Efficient Computation of Scale-Space Features for Deformable Shape Correspondences

Tingbo Hou and Hong Qin

Department of Computer Science, Stony Brook University
Stony Brook, New York, 11794, USA
{thou,qin}@cs.sunysb.edu

Abstract. With the rapid development of fast data acquisition techniques, 3D scans that record the geometric and photometric information of deformable objects are routinely acquired nowadays. To track surfaces in temporal domain or stitch partially-overlapping scans to form a complete model in spatial domain, robust and efficient feature detection for deformable shape correspondences, as an enabling method, becomes fundamentally critical with pressing needs. In this paper, we propose an efficient method to extract local features in scale spaces of both texture and geometry for deformable shape correspondences. We first build a hierarchical scale space on surface geometry based on geodesic metric, and the pyramid representation of surface geometry naturally engenders the rapid computation of scale-space features. Analogous to the SIFT, our features are found as local extrema in the scale space. We then propose a new feature descriptor for deformable surfaces, which is a gradient histogram within a local region computed by a local parameterization. Both the detector and the descriptor are invariant to isometric deformation, which makes our method a powerful tool for deformable shape correspondences. The performance of the proposed method is evaluated by feature matching on a sequence of deforming surfaces with ground truth correspondences.

Key words: Shape feature, Scale space, Deformable shape, SIFT

1 Introduction

In recent years, the rapid development of data acquisition techniques naturally gives rise to a massive collection of 3D scans of the geometric and photometric information of deformable objects. Comparing with 2D images that are generally perspective projections of the scenes, 3D scans capture the shape and the photograph of objects. Feature extraction is an enabling method to organize/index partial 3D scans of an object and track its temporal variations. Strongly motivated by scientific evidences from both physics and biological vision, scale-space features that can appear in multiple scales, are much more desirable with many attractive properties. This paper aims to compute scale-space features directly on deformable surfaces, with a unique application of matching and tracking surfaces undergoing deformation.

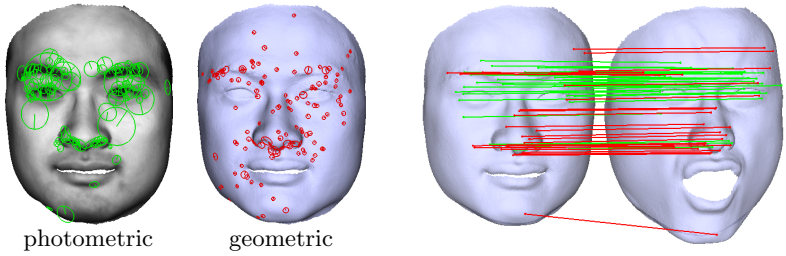


Fig. 1. Photometric (green) and geometric (red) features for deformable shape correspondences (right). Oriented circles indicate the scales and orientations of the features. The geometric information here is the Gaussian curvature map.

Different from 2D images, 3D scans acquired by active scanner, stereo vision, multi-view silhouette, or a mixture of them, are commonly represented as triangular meshes, with specific challenges:

- Deformation: Besides shape geometry, consecutive scans recording deformation frequently have unpredictable changes of topology and boundary.
- Irregular grid: The triangulation of scans is typically irregular, with no restriction on the valence (number of connected edges) of a vertex.
- Metric: On curved surfaces, the metric is often referred to the geodesic distance rather than the Euclidean distance.
- Local access: In a triangular mesh, the global index does not reflect its connectivity, and thus the data can only be accessed locally.

Such challenges create many difficulties towards developing algorithms to compute scale-space features on deformable surfaces, with the purpose of shape correspondence. The state of the art therefore is struggling in finding efficient and robust algorithms for local feature extraction and representation on deformable surfaces. For metric choices, some used the Euclidean metric [1–3] that is not preserved under deformation. Some other [4] employed the geodesic metric that is invariant to isometric deformation. Besides, there are also methods converting surfaces to some intrinsic domains to address deformations, e.g., the parametric domain [5–8] and the frequency domain [9, 10]. Parameterization-based methods are usually accompanied by model cutting, and can be easily affected by topological changes. Frequency-based methods decompose the surface into its globally defined Laplace-Beltrami eigenfunctions (LBE), and thus are not applicable for local features. Moreover, most existing methods only focus on geometric characteristics (e.g., normal, curvature, spectrum, etc.). We are interested in the concept of scalar fields on surfaces [3], which neatly combines geometric and photometric characteristics.

In this paper, we develop an efficient method to compute scale-space features using geodesic metric for deformable shape correspondences. Specifically, the contributions of this paper are as follows:

- We present a hierarchical scale space using the pyramid representation, together with geodesic metric. It downsamples the surface when the scale in-

creases, while controlling the sampling rate by a constant factor in the scale space. This hierarchical scale space elegantly integrates photometric and geometric characteristics, and engenders the computation efficiency.
– We propose a new feature descriptor for deformable surfaces enabled by a local parameterization. This descriptor is a gradient histogram on a local region parameterized by geodesics and polar angles in the local tangent plane.
– We evaluate the performance of feature descriptors via a matching experiment on a dataset, which contains a sequence of deforming surfaces with ground truth correspondences.

An example of photometric (green) and geometric (red) features extracted by our method is shown in Fig. 1(left), where the deformable shape correspondences based on these features are shown on the right. The geometric information used here is the Gaussian curvature map.

2 Previous Work

Extracting distinctive local features for 2D images is a fundamental and long-lasting task in computer vision. Perhaps the most influential method with great impact is the scale invariant feature transformation (SIFT) proposed by Lowe [11]. The success of SIFT primarily lies in its effective strategies including pyramid representation, extremum detection in scale space, orientation assignment, histogram of gradients, etc. According to a performance evaluation of local feature descriptors [12], SIFT-based descriptor could reach the best performance. With continuously-increasing interest, extensions have been made to improve the SIFT in recent years. Ke and Sukthankar [13] applied principal components analysis (PCA) to the normalized gradient patch instead of weighted histograms in SIFT. The PCA-SIFT representation with top eigenvectors is more compact than the gradient image, whereas it requires pre-computation over large amount of training data. Mikolajczyk and Schmid [12] proposed the gradient location-orientation histogram (GLOH) descriptor, which extends the rectangular sampling grid of SIFT to a log-polar sampling grid that is more meaningful under rotation. Tola et al. [14] introduced the descriptor "DAISY" replacing the weighted sums of gradient norms in SIFT by convolutions with several oriented derivatives of Gaussian filters, which can be computed even faster without degrading the performance.

Strongly inspired by the prior success of SIFT-like methods on images, some recent work has been dedicated to compute multi-scale features for surfaces. An intuitive idea is to "flatten" surfaces to 2D images via parameterization, and then compute SIFT features of geometric attributes such as normal [6, 7] and curvatures [8]. The parameterization itself, however, suffers from unpredictable changes of topology and boundary, accompanied by domain cutting and shape distortion. In [2], texture was projected to the tangent plane to locally flatten the surface. This method, however, was designed for surfaces with simple geometric shape such as walls. Purely derived from geometry, some other work constructed scale space directly on the 3D surfaces evolving the scale domain information. In [15], a scale space was formulated via surface variation on point-sampled

surfaces. Line-type features were extracted by a multi-scale classification operator that smoothes the surface at different scales. In [16], an intrinsic geometric scale space (IGSS) of 3D surfaces was proposed for extracting scale-dependent saliency. Using Ricci Flow, the surface gradually changes its curvature via shape diffusion. This scale space, therefore, is invariant to conformal deformation. For scale space on surfaces, Lee et al. [1] adopted 3D Gaussian convolution of curvature maps to compute mesh saliency. The 3D Gaussian scale space is easy to compute, nevertheless, it is based on Euclidean distance and only feasible for rigid objects. To improve this, a geodesic scale space (GSS) [4] was introduced using geodesic-based Gaussian convolution. The cost of computing geodesics, however, is extremely high as the scale increases. Recently, the concept of scalar fields defined on 3D surface has been proposed [3], which nicely combines the photometric and geometric characteristics together. They proposed a 3D feature detector (MeshDoG) and a descriptor (MeshHoG). The MeshDoG, however, was computed by 3D Gaussian convolution in 1-ring neighborhood, which is computationally redundant in scale space and may vary subject to shape deformation.

Among descriptors of local features on surfaces, the spin-images [17] is perhaps the most widely adopted. It maps a local surface patch to a 2D image by the radial distance and the axial distance. However, its widespread use has been limited only for rigid objects. The 3D shape context [18] has also been proved to be a successful shape descriptor. In [19], a statistical approach was proposed to describe surface features, where the neighbors of a feature were organized by a spiral pathway, and modeled by a Hidden Markov Model. Others, like the LBE-based shape signatures [9, 10], are defined globally and dedicated to describe geometric characteristics only.

3 Hierarchical Scale Space on Deformable Surfaces

3.1 Gaussian Scale Space using Geodesic Metric

Let $\mathcal{S}$ be a surface, a 2D (topological) manifold embedded in $\mathbb{R}^3$, and let $T(V, E, F)$ be an irregular triangular mesh of $\mathcal{S}$ with vertex subset V, edge subset E and face subset F. Irregular meshes have no restriction on the valences of vertices, engendering more flexibility for complex geometric features and topology changes. A scalar field $L(v)$ where $v \in V$, has attributes defined on all vertices, e.g., texture, curvature, normal, heat, density, etc.

We build a scale space of the scalar field on the surface using the Gaussian kernel, given by

$$L(v, \sigma) = G(v, \sigma) * L_0(v), \tag{1}$$

where $L_0(v) = L(v, 0)$ is the initial scalar field, and the scale space $L(v, \sigma)$ varies according to the scale σ. The Gaussian kernel on surface is defined as

$$G(v, \sigma) = \frac{1}{2\pi\sigma^2} \exp(-\frac{g(v)^2}{2\sigma^2}), \tag{2}$$

where σ is the standard deviation, and $g(v)$ is the geodesic from vertex v to the Gaussian center. This is a convolution of a family of isometric embeddings that

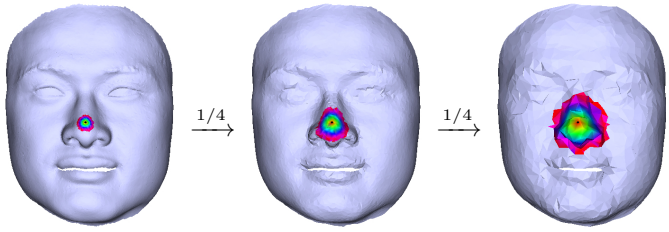


Fig. 2. A pyramid consists of 3 octaves with geodesic neighborhoods of the same radii on different units. It is faster to access large neighbors in higher octave of the pyramid.

preserve geodesic distances from valid neighboring vertices to the center. The discrete Gaussian convolution is then computed in a local region of any v (set as the center)

$$L(v,\sigma) = \frac{\sum_{g(u)<C\sigma} G(u,\sigma)L_0(u)}{\sum_{g(u)<C\sigma} G(u,\sigma)}, \tag{3}$$

where $C\sigma$ is a sufficiently large cut-off (e.g., $C = 2$ in our implementation) and the unit is the average edge length $\overline{e}$.

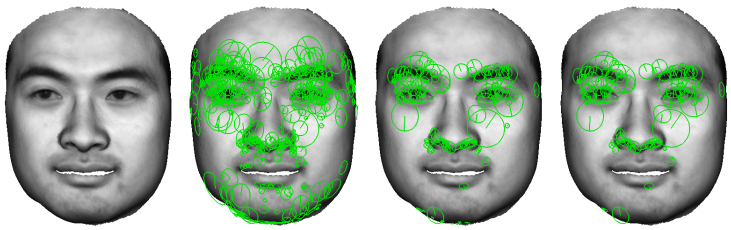
3.2 Pyramid Representation

We propose a hierarchical scale space using geodesic metric. The geodesics on arbitrary triangular meshes can be computed by the fast marching method [20]. It solves the Eikonal equation

$$|\nabla T(x)| = F(x), \tag{4}$$

where the solution $T(x)$ is the shortest time needed to travel from the source to x with $F(x)$ being the time cost. This algorithm has $O(m \log m)$ time complexity for one-source geodesics, where m is the number of traversed vertices. In the discrete Gaussian convolution in Eq. (3), m is related to the area of the neighborhood with radius $C\sigma$, which yields $O(m) = O(\sigma^2)$. Hence, the complexity of computing geodesics for all vertices is $O(n\sigma^2 \log \sigma)$, which is linear w.r.t. the number of vertices n, but quadratic-logarithmic to the scale σ. It implies that the computation cost could be incredibly high when the scale increases.

To address this problem, we introduce a hierarchical scale space using a pyramid representation, which has demonstrated its efficiency in images [11]. A level of the pyramid, or an octave, is obtained by subsampling the previous octave of the pyramid. For triangular meshes, the subsampling can be accomplished by mesh simplification. There is a large amount of literature on mesh simplification using various error metrics. Here we favor approximately uniform sampling during mesh simplification. Therefore, general approaches such as Progressive mesh [21] and QSlim [22] suffice for this purpose. To make sure that the input mesh is approximately uniformly sampled, we investigate the edge lengths,



(a) (b) (c) (d)

Fig. 3. Stages of feature detection: (a) the input scalar field, (b) the initial (494) extrema, (c) selected (150) keypoints by discarding weak responses, and (d) the final (116) keypoints by further removing unstable (i.e., edge, boundary) responses.

and insert vertices on edges with length larger than twice of the $\overline{e}$ using linear interpolation.

The pyramid consists of consecutive octaves $[T_0, T_1, ..., T_O]$, where O is the number of octaves, and T_0 is the original mesh. An octave T_i is subsampled from the lower octave T_{i-1} such that the number of faces $n(F_i)$ is one fourth of $n(F_{i-1})$. The unit of the i-th octave is the average edge length $\overline{e}_i \approx \frac{1}{2}\overline{e}_{i-1}$. Fig. 2 shows a pyramid of 3 octaves. It is faster to access large geodesic neighborhoods in higher octave of the pyramid. Therefore, this computation approach is very efficient. Though the geometry in higher octave is coarser, the lost details are not significant for large scales. Each octave contains S scales. The sampling in the scale space is consistent for all octaves by a constant factor $k = 2^{1/S}$. Besides, the pyramid and geodesics can be computed in the pre-processing stage that may be accomplished by the procedure of triangulation.

4 Feature Detector and Descriptor

4.1 Feature Detector

The entire pipeline of feature detection is highlighted in Fig. 3, with specific processes presented as follows.

Local Extrema. Features are detected by finding extrema in the differences of the scales:

$$\begin{aligned} D(v,\sigma) &= L(v,k\sigma) - L(v,\sigma) \\ &= (G(v,k\sigma) - G(v,\sigma)) * L_0(v), \end{aligned} \tag{5}$$

where $k = 2^{1/S}$ is the factor in scale domain. Assuming the valence of vertex v is n_v, the local extremum of $D(v,\sigma)$ are detected in $(n_v+1) \times (n_v+1) \times (n_v+1)$ neighborhoods of samples. Weak features with small values in $D(v,\sigma)$ are discarded.

Unstable responses. Candidate keypoints with unstable responses (boundaries, edges) are further removed. In regular domain, the edge responses are defined as large ratios of the principal curvatures. In irregular meshes, we can also

apply this strategy using algorithms in [23] that compute curvatures in 1-ring neighborhood. Specifically, we project the 1-ring neighbors $N_1(v)$ of v to its local tangent plane, and use the scalar value as the third dimension to build a new mesh structure. The mean curvature κ_H and Gaussian curvature κ_G on the new mesh are given by

$$\begin{cases} \kappa_H(v) = \frac{1}{4A_{mix}} \left| \sum_{u \in N_1(v)} (\cot \alpha_{vu} + \cot \beta_{vu})(v - u) \right| \\ \kappa_G(v) = \frac{1}{A_{mix}} (2\pi - \sum_{j=1}^{n_v} \theta_j) \end{cases}, \tag{6}$$

where A_{mix} is the area of the generalized Voronoi region for arbitrary meshes, $\cot \alpha_{vu}$ and $\cot \beta_{vu}$ are the well-known "cot" coefficients (please refer to [23] for more details), and θ_j is the angle of the j-th face at the vertex v. The two principal curvatures are then given by

$$\begin{cases} \kappa_1(v) = \kappa_H + \sqrt{\Delta(v)} \\ \kappa_2(v) = \kappa_H - \sqrt{\Delta(v)} \end{cases}, \tag{7}$$

with $\Delta(v) = \kappa_H^2(v) - \kappa_G(v)$.

Refinement. The detected extremum leads to a local region that contains the location of the feature. Thus, the refinement of feature localization is performed to locate the accurate position. To prevent the localization from leaving the surface, we use projected 1-ring neighbors with scalars in the previous stage. A quadratic function on samples of $D(v, \sigma)$ is fitted over spatial and scale domain in the local coordinate system

$$\begin{aligned} D(x, y, \sigma) = a_0 x^2 + a_1 y^2 + a_2 \sigma^2 + a_3 xy + a_4 x\sigma + \\ a_5 y\sigma + a_6 x + a_7 y + a_8 \sigma + a_9, \end{aligned} \tag{8}$$

where $[a_0, a_1, ..., a_9]$ are a group of coefficients that can be estimated by least-square fitting. The localization $\mathrm{x} = [x, y, \sigma]^T$ is updated using

$$\hat{\mathrm{x}} = \mathrm{x} - \frac{\partial^2 D(\mathrm{x})}{\partial \mathrm{x}^2}^{-1} \frac{\partial D(\mathrm{x})}{\partial \mathrm{x}}. \tag{9}$$

4.2 Feature Descriptor

Estimation of Vertex Gradient. For irregular meshes, the gradient of scalar filed $L(v)$ at vertex v, defined as a vector in its local tangent plane, is usually estimated by solving an optimization problem using the finite element method (FEM) [24]. Specifically, let $N_1(v)$ be the 1-ring neighbor of vertex v. The gradient $\nabla L(v)$ can be estimated by minimizing the following error

$$\nabla L(v) = \arg \min_{\nabla L(v)} \sum_{u \in N_1(v)} \left| \nabla L(v)^T P(\overrightarrow{vu}) - \frac{L(u) - L(v)}{g(u)} \right|^2, \tag{10}$$

where $P(\overrightarrow{vu})$ is the projected unit vector of $\overrightarrow{vu}$ in the local tangent plane $T_v\mathcal{S}$.

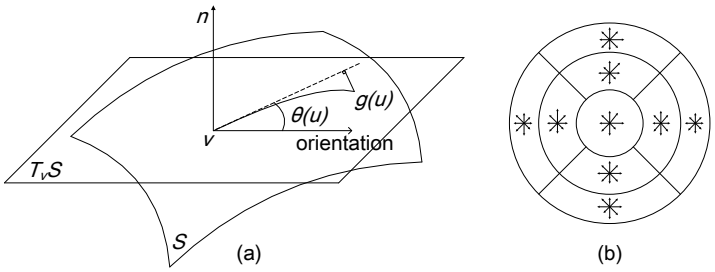


Fig. 4. Local parameterization (a) and descriptor (b). Neighboring vertices are assigned to 9 bins according to their geodesics and polar angles in the local tangent plane T_vS. The histogram of gradients w.r.t. their directions is computed in each bin.

Orientation Assignment. The orientation of the feature is assigned by the dominating direction of gradients in its neighborhood. This assignment makes the detected feature invariant to rotation, which has been a common strategy [11, 3] in computer vision. We divide the local tangent plane into 36 bins, and compute weighed magnitudes of gradients falling into the corresponding bins within a 1.5σ geodesic region. The orientation of the bin with greatest magnitude is assigned as the feature, and a second orientation may also be assigned if there exists a second maximum that is no less than 80% of the highest peak.

Local Parameterization. Previous feature descriptors for curved surfaces (e.g., Spin-images[17], MeshHoG[3]) are typically statistical characteristics distributed in 3D. We propose a new descriptor for deformable surfaces based on a local bivariate parameterization enabled by geodesics and polar angles. Intuitively speaking, any given vertex in a 2D manifold has a neighborhood which is homeomorphic to an open set of a 2D plane. Thus, we parameterize the local region of the vertex v by a polar coordinate system $[g(u), \theta(u)]$ on the surface, where $g(u)$ is the geodesic from vertex v to u, and $\theta(u)$ is the projected polar angle of u from the orientation of v in the local tangent plane $T_v\mathcal{S}$. As shown in Fig. 4(a), this parameterization is completely local, which encodes the geodesic of the destination on the surface and the direction projected in the local tangent plane. Comparing with other local parameterization [25], our method preserves geodesic distances from all neighbors to the origin, and is easy to compute.

Descriptor. A possible drawback of this local parameterization is that the projected polar angles may change subject to severe deformations. To reduce the dependence on the polar angle θ, we quarter the angles in the tangent plane as shown in Fig. 4(b), which affords our shape descriptor to be invariant to most isometric deformations. We use polar grid to cluster vertices into 9 bins, and use tri-linear interpolations to reduce boundary effects. The histogram of gradients w.r.t. eight directions is computed in each bin. The radii of three circles are subject to the ratio of $1 : \sqrt{5} : 3$ with unit σ, so each bin has the same area. The magnitudes of gradients are smoothed by a Gaussian function with 3σ. This descriptor converts histograms of gradients in 2D to a 1D vector that is normalized for matching purpose.

Table 1. Comparison of SIFT-like methods on surfaces.

Methods	Grid	Scale Sample	Pyramid	Descriptor	Deform. Invariance
SIFT	regular	$\sigma_0 2^{o+s/S}$	yes	2D HoG	no
GSS	irregular	$\sigma_0 k^n$	no	Spin-images	no
MeshDoG	regular	$\sigma\sqrt{n}$	no	3D HoG	no
ours	irregular	$\sigma_0 2^{o+s/S}$	yes	2D HoG	yes (isometric)

4.3 Discussion

Most recent multi-scale methods of feature detection on surfaces are inspired by the SIFT on images. It is therefore valuable and illustrative to compare them to clarify their differences, as shown in Table 1. Our method is the most similar one in spirit to the SIFT, and is invariant to isometric deformation. Although all of these methods can produce multi-scale representation of scalar inputs, we found that the tight coupling of samplings in scale and spatial domain is critical to scale-dependent features. The SIFT, which has been shown effective, samples by a constant factor $k = 2^{1/S}$ in the scale space: $\sigma_0 2^{o+s/S}$, where σ_0 is the initial value, S is the number of scales per octave, o and s are the order of octave and scale in the pyramid, respectively. By using the pyramid representation, SIFT increases the sampling intervals by a factor 2 in the spatial domain, which allows us to find extrema in larger neighborhoods for greater scales. The GSS method also samples by a constant factor k in the scale space, while the sampling interval in spatial domain remains unchanged for all scales. The MeshDoG method, which builds the scale space by repeatedly convolving the kernel with the same scale σ, samples in the scale domain as $\sigma\sqrt{n}$. The sampling rate gets smaller when the scale increases, so it generate many more redundant samples, while the sampling rate in spatial domain remains unchanged. Our method has similar sampling strategy to that of SIFT, which makes itself both effective and robust.

5 Experimental Results

In this section, experiments are conducted to evaluate the efficiency and reliability of different methods. The deformable surface data being used are 3D scans from [26] acquired by real-time scanners using structured light, and spacetime faces [27] captured by synchronized video cameras and structured light projectors. Previous methods such as the GSS [4] and the MeshDoG/MeshHoG [3] that fall into the same category as our method are used for the comparison purpose.

5.1 Efficiency Evaluation

We run the three methods on three kinds of data with different resolutions and geometric characteristics: complete model, spacetime face, and 3D scan, and show their results in Fig. 5. We use Gaussian curvature maps for scalar fields (a) of data 1 and 2, and photograph for data 3. The scales of detected features

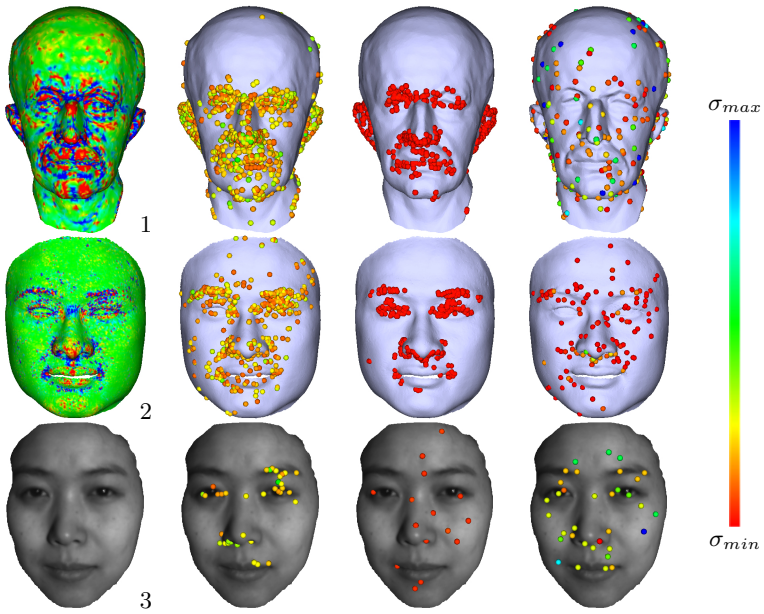


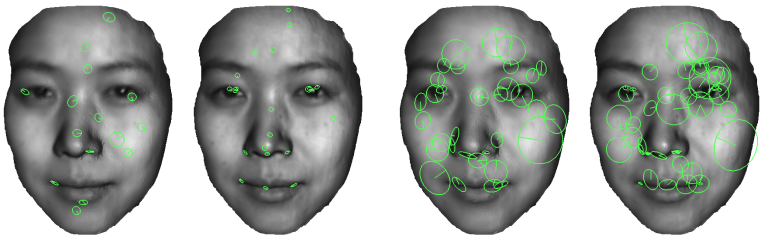
Fig. 5. Scale-space features on three kinds of data: (a) input scalar fields; (b) results by the GSS; (c) results by the MeshDoG; and (d) our results in this paper. We use Gaussian curvature maps for data 1 and 2, and photograph for data 3. Different colors of detected features represent their scales (better viewed in color).

are represented by their colors: blue for large scale, red for small scale. The GSS method (b) contains 32 scales, and the most features detected belong to small and medium scales. We compute 80 scales in the MeshDoG method (c), while the features only show up in the very small scales. For the data used in our experiments, the two methods have redundant samples in spatial domain and scale domain, indicated by the dense spheres and their intensively distributed colors respectively. Our method (d) has 4 octaves (12 scales) for data 1 and data 2, and has 3 octaves (9 scales) for data 3. The detected features by our method appear to be more intuitive in the sense of scale. We also noticed that the geometric features are related to the resolutions of data. It tends to find more large-scale features for coarse meshes (e.g., data 1), and more small-scale features for fine meshes (e.g., data 2). This fact results from the discrete computation of Gaussian curvatures.

The computation time for the three methods in this experiment are shown in Table 2, obtained from a PC with Quad 2.66GHz CPU and 4GB RAM. In the pre-process (pre), we compute geodesics for the GSS, and geodesics and pyramids for our method. And in the running-process (run) we compute feature detectors and descriptors for all three methods. Compared with the GSS that also uses geodesic metric, our method significantly reduces the pre-process time.

Table 2. Computation time of three methods in the experiment of Fig. 5

Methods		**Data 1** (47.1k faces)			**Data 2** (46.9k faces)			**Data 3** (10k faces)		
	scale	feature	pre-(s)	run-(s)	feature	pre-(s)	run-(s)	feature	pre-(s)	run-(s)
GSS	32	1127	5893.9	11.7	478	6252.0	5.7	38	970.9	1.2
MeshDoG	80	540	N/A	31.8	451	N/A	22.2	15	N/A	1.9
ours	12/9	274	47.5	1.8	128	33.7	0.8	36	6.3	0.2



$10k$ faces $32k$ faces $10k$ faces $32k$ faces

Fig. 6. Photometric features over mesh changes and different resolutions. The results by the MeshDoG (left two) are affected by the resolutions. Our results (right two) are more stable under mesh changes and different resolutions.

The MeshDoG only uses 1-ring neighbors to compute Gaussian kernel for all scales, which is actually a bad practice though it saves the pre-process time. In terms of running-process time, our method is still much faster. In fact, for rigid models we can use Euclidean distances to replace geodesics, this will reduce computation time in the pre-process (e.g., about 10 seconds pre-process time for data 1, and the total computation time is about 1/3 of that in the MeshDoG).

Another experiment is conducted to examine the scales of photometric features over mesh changes and different resolutions, with the results shown in Fig. 6. The original scan from [26] has high resolution ($170k$ faces) and moderate accuracy. We downsample them into scans with different resolutions ($10k$ faces and $32k$ faces). The MeshDoG (left) only finds features in the small scales, and thus the results are unstable under resolution changes. Our method (right) is more stable under mesh changes and different resolutions.

5.2 Reliability Evaluation via Feature Matching

The reliability of features for matching purpose can be evaluated by feature matching with ground truth, as in [12, 13]. Thus, we use the spacetime faces which have ground truth correspondences to evaluate the feature descriptors via feature matching. In this experiment, we use the photometric scalar fields, and evaluate two descriptors: MeshHoG and ours. Since the descriptor employed in the GSS is the Spin-images for geometric features on rigid models only, it is not appropriate for our purpose. The matching strategy is a basic nearest neighbor algorithm with cross validation as in [3]. A pair of candidate matches $\{f_i^1, f_j^2\}$ from $\mathcal{S}^1$ and $\mathcal{S}^2$ is identified if they are the nearest neighbor to each other in the

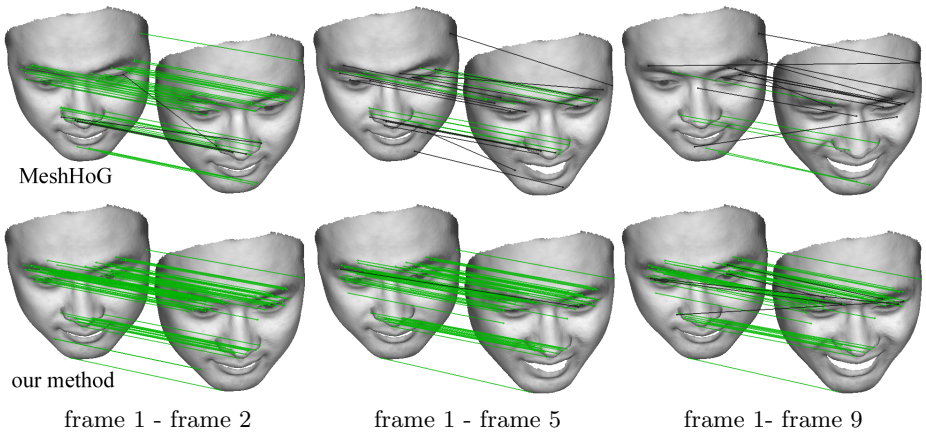
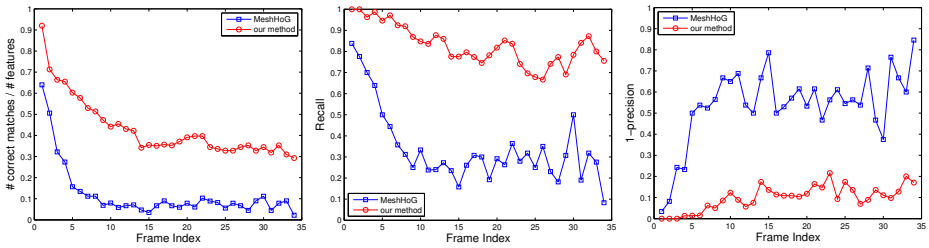


Fig. 7. Selected frames from our evaluation of the MeshHoG (first row) and our method (second row). We match frame 1 to all other frames. Green lines represent correct matches, while black lines represent false matches

Euclidean space of descriptor. And only matches that are discriminative (i.e., the distance of the closest neighbor is γ or less to that of the second-closest neighbor) can be finally accepted.

The spacetime face data has a sequence of a deforming face, where 34 frames are selected for evaluation which last about 5 seconds. We fix thresholds for the two methods respectively, which maintain about ($120 \sim 160$) features detected for each frame. Then we match features in the first frame to the ones in other frames to evaluate the reliability of feature descriptors under deformation. The parameter for matching algorithm is set as $\gamma = 0.9$ for both methods. Some selected frames are shown in Fig. 7, where green lines represent correct matches while black lines represent false matches. The complete result (up to 50 frames) can be found in the supplementary material (in the interest of space). Three criteria are employed: effectiveness, recall, and 1-precision. The effectiveness is defined as the ratio between the number of correct matches and the number of detected features, which reflects how effective for the method to find correct matches. The correct matches are identified if the matched feature is within one unit of $\overline{e}$ of the ground truth. The recall and 1-precision are defined conventionally as that in [13]. The number of positives is determined for the detected features using the same way as correct matches. The evaluation results are shown in Fig. 8. The MeshHoG performs well for the first two frames with small deformation, and its performance severely deteriorates for large frame indices (i.e., large deformation) in our experiment. This is primarily because the detected features by the MeshDoG are not distinctive, and the MeshHoG is not deformation-invariant. Our method is more reliable in terms of high effectiveness, recall, and low 1-precision, and is stable even for large deformations.



(a) effectiveness (b) recall (c) 1-precision

Fig. 8. Evaluation results on reliability of descriptors. The performance of the MeshHoG (blue) severely deteriorates for large deformation, while our method (red) is more stable in terms of high effectiveness, recall, and low 1-precision both consistently and simultaneously.

6 Conclusion

In this paper we have detailed an efficient method to compute scale dependent features on surfaces for deformable shape correspondences, which is a natural generalization of the SIFT. The proposed feature detector and descriptor are invariant to isometric deformation. Unlike previous methods on rigid surfaces, our method takes the 3D scans as scalar fields on deformable manifolds using geodesic metric. By employing a hierarchical scale space and a pyramid shape representation, our method is both efficient and stable, as shown in the experimental results. We have also conducted the comprehensive evaluation of the reliabilities of descriptors via matching features on a sequence of deforming surfaces with ground truth correspondences. Compared with existing work, our method is much more robust and effective under natural deformations. Our on-going research efforts will continue to center on the comprehensive studies on shape matching and registration of deformable surfaces, with new research directions including shape completion in both temporal and spatial domains.

Acknowledgments. This research has been supported in part by NSF grants IIS-0949467 and IIS-0710819. We also wish to thank the University of Washington Graphics and Imaging Laboratory to provide the “spacetime faces” data.

References

1. Lee, C.H., Varshney, A., Jacobs, D.W.: Mesh saliency. ACM Trans. Graph. (TOG) **24** (2005) 659–666
2. Wu, C., Clipp, B., Li, X., Frahm, J.M., Pollefeys, M.: 3d model matching with viewpoint-invariant patches (vip). In: CVPR. (2008)
3. Zaharescu, A., Boyer, E., Varanasi, K., Horaud, R.: Surface feature detection and description with applications to mesh matching. In: CVPR. (2009)
4. Zou, G., Hua, J., Dong, M., Qin, H.: Surface matching with salient keypoints in geodesic scale space. Computer Animation and Virtual Worlds **19** (2008) 399–410

5. Elad, A., Kimmel, R.: On bending invariant signatures for surfaces. TPAMI **25** (2003) 1285–1295
6. Novatnack, J., Nishino, K.: Scale-dependent 3d geometric features. In: ICCV. (2007)
7. Novatnack, J., Nishino, K.: Scale-dependent/invariant local 3d shape descriptors for fully automatic registration of multiple sets of range images. In: ECCV. (2008)
8. Hua, J., Lai, Z., Dong, M., Gu, X., Qin, H.: Geodesic distance-weighted shape vector image diffusion. TVCG **14** (2008) 1643–1650
9. Reuter, M., Wolter, F.E., Peinecke, N.: Laplace-beltrami spectra as "shape-dna" of surfaces and solids. Computer-Aided Design (CAD) **38** (2006) 342–366
10. Rustamov, R.: Laplacebeltrami eigenfunctions for deformation invariant shape representation. In: Symposium of Geometry Processing (SGP). (2007)
11. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV **60** (2004) 91–110
12. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. TPAMI **27** (2005) 1615–1630
13. Ke, Y., Sukthankar, R.: Pca-sift: A more distinctive representation for local image descriptors. In: CVPR. (2004)
14. Tola, E., Lepetit, V., Fua, P.: A fast local descriptor for dense matching. In: CVPR. (2008)
15. Pauly, M., Keiser, R., Gross, M.: Multi-scale feature extraction on point-sampled surfaces. Comput. Graph. Forum (CGF) **22** (2003) 281–289
16. Zou, G., Hua, J., Lai, Z., Gu, X., Dong, M.: Intrinsic geometric scale space by shape diffusion. TVCG **15** (2009) 1193–1200
17. Johnson, A.: Spin-images: A representation for 3-d surface matching. PhD thesis, Carnegie Mellon University (1997)
18. Frome, A., Huber, D., Kolluri, R., Bülow, T., Malik, J.: Recognizing objects in range data using regional point descriptors. In: ECCV. (2004)
19. Castellani, U., Cristani, M., Fantoni, S., Murino, V.: Sparse points matching by combining 3d mesh saliency with statistical descriptors. In: Eurographics. (2008)
20. Kimmel, R., Sethian, J.A.: Computing geodesic paths on manifolds. In: Proceedings of National Academy of Science USA (PNAS). (1998) 8431–8435
21. Hoppe, H.: Progressive meshes. In: SIGGRAPH. (1996) 99–108
22. Garland, M., Heckbert, P.S.: Surface simplification using quadric error metrics. In: SIGGRAPH. (1997) 209–216
23. Meyer, M., Desbrun, M., Schrder, P., Barr, A.H.: Discrete differential-geometry operators for triangulated 2-manifolds. VisMath (2002)
24. Meyer, T.H., Eriksson, M., Maggio, R.C.: Gradient estimation from irregularly spaced data sets. Mathematical Geology **23** (2004) 693–717
25. Shapira, L., Shamir, A.: Local geodesic parametrization: An ants perspective. In: Mathematical Foundations of Scientific Visualization, Computer Graphics, and Massive Data Exploration. Springer Berlin Heidelberg (2009) 127–137
26. Wang, S., Gu, X., Qin, H.: Automatic non-rigid registration of 3d dynamic data for facial expression synthesis and transfer. In: CVPR. (2008)
27. Zhang, L., Snavely, N., Curless, B., Seitz, S.: Spacetime faces: High-resolution capture for modeling and animation. In: SIGGRAPH. (2004)