2008) a SIFT-based descriptor using 3D oriented patches is proposed, namely VIP, or viewpoint-invariant patches, which was used for 3D model matching. Both Starck and Hilton (2007) and Wu et al. (2008) are among the first attempts to devise a descriptor that combines geometric and photometric information. Our proposed approach is similar in spirit to Wu et al. (2008), but instead of back-projecting an image descriptor onto the surface, we propose to detect keypoints and build an associated descriptor directly onto the surface, taking full advantage of the 3D nature of the surface.

Recently, a number of extensions were proposed to our previous work (Zaharescu et al. 2009). Smith et al. (2011) consider the specific scenario when the 3D triangulated mesh and the scalar function are from single range-intensity image. The discrete gradient approximation, inspired from Xu (2004), requires the scalar function to be defined over the faces of a triangulated mesh. This may be difficult and expensive to retrieve from the initial data, or it may not be available at all. Alternatively, we propose a radically different way of approximating gradient computation on any polygonal mesh that only requires that the scalar function is defined at the mesh vertices. This allows us to compute not only gradients of functions but also gradients of directional derivatives of these functions, which are needed for keypoint detection.

2.4 Registration and Recognition

Interestingly, many applications of 3D modelling make use of local features, e.g., rigid and non-rigid registration, object recognition, shape retrieval, etc. Recent work (Furukawa and Ponce 2008; Ahmed et al. 2008; de Aguiar et al. 2007; Varanasi et al. 2008) addressed non-rigid mesh registration using observations from multiple views. The vast majority of the proposed methods (one notable exception being Furukawa and Ponce 2008) use both geometric information extracted from surfaces and photometric data available from images. The latter is first extracted using 2D image descriptors (such as SIFT), and subsequently back-projected onto the mesh. This sparse description is generally used to bootstrap dense matching. Surface descriptors may well be used for 3D object recognition, as it has been already done in Shilane et al. (2008), using the Princeton shape benchmarking database.[1] Our work contributes to these efforts by taking a different, yet complementary approach: image-feature detection and description methodologies are extended to features defined directly onto discrete 2D manifolds.

2.5 Paper Contributions

This paper is an extended version of Zaharescu et al. (2009), which introduced 3D shape descriptors inspired from image descriptors. Unlike an image, which is a regular Euclidean domain, a 3D shape is often defined over an irregular non-Euclidean domain such as a mesh, which may be viewed as a discrete manifold. In Zaharescu et al. (2009) we proposed a surface keypoint detection based on a difference of Gaussian operator (MeshDOG) and a local descriptor based on the histogram of oriented gradients (MeshHOG). Both MeshDOG and MeshHOG require the estimation of a gradient operator and of first- and second-order directional derivative operators. In this paper we propose an improved computational framework for estimating the gradient and the directional derivatives of real-valued functions defined on discrete manifolds, e.g., a mesh. With respect to Zaharescu et al. (2009), we relax the constraint that the mesh vertices must correspond to a regular sampling of the underlying continuous surface; We do this by using the geodesic distance throughout the formulation. Also, the proposed gradient computation is now cast into a least square minimization problem that can be efficiently estimated using a linear solver. The gradient method is inspired from Barth (1993), Mukherjee et al. (2010) and it could, in principle, be applied to any kind of polygonal mesh or to a point cloud. It differs from previous approaches that are based on the eigenfunctions of a discrete Laplace-Beltrami operator, e.g., Xu (2004), Luo et al. (2009). Additionally, we introduce a new dataset as well as an in-depth evaluation of our method, thus testing both the repeatability of the keypoint detector and the robustness of the region descriptor under a large number of deformations. Finally, two other existing datasets are used for comparisons with other existing methods.

3 Differential Mesh Processing

In this section we introduce a computational framework required to estimate interest points and local descriptors of a scalar function defined over a manifold. To this end, we define several operators that can handle an irregular domain, including the gradient operator and the first- and second-order directional derivative operators.

Let $\mathcal{M}$ be a 2D closed manifold (i.e. compact and without boundaries) embedded in $\mathbb{R}^3$ and let M be a discrete

[1] http://shape.cs.princeton.edu/benchmark/.

mesh representation of $\mathcal{M}$ composed of vertices on $\mathcal{M}$ and of convex polygons, i.e. facets. M can be viewed as a graph $M(V, E)$, where $V = \{v_i\}_{i=1}^N$ is the set of mesh vertices and $E = \{e_{ij}\}$ is the set of edges between adjacent vertices. We associate a 3D point $\mathbf{v}_i \in \mathbb{R}^3$ with each mesh vertex v_i. Note that an image can be viewed as a "flat" uniformly sampled mesh *with boundaries*, i.e., a grid of vertices with valence 4 and whose facets are rectangles.

3.1 Gradient

Let $f : \mathcal{M} \to \mathbb{R}$ be a smooth real-valued function defined on $\mathcal{M}$, e.g., photometric data or curvature. In order to estimate the *gradient* $\nabla_{\mathcal{M}} f(\mathbf{v})$ of f at point $\mathbf{v}$, we consider the first order Taylor expansion approximating f at a manifold point $\mathbf{v}_i$ in a neighborhood of point $\mathbf{v}_j$:

$$f(\mathbf{v}_i) \approx f(\mathbf{v}_j) + \nabla_{\mathcal{M}} f(\mathbf{v}_i)^\top (\mathbf{v}_i - \mathbf{v}_j). \quad (1)$$

where the gradient Δ belongs, by definition, to the tangent plane of $\mathcal{M}$ at v_i. In the discrete case one can therefore write:

$$\nabla_M f(\mathbf{v}_i)^\top (\mathbf{v}_i - \mathbf{v}_j) \approx f(\mathbf{v}_i) - f(\mathbf{v}_j), \quad (2)$$

where $\nabla_M f(\mathbf{v})$ denotes the *discrete gradient* of f at $\mathbf{v}$. This expression can be used to estimate the discrete gradient at any mesh vertex $\mathbf{v}_i$ through an error minimization criterion (Sibson 1981). We adopt the least square gradient construction that follows this principle (Barth 1993) and we seek the 3D vector that minimizes the criterion:

$$\nabla_M f(\mathbf{v}_i) = \underset{\mathbf{g}}{\operatorname{argmin}} \left\{ \sum_{v_j \sim v_i} w_{ij} \big(f(\mathbf{v}_i) - f(\mathbf{v}_j) - \mathbf{g}^\top (\mathbf{v}_i - \mathbf{v}_j) \big)^2 \right\}, \quad (3)$$

where the notation $v_j \sim v_i$ means that $v_j \in \mathcal{N}(\mathbf{v}_i)$, i.e., the neighbourhood of $\mathbf{v}_i$ considered in the estimation and where the weights w_{ij} balance the contributions of the neighboring vertices. Both $\mathcal{N}(\mathbf{v}_i)$ and w_{ij} are chosen as follows.

$\mathcal{N}(\mathbf{v}_i)$ is usually the first ring of vertices around $\mathbf{v}_i$. However, in order to make it more robust to non-uniform sampling, $\mathcal{N}(\mathbf{v}_i)$ can be defined as the set of vertices $\mathbf{v}_j \in M$ residing within a geodesic ball centred in $\mathbf{v}_i$ of radius r:

$$\mathcal{N}(\mathbf{v}_i) = \{\mathbf{v}_j | d_g(\mathbf{v}_i, \mathbf{v}_j) < r\}, \quad (4)$$

where $d_g(\mathbf{v}_i, \mathbf{v}_j)$ represents the geodesic distance between $\mathbf{v}_i$ and $\mathbf{v}_j$.

The weight function w_{ij} can be uniform or it can vary with respect to, e.g., areas (Sibson 1981) or inverse distances (Barth 1993) in the neighbourhood of $\mathbf{v}_i$. In Mavriplis (2003) it is shown that weighted gradient estimations based on inverse distances significantly improve over unweighted estimations. In this work, the weight function is a zero-centred Gaussian function:

$$w_{ij} = G_\sigma\big(d_g(\mathbf{v}_i, \mathbf{v}_j)\big) = \exp\big(-d_g^2(\mathbf{v}_i, \mathbf{v}_j)/2\sigma^2\big). \quad (5)$$

Note that vector $\mathbf{g}$ in (3) can be advantageously constrained to belong to the tangent plane of M at $\mathbf{v}_i$, whenever the normal unit vector $\mathbf{n}_i$ to this tangent plane is known:

$$\nabla_M f(\mathbf{v}_i) = \underset{\mathbf{g}}{\operatorname{argmin}} \left\{ \sum_{v_j \sim v_i} w_{ij} \big(f(\mathbf{v}_i) - f(\mathbf{v}_j) - \mathbf{g}^\top (\mathbf{v}_i - \mathbf{v}_j) \big)^2 + \lambda_i \big(\mathbf{g}^\top \mathbf{n}_i\big)^2 \right\}, \quad (6)$$

where the positive scalar λ_i is chosen such that the tangent-plane constraint is emphasized:

$$\lambda_i = \sum_{v_j \sim v_i} w_{ij}.$$

The constrained minimization (6) is a linear least-squares problem that is efficiently solved using standard matrix factorization methods, such as singular value decomposition (Lay 1996).

The current gradient computation formalism is better motivated mathematically than our previously proposed approximation (Zaharescu et al. 2009). Even though the current formulation handles better particular edge cases, both methods behave numerically similar on the average, when dealing with evenly sampled manifold discretizations.

3.2 Directional Derivatives

The *directional derivative* $D_{\mathbf{a}} f(\mathbf{v})$ of f at $\mathbf{v} \in \mathcal{M}$ along vector $\mathbf{a}$ is then, by definition, the projection of the gradient vector $\nabla_{\mathcal{M}} f(\mathbf{v})$ along the direction of $\mathbf{a}$:

$$D_{\mathbf{a}} f(\mathbf{v}) = \nabla_{\mathcal{M}} f(\mathbf{v})^\top \frac{\mathbf{a}}{\|\mathbf{a}\|}, \quad (7)$$

where $\mathbf{a}$ is a vector lying in the tangent plane of $\mathcal{M}$ at $\mathbf{v}$. The discrete directional derivative at $\mathbf{v}_i$ along any direction in the tangent plane at $\mathbf{v}_i$ can therefore be computed using (7) with the discrete gradient $\nabla_M f(\mathbf{v}_i)$ estimated using (6).

Let us now consider *second* discrete directional derivatives along unit vectors $\mathbf{a}$ and $\mathbf{b}$ lying in the tangent plane to the mesh at $\mathbf{v}_i$. Such a second directional derivative can be written as:

$$\begin{aligned} D_{\mathbf{ab}} f(\mathbf{v}_i) &= \nabla_M \big(\nabla_M f(\mathbf{v}_i)^\top \mathbf{b}\big)^\top \mathbf{a}, \\ &= D_{\mathbf{a}}\big(D_{\mathbf{b}} f(\mathbf{v}_i)\big), \end{aligned} \quad (8)$$

which requires an estimate of the gradient of the scalar function $D_{\mathbf{b}} f(\mathbf{v}_i)$, namely $\nabla D_{\mathbf{b}} f(\mathbf{v}_i)$. This can be easily obtained by applying the least square criterion (6) to the directional derivative function. Similarly, one can estimate $D_{\mathbf{aa}} f(\mathbf{v}_i)$, $D_{\mathbf{bb}} f(\mathbf{v}_i)$, $D_{\mathbf{ab}} f(\mathbf{v}_i)$ and $D_{\mathbf{ba}} f(\mathbf{v}_i)$. By further assuming that the vectors $\mathbf{a}$ and $\mathbf{b}$ form an orthonormal basis of the tangent plane at $\mathbf{v}_i$, one obtains the Hessian matrix of f:

$$\mathbf{H}_{a,b}\big(f(\mathbf{v}_i)\big) = \begin{bmatrix} D_{\mathbf{aa}} f(\mathbf{v}_i) & D_{\mathbf{ab}} f(\mathbf{v}_i) \\ D_{\mathbf{ba}} f(\mathbf{v}_i) & D_{\mathbf{bb}} f(\mathbf{v}_i) \end{bmatrix}. \tag{9}$$

Assuming that the directional derivatives of f are continuous, the order of the differentiation does not matter, and hence (by Clairaut's theorem), one should expect the Hessian in (9) to be symmetric, namely that $D_{\mathbf{ab}} f(\mathbf{v}_i) = D_{\mathbf{ba}} f(\mathbf{v}_i)$. However, in our case, the gradients $\nabla D_{\mathbf{a}} f(\mathbf{v}_i)$ and $\nabla D_{\mathbf{b}} f(\mathbf{v}_i)$ are obtained by numerical optimization of (6). Hence, the Hessian is not guaranteed to be symmetric. This means that it is not guaranteed that the two eigenvalues of (9) are real. Therefore, we propose to use the following real symmetric matrix:

$$\tilde{\mathbf{H}}_{a,b}\big(f(\mathbf{v}_i)\big) = \frac{1}{2}\big(\mathbf{H}_{a,b}\big(f(\mathbf{v}_i)\big) + \mathbf{H}^{\top}_{a,b}\big(f(\mathbf{v}_i)\big)\big) \tag{10}$$

which corresponds to the projection of $\mathbf{H}$ onto the linear space of 2×2 symmetric matrices (Horn and Johnson 1994).

3.3 Convolution

Finally, using the same notations, the *normalized convolution* of the function f with a Gaussian kernel G yields:

$$\begin{aligned} F_\sigma(\mathbf{v}_i) &= f \star G_\sigma(\mathbf{v}_i) \\ &= \frac{1}{K_i} \sum_{v_j \sim v_i} f(\mathbf{v}_j) \exp\big(-d_g^2(\mathbf{v}_i, \mathbf{v}_j)/2\sigma^2\big) \end{aligned} \tag{11}$$

where G_σ is the Gaussian function defined in (5) and K_i is a normalization term such that:

$$K_i = \sum_{v_j \sim v_i} G_\sigma\big(d_g(\mathbf{v}_i, \mathbf{v}_j)\big).$$

Using the properties of convolution, one can easily compute the first- and second-order directional derivatives of F_σ, namely:

$$D_{\mathbf{a}} F_\sigma(\mathbf{v}_i) = D_{\mathbf{a}} f \star G_\sigma(\mathbf{v}_i), \tag{12}$$

$$D_{\mathbf{ab}} F_\sigma(\mathbf{v}_i) = D_{\mathbf{ab}} f \star G_\sigma(\mathbf{v}_i). \tag{13}$$

3.4 Numerical Approximations

Geodesics The computation of geodesic distances on arbitrarily triangular meshes can be computed by the fast marching method (Kimmel and Sethian 1998) or by other approximations (Surazhsky et al. 2005). In practice, in the interest of computational speed, we have used a local shortest path approximation on the edge connectivity graph. It has been observed experimentally that, for typical meshes, the variations due to the triangulation are minimal.

Normals At first, a local normal estimation is used, using a one ring neighbourhood. In order to increase the robustness of the normal estimation, a smoothed version is then computed, using the mean estimate in a small geodesic neighbourhood.

Curvatures Instead of using the classical curvature estimation method (Meyer et al. 2002) that employs only a one ring neighbourhood, we employ the more robust method proposed in Dong and Wang (2005), using a 3 ring neighbourhood.

4 Feature Detection (MeshDOG)

Feature detection comprises three steps, as illustrated in Fig. 2. First, the extrema of the function's Laplacian are found across scales using a one-ring neighbourhood. The Laplacian is approximated with the standard difference of Gaussian (DOG) operator. Second, the extrema thus detected are thresholded. Finally, the unstable extrema are eliminated, only retaining the features exhibiting some degree of cornerness.

4.1 Scale-Space Construction

A scale-space representation of any scalar function f defined on a mesh is considered, built by progressive convolutions over f. The scale-space is built over $s = 3$ octaves, covering each octave in $c = 6$ steps. This is accomplished by progressive convolutions with a Gaussian kernel (11) of the original scalar function:

$$F_0 = f, \tag{14}$$

$$F_t = F_{t-1} \star G_{\sigma(t)}, \tag{15}$$

with $t = \{1, 2, \ldots, s \cdot c\}$. The standard deviation parameter $\sigma(t)$ of the Gaussian at iteration t is chosen based on the formula:

$$\sigma(t) = 2^{\frac{1}{c-2}\lceil\frac{t}{c}\rceil} e_{avg}, \tag{16}$$

where e_{avg} represents the average edge length. As it can be observed, the value of $\sigma(t)$ only changes when t starts spanning a new octave. $\sigma(t)$ remains unchanged for the next c iterations, while t covers the current octave, thanks to the term $\lceil\frac{t}{c}\rceil$. This behaviour emulates in spirit the 2D grid downsampling, introduced in Lowe (2004), but without modifying the mesh geometry, which can be an expensive

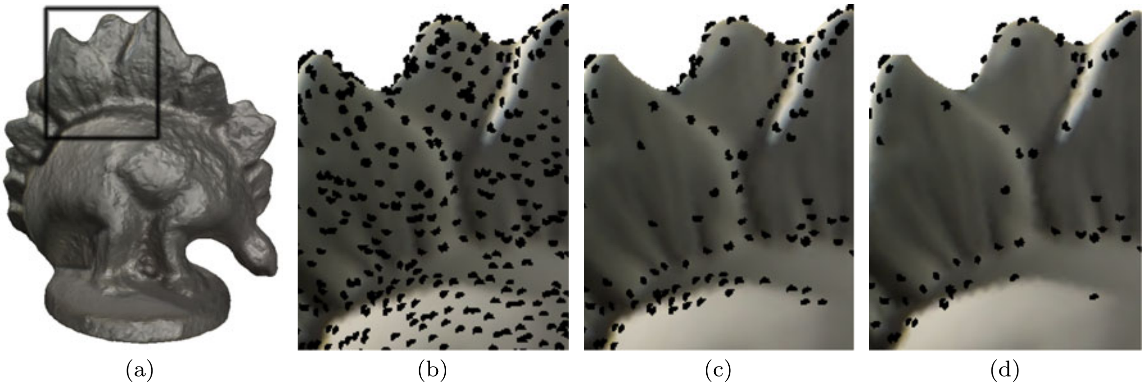


Fig. 2 Feature detection shown with photometric data. (**a**) Original mesh (27240 vertices); (**b**) Scale-space extrema (5760 vertices *left*); (**c**) Thresholding (1360 vertices *left*); (**d**) Corner detection (650 vertices *left*)

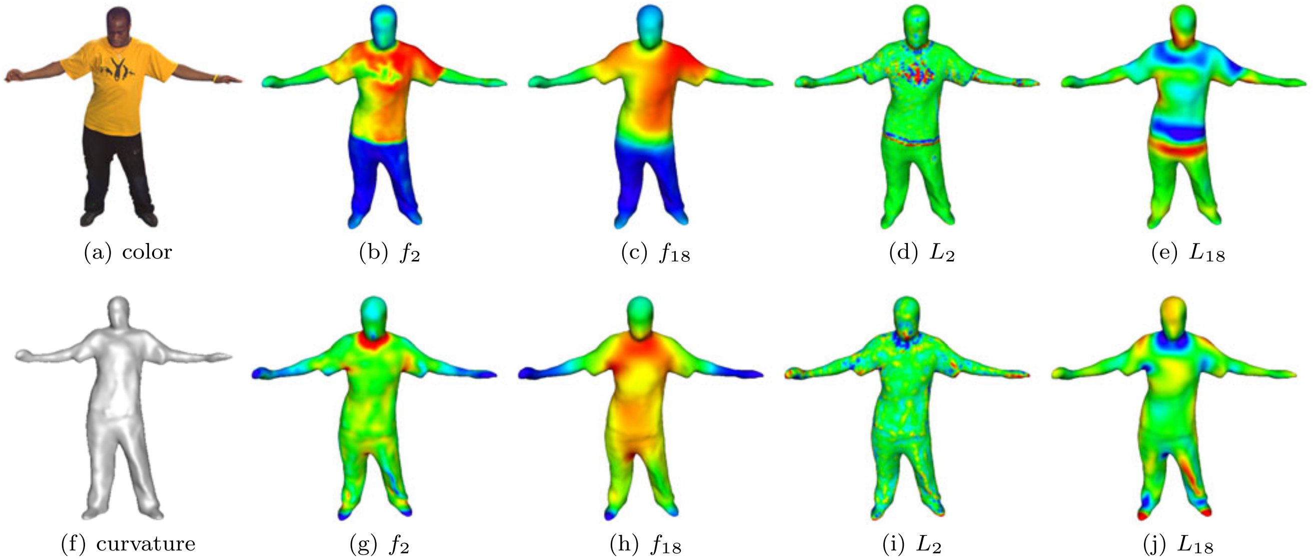


Fig. 3 An example of processing (**a**) color intensity and (**f**) mean curvature. Scale-space representation of color (**b**)–(**e**) and of mean curvature (**g**)–(**j**). The mesh corresponds to frame 30 of the *pop2lock* sequence from the University of Surrey

operation, in the case of non-uniformly sampled meshes. Therefore, an important observation is that, when building the scale space of scalar function defined over the mesh, *the mesh geometry does not change*. This contrasts to other approaches, such as Hou and Qin (2010), that construct the scale-space by generating meshes with different samplings, thus requiring further mesh processing and simplification.

The difference of Gaussian operator is then used as an approximation of the Laplacian operator, built by subtracting adjacent convolved functions:

$$L_t = F_t - F_{t-1}. \tag{17}$$

An example can be observed in Fig. 3, where the data being used correspond to frame 30 of the *pop2lock* sequence from the University of Surrey data set. The features being shown are colour intensity (first row) and mean curvature (second row).

4.2 Feature Detection

Feature points represent a subset of all vertices that can be detected with high repeatability. Using local gradient information is one way to detect a repeatable feature. Therefore, the feature points are selected as the local extrema over one ring neighbourhoods, in the current and in the adjacent scales. Such an example can be observed in Fig. 1(**b**).

From the extrema of the scale space, only the top $\beta = 5\ \%$ of the maximum number of vertices are being considered, sorted by magnitude. We have chosen a percentage value, versus a hard value threshold, in order to keep the detector flexible, no matter which scalar function is being considered, i.e. colour intensity or mean curvature, without the need for normalization. However, when the threshold response is known a priori for a particular scalar function, such as it is the case in Lowe (2004) with image intensity, it can be easily used instead.

Additionally, in order to eliminate more non-stable responses, we only retain the features that exhibit corner characteristics. As proposed in Lowe (2004), this can be done by examining the eigenvalues of the 2 × 2 Hessian matrix of second directional derivates of the difference of Gaussian operator, i.e., Sect. 3.

Let us consider the Difference of Gaussian operator L_i, defined in (17) and the *symmetric* Hessian matrix approximation $\tilde{\mathbf{H}}_{x,y}(L_i(\mathbf{v}_i))$ from (10), where $(\mathbf{x}, \mathbf{y})$ is a pair of orthonormal vectors lying in the tangent plane $\mathcal{T}_i$ of M at $\mathbf{v}_i$. The absolute value of the ratio between the two sorted eigenvalues, $|\mu_1| \geq |\mu_2|$ of this matrix is a good indication of a corner response. By construction, this ratio is independent of the choice of the local coordinate frame, i.e., vectors $\mathbf{x}$ and $\mathbf{y}$. We use $|\mu_1/\mu_2| = 10$ as threshold value to eliminate the non-stable edge responses.

5 Feature Descriptor (MeshHOG)

In association with the detector presented in the previous sections, we propose building a local descriptor, named MeshHOG, similar in spirit to the histogram of gradient descriptor (HOG) (Dalal and Triggs 2005), but extended to 2D manifolds. A 2D image is in essence a 2D regular grid. The regularity assumption does not always hold in the case of a 2D manifold, that can exhibit non regular sampling. For this reason, the support region of the descriptor has to be chosen using a measure invariant to local triangulation, such as the geodesic distance. In addition to invariance to mesh sampling, the descriptor should also exhibit invariance to a number of other transformations, such as rotation and scale. The scale invariance is achieved by considering the gradient information at the scale of the detected interest point. Rotation invariance is achieved by defining a local coordinate system using the normal at the detected interest point, the dominant gradient in the support region and their cross product. Finally, a two level histogram of gradient is computed, both spatially, at a coarse level, in order to maintain a certain high-level spatial ordering, and using orientations, at a finer level. Trilinear interpolation is used in order to minimize the sampling effect in the histogram bins. The fact that the gradient vectors are 3D allows the computation of the histograms in 3D.

5.1 Support Region

The descriptor for vertex $\mathbf{v}_i$ is computed within a support region $\mathcal{N}(\mathbf{v}_i)$, defined using a geodesic ball of radius r, as in (4). The geodesic support region is chosen adaptively based on a global measure, such that the descriptor is robust to scale and to spatial sampling. The value of r is chosen such that it covers a proportion α_r of the total mesh area, where $\alpha_r \in (0, 1)$. By denoting with A_M the total area of the mesh M, which can be computed as the sum of all triangle areas, the radius of the circular support region is:

$$r = \left[\sqrt{\frac{\alpha_r A_M}{\Pi}}\right], \tag{18}$$

where $[x]$ denotes the integer of x and assuming that the surface covering the ring neighbourhood can be approximated with a circle. In practice, we use r such that $\alpha_r = 2\ \%$.

In the current work we assumed that we are dealing with fully reconstructed objects, thus recovering the "true" global object scale. If this is the case, choosing the size of a support region based on (18) makes it scale and sampling invariant. However, when dealing with partially reconstructed objects, the support-region size should be chosen based on the scale reported by the detected interest point, as suggested in Novatnack and Nishino (2008), Bariya and Nishino (2010), Mian et al. (2010).

5.2 Local Coordinate System

As mentioned earlier, a local coordinate system is desirable, in order to make the descriptor invariant to mesh rotations. A local coordinate system can be built using the unit vector $\mathbf{n}_i$ orthogonal to the plane T_i tangent to M at vertex $\mathbf{v}_i$, and a pair of orthonormal vectors residing in this plane. Given an arbitrary unit vector $\mathbf{a}_i \in T_i$, a local coordinate system C is constructed as:

$$C = \{\mathbf{a}_i, \mathbf{n}_i, \mathbf{a}_i \times \mathbf{n}_i\}. \tag{19}$$

It is therefore important to choose the unit vector $\mathbf{a}_i$ based on some *intrinsic* local property of the scalar function, and hence make the choice of the local coordinate system invariant to mesh rotations. The direction corresponding to the most dominant gradient magnitude in the neighbourhood exhibits such a desired behaviour. Therefore, the unit vector $\mathbf{a}_i$ is chosen as the direction associated with the dominant bin in a polar histogram h_a, with $b_a = 36$ bins. The histogram is constructed by considering the projected participating neighbouring vertices $\mathbf{v}_j \in \mathcal{N}(\mathbf{v}_i)$ onto T_i. The vertex contribution $c_i(\mathbf{v}_j)$ to the appropriate bin takes into account the gradient magnitude and the geodesic distance from the centre vertex $\mathbf{v}_i$, weighted by a Gaussian (see (5)):

$$c_i(\mathbf{v}_j) = \left\|\nabla_M f(\mathbf{v}_j)\right\| G_\sigma\big(d_g(\mathbf{v}_i, \mathbf{v}_j)\big), \tag{20}$$

where the standard deviation is: $\sigma = \epsilon r$, with r the support region radius and ϵ the spatial influence. ϵ is set to 0.5 in our experiments.

Therefore, bin $h_a(k, \mathbf{v}_i)$ yields:

$$h_a(k, \mathbf{v}_i) = \sum_{\mathbf{v}_j \sim \mathbf{v}_i} \chi_{h_a(k,\mathbf{v}_i)}(\mathbf{v}_j) c_i(\mathbf{v}_j), \tag{21}$$

where $k = \{1, 2, \ldots, b_a\}$ is the bin index and $\chi_{h_a(k,\mathbf{v}_i)}(\mathbf{v}_j)$ represents the indicator function for bin selection. In general, the indicator function is defined as :

$$\chi_Z(x) = \begin{cases} 1 & \text{if } x \in bin\, Z, \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

In order to reduce aliasing and the boundary effects of binning, votes $c_i(\mathbf{v}_j)$ are interpolated trilinearly between neighbouring bins during the histogram computation. In practice, that means relaxing the indicator function definition (22). The same interpolation technique is used in the next sub-section.

5.3 The HOG Descriptor

Instead of computing full 3D orientation histograms, as proposed in Kläser et al. (2008), we project the gradient vectors onto the three planes associated with the local coordinate system (19), in order to provide a more compact representation of the descriptor. A possible drawback of the approach described below is a decrease in the discriminability of the descriptor. However, given the fact that local surface neighbourhoods and the associated gradients estimated at the mesh vertices typically span a limited subset of the possible 3D spatial/orientation bins, the current compression scheme does not incur any practical setbacks.

For each of the three planes, a two-level histogram $h_{s,o}$ is computed. First, the plane is divided in $b_s = 4$ polar slices h_s, starting with the origin and continuing in the direction dictated by the right hand rule, with respect to the other orthonormal vector. When projected onto the plane, the participating neighbouring vertices $\mathbf{v}_j$ will fall within one of the spatial slices.

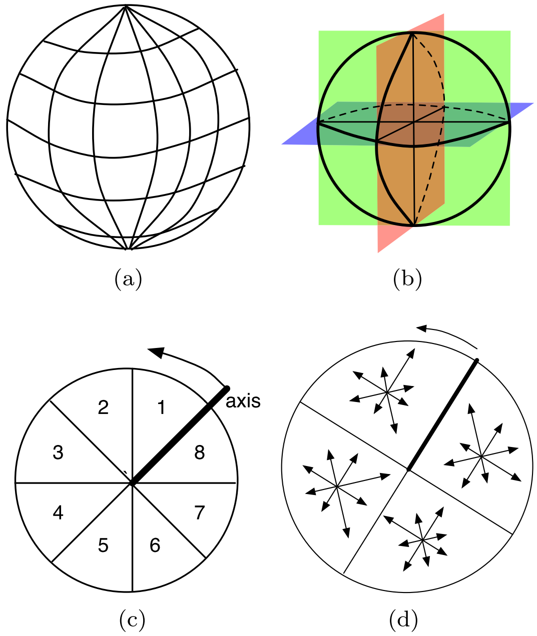
Second, for each spatial slice, the space is divided into $b_o = 8$ orientation slices h_o. The projected gradient vectors $\nabla_M f(\mathbf{v}_j)$ of the vertices $\mathbf{v}_j \in \mathcal{N}(\mathbf{v}_i)$ that projected onto spatial slice are used to determine the orientation slice, as shown in Fig. 4(d).

Similar to the histogram definition (21), the histogram bin $h_{s,o}(e, l, \mathbf{v}_i)$ yields:

$$h_{s,o}(e, l, \mathbf{v}_i) = \sum_{\mathbf{v}_j \sim \mathbf{v}_i} \chi_{h_{s,o}(e,l,\mathbf{v}_i)}(\mathbf{v}_j) c_i(\mathbf{v}_j), \tag{23}$$

where $e = \{1, 2, \ldots, b_s\}$ is the spatial bin index, $l = \{1, 2, \ldots, b_o\}$ is the orientation bin index and $\chi_{h_{s,o}(e,l,\mathbf{v}_i)}$ represents the indicator function for bin selection, defined as:

$$\chi_{h_{s,o}(e,l,\mathbf{v}_i)}(\mathbf{v}_j) = \chi_{h_s(e,\mathbf{v}_i)}(\mathbf{v}_j) \chi_{h_o(l,\mathbf{v}_i)}\big(\nabla_M f(\mathbf{v}_j)\big). \tag{24}$$



(a) (b)

(c) (d)

Fig. 4 (**a**) 3D Histogram—polar mapping used for creating histograms via binning of 3D vectors; (**b**) Choosing three orthogonal planes onto which to project the 3D histogram. (**c**) Polar coordinate system used for creating histograms via binning of 2D vectors, shown in this example with eight polar slices. (**d**) Example of typical spatial and orientation histograms, using four spatial polar slices and eight orientation slices

The final descriptor is obtained by concatenating the $b_s \cdot b_o$ histogram values for each of the 3 orthonormal planes. In order to make the descriptor invariant to mesh sampling, the concatenated histograms are normalized using the L^2 norm. The final descriptor will have $3 \times b_s \times b_o$ elements. Given the previous choice of parameters, the dimensionality of the descriptor is $3 \times 4 \times 8 = 96$.

Whenever reducing the descriptor dimensionality is a requirement, keeping only histograms computed in the tangent plane T_i is a possibility, thus shrinking the descriptor to 32 elements. Also, if multiple scalar functions are available, an aggregate descriptor can be built by concatenating multiple individual descriptors.

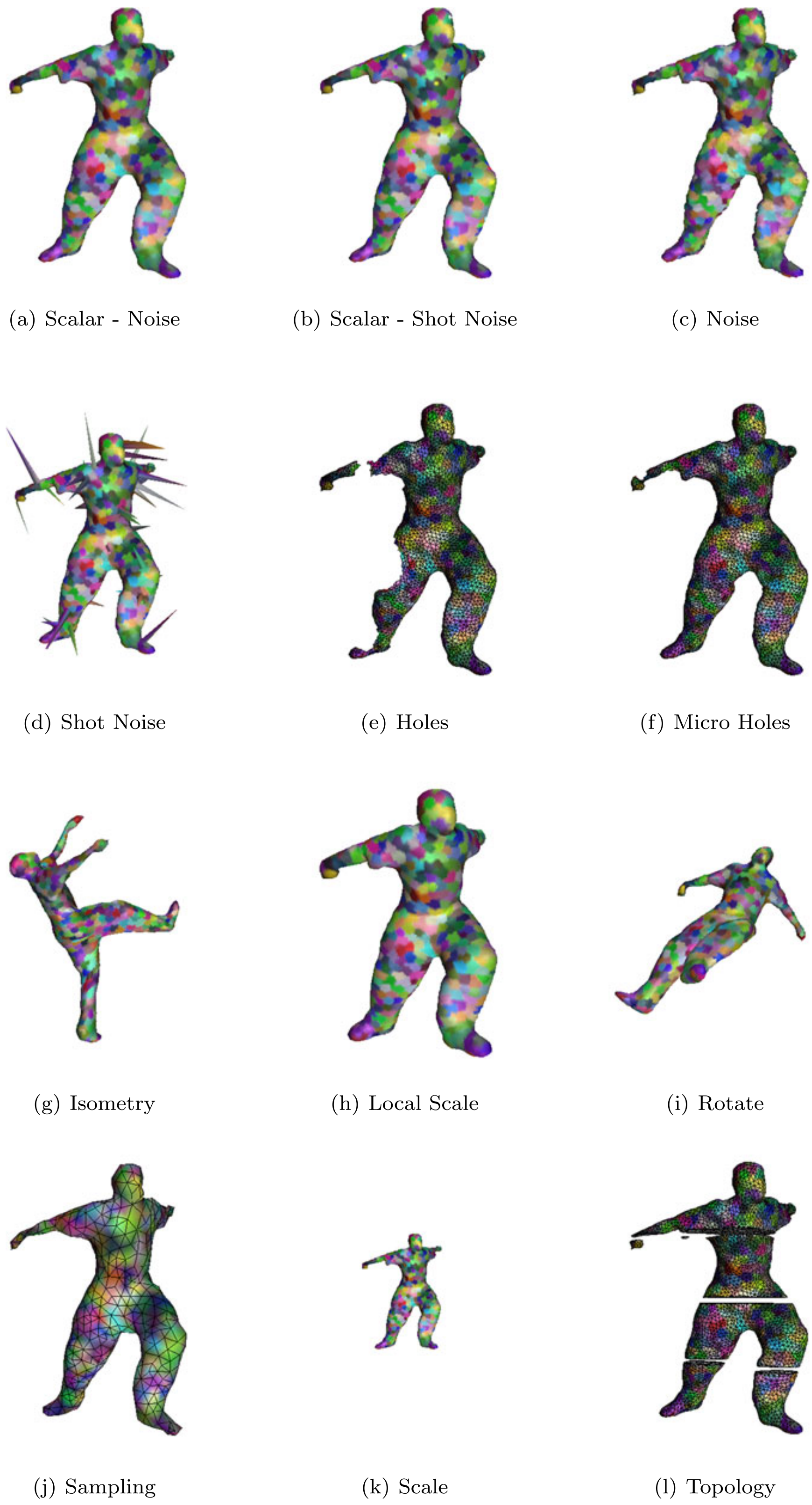
The method has been implemented in C++. The source code has been made available under a GPL license and it can be downloaded from.[2]

6 Performance Evaluation

In this section an extensive evaluation will be performed, covering both the interest point detector and the region descriptor. The original paper (Zaharescu et al. 2009) introduced some preliminary empirical results, in the context of sparse mesh matching.

[2] http://mvviewer.gforge.inria.fr.

Fig. 5 Examples of possible transformations of the null shape (shown in strength 3 out of 5) for the PHOTOMESH dataset



(a) Scalar - Noise

(b) Scalar - Shot Noise

(c) Noise

(d) Shot Noise

(e) Holes

(f) Micro Holes

(g) Isometry

(h) Local Scale

(i) Rotate

(j) Sampling

(k) Scale

(l) Topology

Results are presented in four different scenarios. In Sect. 6.1, the performance is evaluated on the newly proposed PHOTOMESH dataset, which consists of deformations on meshes equipped with photometric information. In Sect. 6.2 the performance of the currently proposed method is compared with other state of the art methods on the

Table 1 Transformations and noise levels of the PHOTOMESH dataset

Transformation	Type	Noise (strength $x = \{1, 2, 3, 4, 5\}$)
Noise	Color	Gaussian Noise with $\sigma_x = \{0.002t, 0.005t, 0.01t, 0.02t, 0.05t\}$, $t = 255$
Shot Noise	Color	Shot noise with signal to noise ratio $SNR_x = \{0.002, 0.005, 0.01, 0.02, 0.05\}$ and noise amplitude $\sigma = 50$.
Noise	Geometry	Gaussian Noise with $\sigma_x = \{0.1t, 0.2t, 0.3t, 0.4t, 0.5t\}$, where $t = e_{avg}$.
Shot Noise	Geometry	Shot noise with signal to noise ratio $SNR_x = \{0.002, 0.005, 0.01, 0.02, 0.05\}$ and noise amplitude $\sigma = 20\ e_{avg}$.
Rotation	Geometry	Gaussian noise with $\sigma_x = \{0.1t, 0.2t, 0.3t, 0.4t, 0.5t\}$, $t = \pi$.
Scale	Geometry	Scale factor $s_x = \{0.5, 0.83, 1.25, 1.62, 2.0\}$.
Local Scale	Geometry	The input mesh is dilated $3 * x$ times. At each iteration, the vertices are moved along the their normal by $e_{avg}/3$.
Sampling	Geometry	The mesh is resampled to attain a desired edge size $e_{avg}(1.0 + x)$, using edge split, edge collapse and edge swap operations, as described in Zaharescu et al. (2011).
Holes	Geometry	x random round holes are created, each having area corresponding to 5 % of the total initial mesh area.
Micro-Holes	Geometry	$3 * x$ random round micro-holes are created, each having an area corresponding to 3 neighbouring rings from the chosen centre vertex.
Topology	Geometry	The input mesh is sliced with x equidistant planes into closed non-connected components, using Zaharescu et al. (2011).
Isometry & Noise	Mixed	Meshes are chosen from the captured 3D temporal sequences. x does not encode noise amplitude. Noise is inherently introduced by the multi-camera mesh reconstruction method and the multi-image colour estimation process.

SHREC 2010 features database (Bronstein et al. 2010), containing only geometric deformations. Additional comparisons with other methods are presented in Sect. 6.3, using the database introduced in Kovnatsky et al. (2011), containing both geometric and photometric deformations, but in the context of mesh retrieval. Finally, additional results are presented in Sect. 6.3 in the context of rigid and non-rigid sparse mesh matching.

6.1 Benchmark on the PHOTOMESH Dataset

While there already exist datasets that test a number of geometrical deformations (see the next section), they do not contain photometric meshes and limit scalar functions to measures of the surface geometry alone. Therefore, a new dataset is proposed, named PHOTOMESH, aimed at testing the repeatability of the detector and the robustness of the descriptor under both photometric and geometric deformations.

6.1.1 Dataset

The dataset consists of 3 base shapes. also called *null* shapes, endowed with photometric information at each vertex. Simulated transformations are applied to them. Such an example is shown in Fig. 5. Two of the null meshes are obtained from multi-view stereo reconstruction algorithms, whereas one is generated with a modelling program. The photometric transformations are noise and shot noise. The geometric transformations are noise, shot noise, rotation, scale, local scale, sampling, holes, micro-holes, topology and isometry. Each transformation has 5 levels of noise applied to it. Therefore, for one base shape, a total of $13 \cdot 5 = 65$ shapes are obtained. Hence, the database contains 135 shapes.

Noise Generally, the noise level corresponds to the noise amplitude. For more information on how the noise is generated for the various transformations, please consult Table 1. For *isometries*, groundtruth is obtained either by means of non-rigid semi-elastic transformations (Cagniart et al. 2010), when using real data, or manually, in a modelling program, such as Blender, when using fully synthetic data. Additional noise is inherently introduced by the multi-view image-based mesh reconstruction and mesh colour estimation process. In addition to the colour differences between views, there are regions that are not visible in any camera view, but which are still reconstructed in 3D due to the interpolation process, such as the sole of a foot.

6.1.2 Evaluation Methodology

The evaluation was performed for feature detection and for region description. The performance is measured by comparing the features and descriptors obtained for the null shape with the ones obtained for the different transformations.

Feature Detection The criteria employed for quantifying the quality of feature detection is *repeatability*. Given that the ground-truth (one-to-one correspondence) is known for each transformed shape B of the null shape A, the repeatability is calculated as the percentage of detected feature points in B that are within a geodesic ball of radius $r = 1\ \%$ of the surface area, from one of the detected interest points in A.

Feature Description The quality of the feature description is measured as the average normalized L^2 distance between descriptors corresponding to matched feature points.

6.1.3 Results and Discussion

Results are presented for three different scalar fields, defined over the manifolds: colour intensity (f_I), mean curvature (f_M) and Gaussian curvature (f_G). Tables 3, 4, 5 summarize the repeatability of the detector, whereas Tables 6, 7, 8 show the robustness of the descriptor.

In the case of *colour noise* and *color shot noise* performance slowly degrades if the scalar function utilizes the colour information (f_I—Tables 3 and 6). In the case of *noise* and *shot noise*, the performance slowly degrades for all functions. Notice, however, that f_I leads to better results than f_M and f_G, because the last two scalar functions use geometric information both when computing the gradient and the actual curvature function.

Rotation and *scale* transformations prove experimentally that the method is indeed invariant to rigid transformations.

Holes, *micro-holes* and *topological* transformations affect the performance of the method linearly. That is partly because these transformations modify the total surface area, which in turn changes the local support area of the descriptor. Also, topological changes introduce new keypoints with the new structures. In the case of holes, some of the keypoints are simply missing.

Even though some invariance to the density of the mesh discretization is built into the method, *sampling* still affects the performance. One aspect is related to the fact that, even though the detection method has invariance built in, during the re-sampling process (when generating the transformation) some of the features could have moved further away than 1 % and they are now detected as incorrect matches. Also, resampling will affect the computation of curvature. Lastly, when computing the descriptor histograms, as the sampling decreases, more bins can become empty, due to the increased sparsity of the participating vertices. One potential way to overcome this effect is to ensure sufficient sampling, by sub-sampling densely enough to guarantee that at least one sample is available for each bin.

Table 2 Performance of different scalar functions under purely isometric transformations

Measure	f_I	f_M	f_G
Repeatability	0.99	0.98	0.98
Robustness	0.07	0.15	0.16

The method also exhibit quasi-invariance to *isometric* transformations. In theory, pure isometric transformations do not affect Gaussian curvature, but they do affect mean curvature and the estimation of normals. In practice, how-

Table 3 Repeatability of MeshDOG (photometric)

Transform.	Strength				
	1	<2	<3	<4	<5
Color Noise	1.00	0.99	0.99	0.97	0.93
Color Shot Noise	0.98	0.96	0.91	0.86	0.76
Geometry Noise	1.00	1.00	1.00	0.99	0.99
Geometry Shot Noise	1.00	0.99	0.99	0.99	0.98
Rotation	1.00	1.00	1.00	1.00	1.00
Scale	1.00	1.00	1.00	1.00	1.00
Local Scale	1.00	1.00	0.99	0.99	0.99
Sampling	0.96	0.96	0.95	0.90	0.94
Holes	1.00	1.00	0.99	0.99	0.97
Micro-Holes	1.00	1.00	0.99	0.99	0.99
Topology	0.93	0.86	0.82	0.82	0.78
Isometry + Noise	0.95	0.97	0.97	0.93	0.96
Average	0.98	0.98	0.97	0.95	0.94

Table 4 Repeatability of MeshDOG (mean curvature)

Transform.	Strength				
	1	<2	<3	<4	<5
Color Noise	1.00	1.00	1.00	1.00	1.00
Color Shot Noise	1.00	1.00	1.00	1.00	1.00
Geometry Noise	0.96	0.93	0.91	0.90	0.89
Geometry Shot Noise	0.99	0.98	0.96	0.95	0.94
Rotation	1.00	1.00	1.00	1.00	1.00
Scale	1.00	1.00	1.00	1.00	1.00
Local Scale	0.99	0.98	0.97	0.96	0.96
Sampling	0.92	0.89	0.91	0.88	0.92
Holes	0.99	0.99	0.99	0.98	0.98
Micro-Holes	1.00	1.00	0.99	0.99	0.98
Topology	0.90	0.83	0.75	0.62	0.76
Isometry + Noise	0.95	0.96	0.94	0.94	0.93
Average	0.97	0.96	0.95	0.93	0.95

Table 5 Repeatability of MeshDOG (Gaussian curvature)

Transform.	Strength				
	1	<2	<3	<4	<5
Color Noise	1.00	1.00	1.00	1.00	1.00
Color Shot Noise	1.00	1.00	1.00	1.00	1.00
Geometry Noise	0.97	0.93	0.87	0.83	0.79
Geometry Shot Noise	0.99	0.98	0.97	0.96	0.92
Rotation	1.00	1.00	1.00	1.00	1.00
Scale	1.00	1.00	1.00	1.00	1.00
Local Scale	0.98	0.98	0.97	0.96	0.95
Sampling	0.88	0.88	0.91	0.94	0.92
Holes	0.99	0.99	0.99	0.97	0.97
Micro-Holes	1.00	0.99	0.99	0.98	0.97
Topology	0.85	0.70	0.65	0.58	0.64
Isometry + Noise	0.95	0.96	0.95	0.92	0.93
Average	0.97	0.95	0.94	0.93	0.92

Table 6 Robustness of MeshHOG (photometric)

Transform.	Strength				
	1	<2	<3	<4	<5
Color Noise	0.02	0.04	0.07	0.10	0.16
Color Shot Noise	0.04	0.11	0.17	0.24	0.31
Geometry Noise	0.18	0.23	0.26	0.28	0.30
Geometry Shot Noise	0.03	0.06	0.11	0.16	0.24
Rotation	0.01	0.01	0.01	0.01	0.01
Scale	0.01	0.01	0.01	0.01	0.00
Local Scale	0.12	0.15	0.18	0.19	0.21
Sampling	0.26	0.30	0.33	0.35	0.34
Holes	0.01	0.02	0.06	0.04	0.06
Micro-Holes	0.01	0.01	0.05	0.05	0.05
Topology	0.14	0.22	0.24	0.26	0.28
Isometry + Noise	0.19	0.20	0.19	0.21	0.21
Average	0.09	0.11	0.14	0.16	0.18

Table 7 Robustness of MeshHOG (mean curvature)

Transform.	Strength				
	1	<2	<3	<4	<5
Color Noise	0.00	0.00	0.00	0.00	0.00
Color Shot Noise	0.00	0.00	0.00	0.00	0.00
Geometry Noise	0.24	0.28	0.30	0.32	0.34
Geometry Shot Noise	0.05	0.10	0.17	0.25	0.36
Rotation	0.01	0.01	0.01	0.01	0.01
Scale	0.01	0.01	0.01	0.01	0.00
Local Scale	0.20	0.25	0.28	0.30	0.31
Sampling	0.30	0.34	0.35	0.36	0.36
Holes	0.01	0.02	0.06	0.06	0.06
Micro-Holes	0.01	0.01	0.06	0.07	0.08
Topology	0.15	0.24	0.26	0.26	0.29
Isometry + Noise	0.23	0.24	0.22	0.25	0.24
Average	0.10	0.12	0.14	0.16	0.17

Table 8 Robustness of MeshHOG (Gaussian curvature)

Transform.	Strength				
	1	<2	<3	<4	<5
Color Noise	0.00	0.00	0.00	0.00	0.00
Color Shot Noise	0.00	0.00	0.00	0.00	0.00
Geometry Noise	0.26	0.29	0.31	0.33	0.34
Geometry Shot Noise	0.04	0.09	0.14	0.21	0.29
Rotation	0.01	0.01	0.01	0.01	0.01
Scale	0.01	0.01	0.01	0.01	0.00
Local Scale	0.21	0.25	0.28	0.30	0.31
Sampling	0.31	0.34	0.34	0.36	0.36
Holes	0.02	0.02	0.07	0.07	0.07
Micro-Holes	0.01	0.01	0.07	0.07	0.08
Topology	0.13	0.20	0.22	0.25	0.28
Isometry + Noise	0.23	0.24	0.22	0.25	0.25
Average	0.10	0.12	0.14	0.15	0.17

ever, the transformations are not purely isometric, due to errors introduced during the mesh tracking or mesh deformation process. In addition, when using real meshes obtained from the multi-view stereo reconstruction, a significant amount of colour noise is implicitly introduced, since colours are interpolated in areas that are non-visible. When considering just the synthetic mesh isometry, the results are a lot more accurate, as presented separately in Table 2.

Overall, the best results are obtained when using the detector/descriptor in conjunction with the photometric information (Tables 3 and 6).

6.2 Benchmark on SHREC 2010 Features Dataset—Non-photometric Meshes

An evaluation was also performed on the Shape Retrieval Contest (SHREC) 2010 dataset[3] (Bronstein et al. 2010), in order to be able to compare with other existing methods.

[3] http://tosca.cs.technion.ac.il/book/shrec_feat.html.

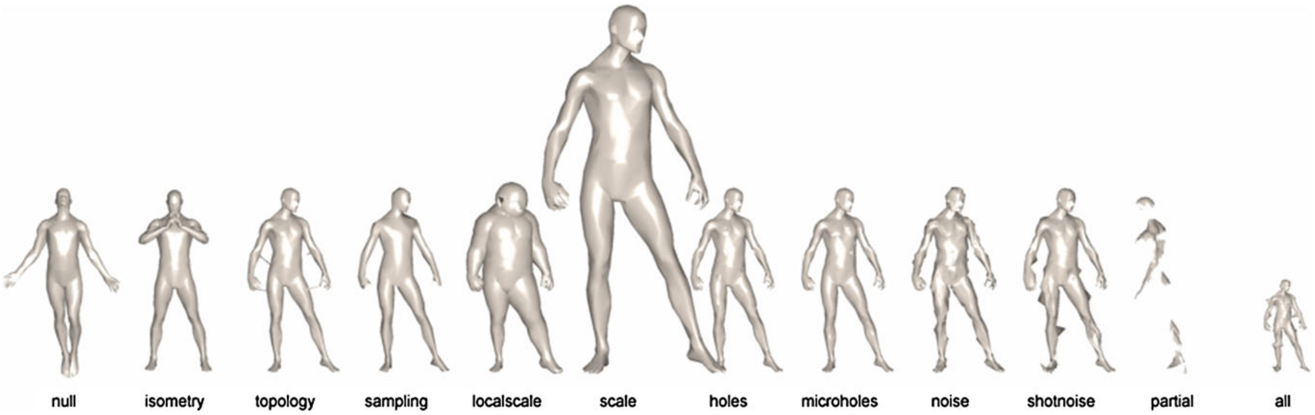


Fig. 6 Examples of possible transformations of the null shape (shown in strength 5 out of 5) for the SHREC 2010 Features dataset. Image taken from Bronstein et al. (2010)

6.2.1 Dataset

The SHREC 2010 feature dataset is similar to the previously proposed benchmark dataset from Sect. 6.1, except for the fact that it does not contain scalar fields defined over the manifolds (i.e. textures), nor any of scalar function transformations. In addition, the null shapes in the current dataset are all synthetic, whereas 66 % of null shapes from the PHOTOMESH dataset are obtained from multi-view stereo. The geometric transformations are similar to the ones introduced in the Sect. 6.1.1. An example can be seen in Fig. 6.

6.2.2 Evaluation Methodology

The evaluation methodology is similar to the one presented in Sect. 6.1.2. The measures tested are the repeatability of feature detections and the robustness of the descriptors.

6.2.3 Results

The results of the proposed method are presented in Tables 9 and 14. The Gaussian curvature was used as the scalar function. The results using mean curvature are very similar, but they are omitted in the interest of space.

Results for other top-performing methods, taken from Bronstein et al. (2010), are also included. For feature detection, the following other methods are included: two heat kernels variants based on the work of Sun et al. (2009)—HK1 (Table 10) and HK2 (Table 11); a Harris 3D corner detection method (Sipiran and Bustos 2010)—H1 (Table 12) and a saliency based method (Castellani et al. 2008)—SP3 (Table 13). For region description, the following methods are included: two sparse heat kernel signature variants based on the work of Sun et al. (2009)—SHK1 (Table 15) and SHK2 (Table 16) and the spin images method (Johnson and Hebert 1999)—SI (Table 17). For more information about the competing methods and for additional results, please consult (Bronstein et al. 2010).

Table 9 Repeatability of MeshDOG: feature detection algorithm using Gaussian curvature as the scalar field. Average number of detected points: 129

Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isometry	97.44	98.72	98.03	98.49	98.49
Topology	97.44	97.44	97.44	97.40	97.41
Holes	96.50	96.50	96.26	95.91	95.55
Micro holes	97.31	97.24	97.22	97.08	96.95
Scale	97.44	97.44	97.35	97.24	97.18
Local scale	94.62	91.67	89.27	85.99	82.62
Sampling	88.08	84.94	81.20	77.82	72.92
Noise	91.92	91.92	90.09	88.59	87.10
Shot noise	97.44	97.50	97.44	97.40	97.38
Average	95.35	94.82	93.81	92.88	91.73

Table 10 Repeatability of HK1: heat kernel based feature detection algorithm. Average number of detected points: 23

Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isometry	98.08	98.72	98.01	97.88	98.04
Topology	97.44	96.10	92.26	91.22	88.64
Holes	91.48	90.60	86.78	83.73	81.86
Micro holes	98.08	96.69	96.00	95.52	94.87
Scale	99.36	99.36	98.50	97.90	97.68
Local scale	98.08	94.83	90.09	83.05	78.31
Sampling	97.05	97.88	97.39	96.27	92.35
Noise	95.30	92.78	91.67	89.24	87.62
Shot noise	98.08	96.22	93.39	90.45	87.32
Average	96.99	95.91	93.79	91.70	89.63

In the context of feature detection, the proposed MeshDOG method performs very well: it has the top average re-

Table 11 Repeatability of HK2: heat kernel based feature detection algorithm. Average number of detected points: 9

Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isometry	100.00	100.00	100.00	100.00	100.00
Topology	94.44	90.38	87.45	88.70	85.76
Holes	80.54	79.00	75.25	72.10	69.99
Micro holes	100.00	100.00	98.15	96.58	95.64
Scale	100.00	100.00	100.00	98.61	97.78
Local scale	97.44	96.79	93.02	87.25	82.90
Sampling	100.00	100.00	100.00	100.00	96.20
Noise	100.00	95.19	93.16	89.37	85.77
Shot noise	100.00	95.30	90.03	82.10	74.38
Average	96.94	95.19	93.01	90.52	87.60

Table 12 Repeatability of H1: Harris 3D feature detection algorithm. Average number of detected points: 303

Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isometry	90.47	91.94	91.71	91.88	92.10
Topology	90.33	90.21	89.93	89.97	89.82
Holes	89.59	89.41	89.25	88.82	88.49
Micro holes	90.42	90.40	90.36	90.33	90.31
Scale	92.21	91.61	90.67	89.55	88.19
Local scale	88.08	86.49	83.64	80.99	78.98
Sampling	84.81	84.80	82.37	78.76	70.68
Noise	89.27	87.36	83.20	79.76	74.53
Shot noise	90.73	90.84	89.43	87.94	86.37
Average	89.55	89.23	87.84	86.44	84.38

sults for noise levels 3 to 5 and scores slightly worse (<2 %) than HK1 and HK2 for noise levels 1 and 2. In the context of region description, MeshHOG's performance is affected by sampling and noise, which pulls down the average performance. As mentioned earlier, the method is not fully robust to sampling errors, mostly due to the fact that for very sparse samplings, a large number of bins are left empty. Proper dense re-sampling of the input mesh would alleviate this problem.

6.3 Benchmarking Using the SHREC Photometric Dataset

A new dataset of photometric meshes is proposed in Kovnatsky et al. (2011) in the context of shape retrieval.

6.3.1 Dataset

The query set consists of 270 real-world human shapes from 5 classes, obtained from multi-view stereo reconstruction, to which a number of transformations have been applied. Geometric transformations are divided into isometry + topology (real articulations and topological changes due to acquisition imperfections), and partiality (occlusions and addition of clutter). Photometric transformations include contrast, brightness, hue, saturation and color noise. Mixed transformations include isometry + topology transformations in combination with two randomly selected photometric transformations. For each class, there are 5 different transformation strength levels, adding up to 54 instances per shape.

The null shape of each of the 5 classes was added to the queried corpus, in addition to the other 75 shapes used as clutter. Retrieval was performed by matching 270 transformed queries to the 75 + 5 null shapes. Each query had exactly one correct corresponding null shape in the dataset.

Table 13 Repeatability of SP3: salient points feature detection algorithm. Average number of detected points: 409

Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isometry	86.17	87.42	87.24	87.76	88.15
Topology	86.18	85.63	85.58	85.56	85.56
Holes	85.72	85.10	84.34	83.56	82.58
Micro holes	68.52	62.27	57.96	54.75	51.99
Scale	89.80	88.28	86.82	85.14	83.70
Local scale	85.73	84.97	84.48	83.33	82.12
Sampling	85.02	83.15	82.21	79.94	77.61
Noise	87.31	85.43	83.28	81.36	79.40
Shot noise	85.95	84.42	82.77	81.76	81.23
Average	84.49	82.96	81.63	80.35	79.15

Table 14 Robustness of MeshHOG feature description algorithm based on features detected by MeshDOG (average L_2 distance between descriptors at corresponding points). Average number of points: 129

Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isometry	0.08	0.07	0.08	0.08	0.08
Topology	0.08	0.08	0.08	0.08	0.08
Holes	0.12	0.13	0.13	0.14	0.15
Micro holes	0.09	0.09	0.09	0.10	0.11
Scale	0.08	0.08	0.08	0.08	0.08
Local scale	0.18	0.25	0.27	0.28	0.31
Sampling	0.37	0.38	0.39	0.40	0.42
Noise	0.37	0.38	0.38	0.38	0.38
Shot noise	0.11	0.11	0.11	0.11	0.11
Average	0.16	0.17	0.18	0.18	0.19

Table 15 Robustness of SHK1: heat kernel signature feature description algorithm based on featured detected by HK1 (average L2 distance between descriptors at corresponding points). Average number of points: 23

Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isometry	0.05	0.04	0.04	0.04	0.04
Topology	0.05	0.06	0.12	0.14	0.19
Holes	0.07	0.07	0.07	0.08	0.09
Micro holes	0.05	0.05	0.06	0.06	0.06
Scale	0.05	0.05	0.05	0.05	0.05
Local scale	0.07	0.09	0.10	0.12	0.14
Sampling	0.06	0.06	0.06	0.08	0.13
Noise	0.08	0.09	0.11	0.12	0.13
Shot noise	0.05	0.06	0.10	0.16	0.25
Average	0.06	0.06	0.08	0.09	0.12

Table 16 Robustness of SHK2: heat kernel signature feature description algorithm based on featured detected by HK2 (average L2 distance between descriptors at corresponding points). Average number of points: 9

Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isometry	0.04	0.03	0.04	0.04	0.04
Topology	0.04	0.06	0.11	0.13	0.18
Holes	0.06	0.07	0.08	0.08	0.09
Micro holes	0.04	0.04	0.05	0.05	0.05
Scale	0.04	0.04	0.04	0.04	0.04
Local scale	0.07	0.08	0.10	0.13	0.16
Sampling	0.05	0.05	0.05	0.07	0.14
Noise	0.08	0.09	0.11	0.12	0.13
Shot noise	0.05	0.08	0.15	0.24	0.31
Average	0.05	0.06	0.08	0.10	0.13

Table 17 Robustness of Spin Images feature description algorithm based on features detected by SP2 (average L2 distance between descriptors at corresponding points). Average number of points: 205

Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isometry	0.12	0.10	0.10	0.10	0.10
Topology	0.11	0.11	0.11	0.11	0.11
Holes	0.12	0.12	0.12	0.12	0.12
Micro holes	0.15	0.15	0.16	0.16	0.16
Scale	0.18	0.15	0.15	0.15	0.15
Local scale	0.12	0.13	0.14	0.15	0.17
Sampling	0.13	0.13	0.13	0.13	0.15
Noise	0.13	0.15	0.17	0.19	0.20
Shot noise	0.11	0.13	0.16	0.17	0.18
Average	0.13	0.13	0.14	0.14	0.15

Table 18 Performance (mAP in %) of BoFs using MeshHOG descriptors (photometry)

Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isom + Topo	100.00	95.00	96.67	94.17	95.33
Partial	75.00	61.15	69.93	68.28	68.79
Contrast	100.00	100.00	100.00	98.33	94.17
Brightness	100.00	100.00	100.00	100.00	99.00
Hue	100.00	100.00	100.00	100.00	100.00
Saturation	100.00	100.00	100.00	98.75	99.00
Noise	100.00	100.00	88.89	83.33	78.33
Mixed	100.00	100.00	100.00	93.33	83.40

6.3.2 Evaluation Methodology

Performance was evaluated using the precision-recall characteristic for the shape retrieval task. Precision $P(r)$ is defined as the percentage of relevant shapes in the first r top-ranked retrieved shapes. Mean average precision (mAP) was used as a single measure of performance, where $mAP = \sum_r P(r) \cdot rel(r)$ and $rel(r)$ is the relevance of a given rank.

6.3.3 Results

Results for the proposed method are summarized in Table 18. Photometric information was used in the scalar field.

Results for other top-performing methods are also presented in Tables 19–24. For more information, please consult (Kovnatsky et al. 2011). The methods using bag of features with heat kernel signatures (Tables 19) and the spectral distance (Table 20) are purely geometric, which makes them automatically invariant to the photometric noise (*contrast, brightness, hue, saturation, noise*). Conversely, the colour histogram method (Table 21) uses purely photometric information, thus making it invariant to purely geometric transformations (*isometry + topology* and *partial*). The methods presented in Tables 22–24 use spectral decomposition methods in conjunction with both photometric and geometric information.

Comparing the overall results, the proposed method (Table 18) performs very well. It provides the best results for the partial geometric transformations (occlusions and the addition of clutter), as well as the best overall results for noise strength 1. While overall top ranking results are obtained by the method described in Table 23, the proposed method comes to a very close second.

Table 19 Performance (mAP in %) of ShapeGoogle using BoFs with HKS descriptors (purely geometric)

Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isom + Topo	100.00	100.00	96.67	95.00	90.00
Partial	66.67	60.42	63.89	63.28	63.63
Contrast	100.00	100.00	100.00	100.00	100.00
Brightness	100.00	100.00	100.00	100.00	100.00
Hue	100.00	100.00	100.00	100.00	100.00
Saturation	100.00	100.00	100.00	100.00	100.00
Noise	100.00	100.00	100.00	100.00	100.00
Mixed	90.00	95.00	93.33	95.00	96.00

Table 20 Performance (mAP in %) of pure geometric spectral shape distance

Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isom + Topo	80.00	90.00	88.89	86.67	89.33
Partial	56.25	65.62	61.61	58.71	61.13
Contrast	100.00	100.00	100.00	100.00	100.00
Brightness	100.00	100.00	100.00	100.00	100.00
Hue	100.00	100.00	100.00	100.00	100.00
Saturation	100.00	100.00	100.00	100.00	100.00
Noise	100.00	100.00	100.00	100.00	100.00
Mixed	66.67	73.33	78.89	81.67	81.33

Table 21 Performance (mAP in %) of color histograms (purely photometric)

Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isom + Topo	100.00	100.00	100.00	100.00	100.00
Partial	100.00	100.00	100.00	100.00	100.00
Contrast	100.00	90.83	80.30	71.88	63.95
Brightness	88.33	80.56	65.56	53.21	44.81
Hue	11.35	8.38	6.81	6.05	5.49
Saturation	17.47	14.57	12.18	10.67	9.74
Noise	100.00	100.00	93.33	85.00	74.70
Mixed	28.07	25.99	20.31	17.62	15.38

6.4 Shape Matching

The current subsection introduces results of the currently proposed approach in a shape matching application.

Table 22 Performance (mAP in %) of BoFs with cHKS descriptors using fixed colorspace scaling factor ($w = 0.1$)

Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isom + Topo	90.00	95.00	96.67	97.50	96.00
Partial	81.25	74.38	71.11	64.48	65.08
Contrast	100.00	100.00	100.00	98.75	98.00
Brightness	100.00	100.00	100.00	100.00	100.00
Hue	100.00	95.00	96.67	97.50	98.00
Saturation	100.00	96.00	84.51	76.53	71.39
Noise	100.00	100.00	86.33	81.62	76.03
Mixed	86.67	79.76	76.17	78.38	71.81

Table 23 Performance (mAP in %) of ShapeGoogle using w-multi-scale BoFs with cHKS descriptors

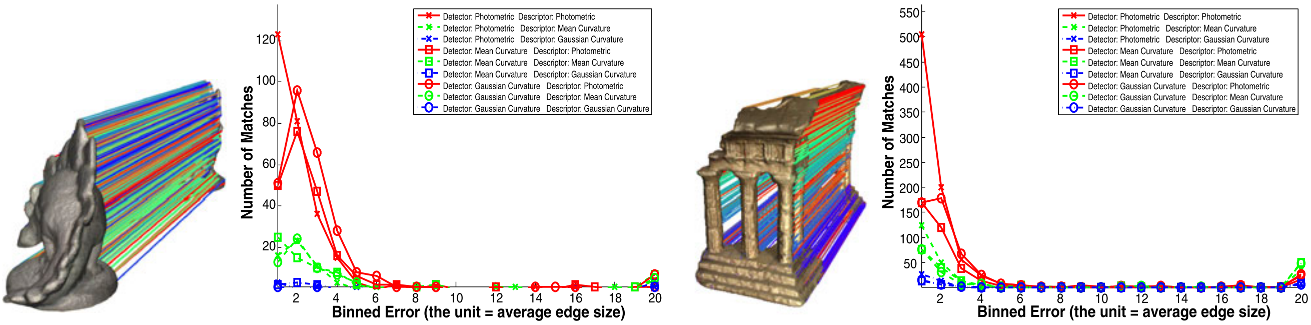
Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isom + Topo	100.00	100.00	96.67	97.50	94.00
Partial	68.75	68.13	69.03	67.40	67.13
Contrast	100.00	100.00	100.00	100.00	100.00
Brightness	100.00	100.00	100.00	100.00	100.00
Hue	100.00	100.00	100.00	100.00	100.00
Saturation	100.00	100.00	100.00	100.00	100.00
Noise	100.00	100.00	100.00	100.00	100.00
Mixed	100.00	100.00	96.67	97.50	98.00

Table 24 Performance of (mAP in %) of the multiscale joint geometric-photometric spectral distance

Transform.	Strength				
	1	≤2	≤3	≤4	≤5
Isom + Topo	100.00	100.00	100.00	100.00	100.00
Partial	62.50	72.92	65.97	62.50	67.50
Contrast	100.00	100.00	100.00	100.00	100.00
Brightness	100.00	100.00	100.00	100.00	100.00
Hue	100.00	100.00	100.00	100.00	100.00
Saturation	100.00	100.00	100.00	100.00	100.00
Noise	100.00	100.00	100.00	100.00	100.00
Mixed	100.00	93.33	95.56	96.67	93.70

6.4.1 Method

Let A and B be two meshes of the same object. The two meshes do not necessarily have the same number of vertices. Using the proposed approach, n_A interest points are detected on A, characterized by descriptors $\mathbf{t}_i^A$, with $i \in [1..n_A]$. Sim-



(a) Dino (b) Dino (c) Temple (d) Temple

Fig. 7 Rigid matching results—*Dino* and *Temple* datasets. (**a**), (**c**) Results when using photometric information for both detection and description; (**b**), (**d**) Error distribution when using different combinations of scalar functions

ilarly, n_B interest points are detected on B, characterized by descriptors $\mathbf{t}_j^B$, with $j \in [1..n_B]$.

For each descriptor $\mathbf{t}_i^A$ from surface A, the best matching descriptor $\mathbf{t}_j^B$ from surface B is found, in terms of the Euclidean distance $d_{ij} = \|\mathbf{t}_i^A - \mathbf{t}_j^B\|$. Cross validation is performed, by checking that $\mathbf{t}_j^B$'s best match is indeed $\mathbf{t}_i^A$. Finally, the candidate match is only accepted if the second best match is significantly worse ($\gamma = 0.7$ or less from the best match score). This method is not meant to fully solve the matching problem, as would a global approach (e.g. Starck and Hilton 2007). It is intended to allow further validation and evaluation of the proposed detector and descriptor.

6.4.2 Datasets

In the evaluation, the following scenarios are considered: (i) the two meshes are representations of the same rigid object, which can thus be aligned using a rigid transformation; (ii) the two shapes are representations of the same non-rigid object, i.e. a moving person. In this context, the following datasets are introduced:

- *Rigid Objects*: Reconstructions of the same object are considered, using different camera sets. In particular, different mesh reconstructions are obtained using (Zaharescu et al. 2011) on the publicly available datasets from the Middlebury Multi-View Stereo site (Seitz et al. 2006). The *Dino* datasets contains two meshes, one with 27240 vertices obtained from 16 cameras and the other of 31268 vertices generated from 47 cameras. Similarly, the *Temple* datasets contains two meshes, one with 78019 vertices obtained from 16 cameras and the other of 80981 vertices generated from 47 cameras.
- *Synthetic Non-Rigid Objects*: A synthetically generated dataset is considered, entitled *Synth-Dance*, of a human mesh with 7061 vertices moving across 200 frames.
- *Real Non-Rigid Objects*: Additionally, frames 515–550 from the INRIA *Dance-1* sequence[4] are used, where the same reconstruction method (Zaharescu et al. 2011) was employed to recover models using 32 cameras. The meshes have vertices ranging between 16212 and 18332.

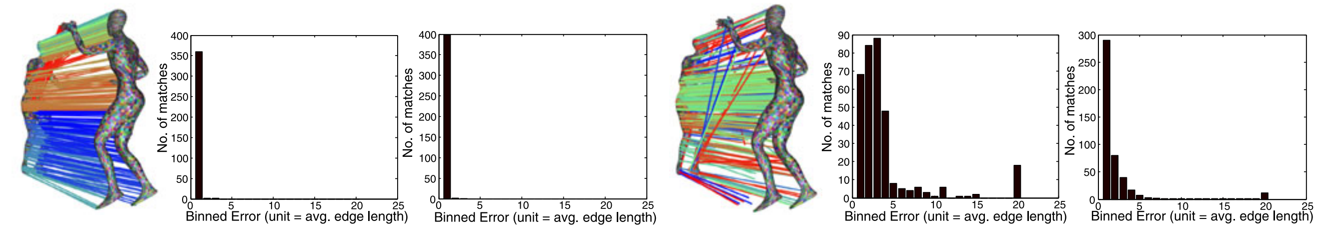
Photometric information The colour of each vertex of the surface is computed by considering the median colour in the visible images. It is assumed that the colours of a vertex follows a non-Gaussian distribution, due to errors that can occur around occluding contours. In the *Synth-Dance* dataset the vertices are randomly coloured.

6.4.3 Evaluation Methodology

For each of the cases, a number of matches are produced by the above mentioned matching algorithm. In the case of rigid objects and synthetic non-rigid objects, the match groundtruth is readily available. Therefore, error distributions of the matches can be computed. They accumulate binned error distance from the groundtruth match, with the size of a bin being the average edge length.

Comparison with 2D SIFT In addition, for the non-rigid motion cases, a comparison is presented between the proposed mesh matching framework using the MeshHOG descriptor and a similar matching framework, based on 2D image detectors and descriptors back-projected onto the mesh. In the image-based framework, the matching is performed in images and only then the 2D matches are back-projected onto the surface. The SIFT (Lowe 2004) keypoint detector and image descriptor was used in 2D. When matching the two surfaces, only matches from the same cameras are considered. In order to carry such a comparison for the *Synth-Dance* dataset, 16 virtual cameras and the associated images have been generated. The virtual cameras are distributed in a circular pattern around the object.

[4] http://4drepository.inrialpes.fr/.



(a) MeshHOG (b) MeshHOG (c) MeshHOG (d) SIFT (e) SIFT (f) SIFT

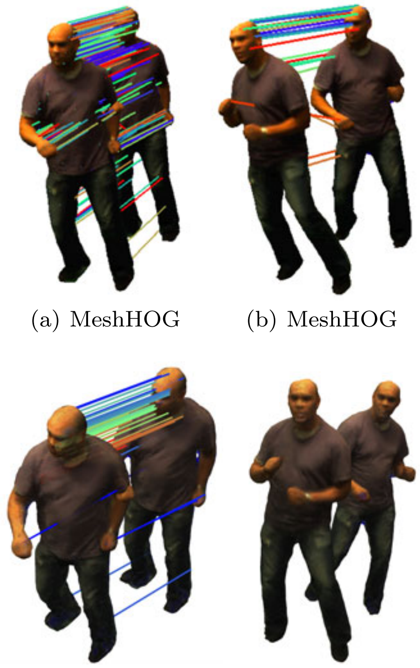
Fig. 8 Non Rigid matching using synthetic data—*dancer-synth* dataset. Comparison between MeshHOG and SIFT matching results. Matches between frames 1 and 50 are visually depicted in (**a**), (**d**). There are 364 matches for MeshHOG and 343 matches for SIFT. They are also quantified in the error histograms (**b**), (**e**). The histogram bins are of size equal to e_{avg}. The last bin groups all the errors greater than $20 * e_{avg}$. Additionally, the average histogram errors are shown in (**c**), (**f**) for matching frame 1 with x, where $x \in [2..200]$

6.4.4 Results

Rigid Matching Results Figure 7 presents results on the *Dino* and *Temple* datasets, with different possible combinations of scalar functions for both detection and description: photometric information, mean curvature and Gaussian curvature. As the results show, the biggest benefit is obtained from using the photometric information for region description, irrespective of which scalar function is used for detection, e.g. the red curves in the graphs from Fig. 7. However, the best overall results are obtained when photometric information is used for both keypoint detection and region description. Mean curvature seems to be the second most informative measure when used for region description (the green curves in the graphs from Fig. 7), whereas the Gaussian curvature (the blue curves) is the worst performer.

Given the datasets and the matching procedure, it is to be expected that photometric information provides the best choice for region description. Geometry alone does not provide sufficient unique regions. Both the *Dino* and *Temple* meshes exhibit a large number of repetitive geometric features, that can only be disambiguated due to the slightly different photometric information.

Non-Rigid Matching Results Synthetic comparative results are presented in Fig. 8. The mesh in the first frame was matched with all the other 199 meshes across the sequence. Observe that the MeshHOG descriptor generates very few false positives in comparison with the SIFT equivalent, clearly demonstrating the advantages of the proposed approach. In addition, empirical results are presented in Fig. 9 for the INRIA *Dance-1* sequence. There are only 54 matches found using the SIFT back-projected method between frame 525 and 526, whereas MeshHOG finds 174 matches. Even when matching across distant frames (530 and 550), our proposed method finds 27 correct matches, whereas the SIFT back-projected method fails completely. It is to be expected, since most of the inter-frame matches are due to local creases formed by the clothes. The head is the only unique feature that can be robustly matched across time.



(a) MeshHOG (b) MeshHOG

(c) SIFT (d) SIFT

Fig. 9 Non Rigid matching using real data—*Dance-1* sequence. (**a**) Matches between frames 525 and 526 using MeshHOG (174 matches); (**b**) Matches between frames 530 and 550 using MeshHOG (27 matches); (**c**) Matches between frames 525 and 526 using SIFT (54 matches); (**d**) Matches between frames 530 and 550 using SIFT (0 matches)

7 Conclusion

We have introduced MeshDOG and MeshHOG, a new 3D interest point detector and a new 3D descriptor, defined on triangular meshes endowed with a scalar function. The descriptor is able to capture the properties of both the local geometry and of the scalar function in a succinct fashion. It is robust to changes in orientation, rotation, translation and scale. The performance of both the interest point detector and the feature descriptor was tested extensively, achieving very competitive results, comparable with the state of the art.

As a future direction, we plan to investigate how to further extend the descriptor to take into account the temporal dimension, considering the context of densely tracked surfaces.

Acknowledgements We would like to thank Cedric Cagniart and Artiom Kovnatsky for their help with the datasets.

References

de Aguiar, E., Theobalt, C., Stoll, C., & Seidel, H. P. (2007). Markerless 3D feature tracking for mesh-based human motion capture. In *Human motion—understanding, modeling, capture and animation* (pp. 1–15).

Ahmed, N., Theobalt, C., Rossl, C., Thrun, S., & Seidel, H. P. (2008). Dense correspondence finding for parametrization-free animation reconstruction from video. In *Proceedings of IEEE conference on computer vision and pattern recognition*.

Bariya, P., & Nishino, K. (2010). Scale-hierarchical 3d object recognition in cluttered scenes. In *Proc. of IEEE computer vision and pattern recognition* (pp. 1657–1664).

Barth, T. (1993). A 3-D least-squares upwind Euler solver for unstructured meshes. In *Lecture notes in physics: Vol. 414. Thirteenth international conference on numerical methods in fluid dynamics* (pp. 240–244). Berlin: Springer.

Bay, H., Ess, A., Tuytelaars, T., & Gool, L. V. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, *110*(3), 346–359.

Bolles, R. C., & Cain, RA (1982). Recognizing and locating partially visible objects, the Local-Feature-Focus method. *The International Journal of Robotics Research*, *1*(3), 57–82.

Bolles, R. C., & Horaud, R. (1986). 3DPO: A three-dimensional part orientation system. *The International Journal of Robotics Research*, *5*(3), 3–26.

Bronstein, A. M., Bronstein, M. M., Bustos, B., Castellani, U., Crisani, M., Falcidieno, B., Guibas, L. J., Kokkinos, I., Murino, V., Ovsjanikov, M., Patané, G., Sipiran, I., Spagnuolo, M., & Sun, J. (2010). Shrec 2010: robust feature detection and description benchmark. In *Proc. EUROGRAPHICS workshop on 3D object retrieval (3DOR)*.

Bronstein, A. M., Bronstein, M. M., Ovsjanikov, M., & Guibas, L. J. (2011). Shape Google: geometric words and expressions for invariant shape retrieval. *ACM Transactions on Graphics*, *30*(1), 1–20.

Bustos, B., Keim, D. A., Saupe, D., Schreck, T., & Vranic, D. V. (2005). Feature-based similarity search in 3D object databases. *ACM Computing Surveys*, *34*(4), 345–387.

Cagniart, C., Boyer, E., & Ilic, S. (2010). Probabilistic deformable surface tracking from multiple videos. In *Proceedings of European conference on computer vision*.

Castellani, U., Cristani, M., Fantoni, S., & Murino, V. (2008). Sparse points matching by combining 3D mesh saliency with statistical descriptors. *Computer Graphics Forum*, *27*(2), 643–652.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proceedings of IEEE conference on computer vision and pattern recognition* (Vol. 1, pp. 886–893).

Dong, C. S., & Wang, G. Z. (2005). Curvatures estimation on triangular mesh. *Journal of Zhejiang University SCIENCE*, *6A*(1), 128–136.

Dufournaud, Y., Schmid, C., & Horaud, R. P. (2004). Image matching with scale adjustment. *Computer Vision and Image Understanding*, *93*(2), 175–194.

Frome, A., Huber, D., Kolluri, R., Bulow, T., & Malik, J. (2004). Recognizing objects in range data using regional point descriptors. In *Proceedings of European conference on computer vision*.

Furukawa, Y., & Ponce, J. (2008). Dense 3D motion capture from synchronized video streams. In *Proceedings of IEEE conference on computer vision and pattern recognition*.

Horn, R. A., & Johnson, C. A. (1994). *Matrix analysis*. Cambridge: Cambridge University Press.

Hou, T., & Qin, H. (2010). Efficient computation of scale-space features for deformable shape correspondences. In *Proceedings of European conference on computer vision*.

Hua, J., Lai, Z., Dong, M., Gu, X., & Qin, H. (2008). Geodesic distance-weighted shape vector image diffusion. *IEEE Transactions on Visualization and Computer Graphics*, *14*(6), 1643–1650.

Johnson, A. E., & Hebert, M. (1999). Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(5), 433–449.

Kimmel, R., & Sethian, J. (1998). Computing geodesic paths on manifolds. In *Proceedings of national academy of science* (pp. 8431–8435).

Kläser, A., Marszałek, M., & Schmid, C. (2008). A spatio-temporal descriptor based on 3D-gradients. In *Proceedings of the British machine vision conference*.

Körtgen, M., Park, G. J., Novotny, M., & Klein, R. (2003) 3D shape matching with 3D shape contexts. *Central European Seminar on Computer Graphics*.

Kovnatsky, A., Bronstein, M. M., Bronstein, A. M., & Kimmel, R. (2011). Photometric heat kernel signatures. In *Proceedings of conference on scale space and variational methods in computer vision*.

Laptev, I. (2005). On space-time interest points. *International Journal of Computer Vision*, *64*(2–3), 107–123.

Lay, D. (1996). *Linear algebra and its applications*. Reading: Addison-Wesley.

Lee, C. H., Varshney, A., & Jacobs, D. (2005) Mesh saliency. *Proceedings of SIGGRAPH*.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

Luo, C., Safa, I., & Wang, Y. (2009). Approximating gradients for meshes and point clouds via diffusion metric. In *Proceedings of the Eurographics symposium on geometry processing*.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London*, *B207*, 187–217.

Matas, J., Chum, O., Urban, M., & Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, *22*(10), 761–767.

Mavriplis, D. (2003). Revisiting the least-squares procedure for gradient reconstruction on unstructured meshes. In *Proc. of the 16th AIAA computational fluid dynamics conference*, Orlando, FL.

Meyer, M., Desbrun, M., Schröder, P., & Barr, A. H. (2002). Discrete differential geometry operators for triangulated 2-dimensional manifolds. In *Proceedings of VisMath*.

Mian, A., Bennamoun, M., & Owens, R. (2010). On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes. *International Journal of Computer Vision*, *89*, 348–361.

Mikolajczyk, K., & Schmid, C. (2004). Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, *60*(1), 63–86.

Mikolajczyk, K., & Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(10), 1615–1630.

Mukherjee, S., Wu, Q., & Zhou, D. X. (2010). Learning gradients on manifolds. *Bernoulli*, *16*(1), 181–207.

Novatnack, J., & Nishino, K. (2007). Scale-dependent 3D geometric features. In *Proceedings of international conference on computer vision*.

Novatnack, J., & Nishino, K. (2008). Scale-dependent/invariant local 3D shape descriptors for fully automatic registration of multiple sets of range images. In *Proceedings of European conference on computer vision* (Vol. III, pp. 440–453).

Rothganger, F., Lazebnik, S., Schmid, C., & Ponce, J. (2006). 3D object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints. *International Journal of Computer Vision*, *66*(3), 231–259.

Ruggeri, M. R., Patanè, G., Spagnuolo, M., & Saupe, D. (2010). Spectral-driven isometry-invariant matching of 3D shapes. *International Journal of Computer Vision*, *89*(2–3), 248–265.

Schlattmann, M., Degener, P., & Klein, R. (2008). Scale space based feature point detection on surfaces. *Journal of WSCG*, *16*(1–3).

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., & Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proceedings of IEEE conference on computer vision and pattern recognition* (Vol. 1, pp. 519–526).

Shilane, P., Min, P., Kazhdan, M., & Funkhouser, T. (2008). The Princeton shape benchmark. In *Shape modeling international*.

Sibson, R. (1981). *A brief description of natural neighbour interpolation* (Vol. 21, pp. 21–36). New York: Wiley.

Sipiran, I., & Bustos, B. (2010). A robust 3D interest points detector based on Harris operator. In *3DOR* (pp. 7–14).

Smith, E. R., Radke, R. J., & Stewart, C. V. (2011). Physical scale keypoints: Matching and registration for combined intensity/range images. *International Journal of Computer Vision*.

Starck, J., & Hilton, A. (2007). Correspondence labelling for widetime free-form surface matching. In *Proceedings of international conference on computer vision*.

Sun, J., Ovsjanikov, M., & Guibas, L. (2009). A concise and provably informative multi-scale signature based on heat diffusion. In *Proceedings of the symposium on geometry processing*.

Surazhsky, V., Surazhsky, T., Kirsanov, D., Gortler, S., & Hoppe, H. (2005). Fast exact and approximate geodesics on meshes. *Proceedings of SIGGRAPH*.

Tangelder, J. W. H., & Veltkamp, R. C. (2004) A survey of content based 3D shape retrieval methods. In *Shape modeling international* (pp. 145–156).

Varanasi, K., Zaharescu, A., Boyer, E., & Horaud, R. P. (2008). Temporal surface tracking using mesh evolution. In *Proceedings of European conference on computer vision*.

Wong, S. F., & Cipolla, R. (2007). Extracting spatiotemporal interest points using global information. In *Proceedings of international conference on computer vision*.

Wu, C., Clipp, B., Li, X., Frahm, J. M., & Pollefeys, M. (2008). 3D model matching with viewpoint invariant patches (vips). In *Proceedings of IEEE conference on computer vision and pattern recognition*.

Xu, G. (2004). Convergent discrete Laplace-Beltrami operators over triangular surfaces. In *Proceedings of geometric modeling and processing* (pp. 195–204).

Zaharescu, A., Boyer, E., Varanasi, K., & Horaud, R. (2009). Surface feature detection and description with applications to mesh matching. In *Proceedings of IEEE conference on computer vision and pattern recognition*, Miami, USA (pp. 373–380).

Zaharescu, A., Boyer, E., & Horaud, R. (2011). Topology-adaptive mesh deformation for surface evolution, morphing, and multiview reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(4), 823–837.

Zhong, Y. (2009). Intrinsic shape signatures: A shape descriptor for 3D object recognition. In *IEEE international conference on computer vision (3D) representation and recognition workshop*.