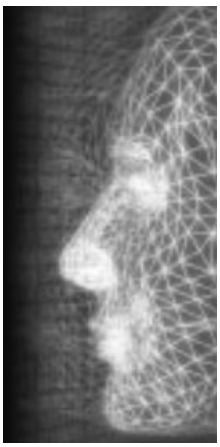
COMPUTER ANIMATION AND VIRTUAL WORLDS
Comp. Anim. Virtual Worlds 2008; **19**: 399–410
Published online 6 August 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/cav.244

Surface matching with salient keypoints in geodesic scale space



By Guangyu Zou, Jing Hua, Ming Dong and Hong Qin*

This paper develops a new salient keypoints-based shape description which extracts the salient surface keypoints with detected scales. Salient geometric features can then be defined collectively on all the detected scale normalized local patches to form a shape descriptor for surface matching purpose. The saliency-driven keypoints are computed as local extrema of the difference of Gaussian function defined over a curved surface in geodesic scale space. This method can properly function on either manifold or non-manifold surface without resorting to any surface mapping or parameterization procedures. Therefore, it has a wide utility in many applications such as shape matching, classification, and recognition. Our experiments on 3D shapes demonstrate that the salient keypoints and local feature descriptors are robust and stable to noisy input and insensitive to resolution change. We have applied our technique to the tasks of 3D shape matching, and the experimental results showed good performance and the effectiveness of this new method. Copyright © 2008 John Wiley & Sons, Ltd.

Received: 25 June 2008; Accepted: 26 June 2008

KEY WORDS: surface matching; scale space; saliency

Introduction

Three-dimensional (3D) surface matching is a fundamental research issue in computer graphics and computer vision useful for many applications such as shape registration,[1,2] motion capture, animation synthesis, partial scan alignment, 3D object recognition, and 3D shape retrieval.[3–5] Generally, the crux of 3D shape matching is to find out good shape representations, allowing us to match two surfaces by comparing their shape representations. Compared with shape matching between 2D images, one of the essential differences is that 3D surface models typically lack a regular and uniform parametrization domain. Therefore, the extensions of 2D algorithms into 3D are oftentimes non-trivial. Recently, geometric mapping-based techniques such as spherical mapping and conformal mapping[1,6] are popularly used to map 3D surfaces onto parametric domains for subsequent analyses, as conducted in 2D images. However, these methods may introduce inevitable, large distortion when mapping large, complex, and topologically complicated surfaces to a canonical domain. This unwanted distortion changes the characteristics of actual 3D shapes and affects the performance in shape matching or other shape analysis. In addition, it may be very difficult to compute parameterization of certain high-genus surfaces. Generally, these methods do not perform well when processing large-scale models.

While a straightforward approach is to resort to mesh simplification techniques for a compromise between geometric accuracy and computational cost, there is another alternative which focuses only on a set of selected salient sub-parts that are descriptive enough to well characterize the whole shape. In our framework, the essential motivation to incorporate the concept of saliency into geometric processing is to derive a more compact shape representation. Apparently, it is very important that the saliency needs to be identified at all scales in order to represent all the shape characteristics. Naturally, scale-space processing seems a viable solution to this end. Scale-space theory is popularly used to handle image structures at different scales, in which the image is represented as a sequence of smoothed images, parameterized by the size of the smoothing kernel used for suppressing fine-scale structures.[7] Recently,

*Correspondence to: J. Hua, Graphics and Imaging Laboratory, Wayne State University, 5143 Cass Avenue, 431 State Hall, Detroit, MI 48202, USA. E-mail: jinghua@wayne.edu

scale-space processing has been introduced to identify mesh saliency[8] and multiscale geometric features.[9] However, how to generalize it as a stable, robust shape descriptor remains an open problem.

Based on the psychological finding that human perceive and recognize real-world objects through different structures or features at different scales,[10] visual data are more reasonable to be represented as a decomposition across multiple scales, particularly when the appropriate scale for a structure of interest can not be determined *a priori*. Toward this end, we present a novel non-parametric method based on a construction of geodesic scale space. The main merits of our approach are that it can directly identify stable, representative, salient keypoints on surfaces by computing local extrema of the Difference-of-Gaussian function defined over the curved surfaces in geodesic scale space. Scale invariant geometric features can then be extracted to represent and characterize the scale-normalized local patches whose sizes are defined by the scale selection. Collectively, they form a stable, scalable, and compact shape representation which can effectively support matching and comparison of large-scale surfaces. This method directly works on manifold or non-manifold surface without any surface mapping or parametrization procedure. Therefore, it has better scalability, facilitating a wider use in applications. Our experiments on 3D shapes demonstrate that the salient keypoints and local feature descriptors are robust and stable to noisy input and insensitive to resolution change. We have successfully applied our approach in face matching, surface alignment, and stitching.

Related Work

Our work is related to the research work in geometric saliency and the scale-space theory. This section reviews some closely related work in these fields, from which more complete literature can be found.

Geometric Saliency

Recently, Gal *et al*. proposed the construction of salient geometric features for a range of shape comparison applications, in which the saliency grade was defined as a summation of the relative curvature to their surroundings and the variance of curvature values in a clustered patches.[4] In References [8,11], low-level human visual attention was modeled by a center-surround operator on Gaussian-weighted mean curvatures in a scale-dependent manner, and applied to mesh simplification and optimal view selection. Most recently, mesh saliency was also implemented through Morse theory by Liu *et al*.[12] Salient critical points were extracted for topological simplification and hierarchical representation of 3D shapes. Motivated by the success of such mechanisms applied in the 2D scenario[7,10] as well as a few limited extensions to 3D surfaces, visual saliency has been paid increasing attention in the fields of geometry processing and shape understanding. One of our goals in this paper is to further explore effective methods for computing geometric saliency which can be used to establish shape correspondence in 3D. Ideally, extracted saliency should be the same as the ones involved in the human perception.

Scale Space

This scale-space processing has been successfully applied to a variety of topics in computer vision including feature detection, computation of shape cues, and object recognition.[7,10] Because similar issues regarding the scale variation of geometric structures are often observed in 3D shape analysis as well, extensions of basic scale-space theory onto 3D shapes have received increasing research interests in recent years. Based on the geodesic curvature flow, a scale space for images painted on surfaces was formulated in Reference [13]. However, it only dealt with surfaces that were given as parameterized functions on a bounded domain. Li *et al*. employed a scale space representation in the detection of feature point for point-based surface alignment.[14] The scale-space was generated by iteratively smoothing the original shape itself with increasing scales, in which there is no clue how the behavior can be quantitatively controlled. Similarly, a scale-space representation defined as different levels of smoothness for the point-based surface was given in Reference [9], focusing its applications on multiscale surface editing. In Reference [15], the surface is first parameterized onto a 2D plane, and then the scale space is built upon a regularly sampled normal map in a manner similar to 2D scale-space processing.[7] The geodesic scale space we propose in the following section is formulated as a sequence of the geodesic Gaussian smoothing directly on the surface attributes. Similar method is used for salient keypoint detection in Reference [16]. However, the Gaussian filtering is directly applied to the vertex coordinates in 3D, which is fundamentally different from our method.

DOI: 10.1002/cav

Geodesic Scale Space

In order to define a scale space, it is of utter importance that new structures must not be introduced from a fine scale to any coarser scale. Suppose L is a time-varying function defined on a certain domain. In general, the diffusion equation

$$\partial_t L = c\nabla^2 L,$$

provides a unified framework for the construction of continuous and discrete scale spaces with initial condition $L(x, y, 0) = f(x, y)$,[10,13] where c is the diffusion factor. In the 2D case, it has been shown that it is equivalent to successively convolve the image by a Gaussian kernel of increasing standard deviation.[10] With further requirements regarding linearity and shift invariance, the Gaussian kernel as well as its derivatives are the only possible smoothing kernels.[10] Therefore, in this paper we use the Gaussian smoothing and geodesic distance metric to construct a scale space of the geometric properties on the surface.

The surface geometry in 3D can be represented as multiple intrinsic properties defined on the domain.[15,17] Features of different scales can thus be extracted from a multi-scale representation of certain property maps. We regard geometric attribute maps as a set of function spaces defined on a two-manifold domain. Let M denote a two-manifold. The function $f : M \to \mathbb{R}^n (n = 1, 2, \ldots)$ defines a map. Different measurements of surface geometry, i.e., mean curvature, Gaussian curvature, and maximal/minimal curvature ratio, and the normals, are therefore modeled as surface maps.

In this paper, a scale space is defined on a curved surface based on geodesic distance, namely the *geodesic scale space* (GSS). Suppose $M(\rho, \theta)$ is a local patch of surface S, where ρ is the geodesic radial distance from the central point, and θ is the angular coordinate with respect to a preassigned tangential direction. $\mathcal{P}(\rho, \theta)$ denotes the certain surface property residing at $M(\rho, \theta)$. The scale space is locally defined as a function, $L(\rho, \theta, \sigma)$, which is produced from the convolution of a variable-scale Gaussian, $G(\rho, \theta, \sigma)$, with $\mathcal{P}(\rho, \theta)$

$$L(\rho, \theta, \sigma) = G(\rho, \theta, \sigma) * \mathcal{P}(\rho, \theta),$$

where

$$G(\rho, \theta, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\rho^2/2\sigma^2},$$

and $*$ is the convolution operation, formulated as

$$(f * g)(\rho, \theta) = \int_0^{\pi} \int_{-\infty}^{+\infty} f(\tau, \phi) g(\rho - \tau, \theta - \phi) d\tau\, d\phi.$$

Note that the integration of Gaussian kernel on a curved domain is usually larger than 1. To eliminate the unwanted scaling effect from the Gaussian convolution and restrict the computation in a local area, in practice, we set cut-off for the Gaussian filter at twice of the standard deviation and normalize the convolution by the integration of the corresponding Gaussian kernel. The actual computation is performed using

$$L(\rho, \theta, \sigma) = \frac{\int_0^{\pi} \int_{-2\sigma}^{+2\sigma} \mathcal{P}(\tau, \phi) G(\rho - \tau, \theta - \phi) d\tau\, d\phi}{\int_0^{2\pi} \int_0^{2\sigma} G(\rho, \theta, \sigma) d\rho\, d\phi}.$$

Hence, the generated scale space is invariant to bending of the surface. In addition, we compute the geodesic distance on triangulated manifolds by solving the Eikonal equation

$$|\nabla T| = F,$$

in a fast marching manner as shown in Reference [18], where the propagating speed F is assigned to a constant, e.g., 1.

Without prior knowledge of the appropriate scales for each potential feature, a scheme for the automatic scale selection and scale invariant feature detection is necessary. In our setting, the scale is selected based on the local extrema over scales of the scale-normalized Laplacian of Gaussian, $\sigma^2\nabla^2 G$, in a geodesic manner. All the discrete approximation are explained in detail in the following sections. Figure 1 shows the geodesic Gaussian smoothing of the curvature map of a face model with scales of σ_{2i}.

Salient Keypoint Detection

Our approach searches local extrema in geodesic scale space to define the salient points on surfaces. These points are further pruned down by applying a threshold based on contrast and anisotropy to eliminate pseudo-salient points. Scale invariant features are then extracted from the detected regions for surface matching. In this section, we will detail each step for the detection of salient keypoints on 3D surfaces. Since the triangle mesh is the most popular approximation of the continuous

DOI: 10.1002/cav

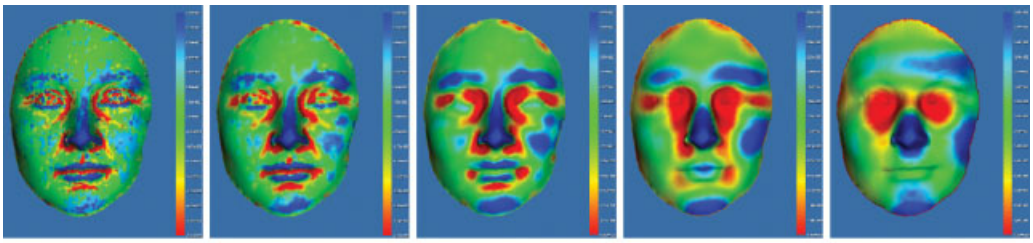


Figure 1. Geodesic Gaussian smoothing of the curvature map of a face model with scales of σ_{2i}. (a) $\sigma_2 = 0.84$ (b) $\sigma_6 = 1.85$ (c) $\sigma_{10} = 4.08$ (d) $\sigma_{14} = 9.01$ (e) $\sigma_{18} = 19.89$.

surface, we present our method in this discrete form. In the remainder of this paper, no distinction will be made among terms: shape, surface, and mesh, unless where it is noted. A few relevant notations are given as follows: M denotes a mesh, v a vertex of M, e an edge of M, $N_1(v)$ the 1-ring neighborhood of v, and $f(v)$ a certain attribute associated with $v \in M$.

GSS-Extrema Detection

It has been found that extrema of normalized Laplacian-of-Gaussian (LoG) produce the most stable salient features.[10] Since Difference-of-Gaussian (DoG) function can provide a close approximation to the scale-normalized LoG,[7] the salient keypoints can be detected by searching the local extrema in both spatial and scale domains in the DoG function. In order to efficiently detect the stable keypoints in the GSS, we first discretize the scale by a constant multiplicative factor k and search for the extrema of the DoG function in the discretized space,

$$\begin{aligned} D(\rho, \theta, \sigma) &= (G(\rho, \theta, k\sigma) - G(\rho, \theta, \sigma)) * \mathcal{P}(\rho, \theta) \\ &= L(\rho, \theta, k\sigma) - L(\rho, \theta, \sigma) \end{aligned} \quad (1)$$

Let $\mathcal{G}(v, q)$ denote the geodesic distance from vertex v to q. At a given scale σ_i, the neighborhood of v is defined as

$$N(v, \sigma_i) = \{q | \mathcal{G}(v, q) < \sigma_i\} \quad (2)$$

If the attribute value associated with v at scale σ_i is larger or smaller than any other vertices in $N(v, \sigma_i) \cup N(v, \sigma_{i-1}) \cup N(v, \sigma_{i+1})$, v is selected as a local extremum.

The frequencies of sampling in scale and spatial domains determine the number of detected stable keypoints in matching and recognition tasks. In order to effectively extract keypoints in GSS, here we empirically define an automatic scheme to determine the minimum and maximum scales, as well as the number of scale samples in the range as follows:

$$\begin{aligned} \sigma_{\min} &= \frac{\min\{\|e\| | e \in M\}}{2} \\ \sigma_{\max} &= \left(\sum_{T \in M} area(T)\right)^{\frac{1}{2}} \Big/ 5 \\ \sigma_i &= \sigma_{\min} \cdot t^i \end{aligned} \quad (3)$$

where M denotes the mesh, e and T denote the edge and face, respectively, and t is calculated by $t = \left(\frac{\sigma_{\max}}{\sigma_{\min}}\right)^{\frac{1}{32}}$, which divides the scale axis into 32 logarithmically uniform units from $\sigma_{\min}$ to $\sigma_{\max}$.

Location Refinement of Keypoints

Since the extrema identified as in Section "GSS-Extrema Detection" are computed based on the discretized function, their locations might not be the actual positions of the DoG extrema. Thus, the detected GSS-Extrema $D(v, \sigma_i)$ only indicate that salient keypoints are likely to be around. In order to obtain more precise estimation about the actual location, scale, and the value of the DoG extrema, a quadratic fitting is performed around the neighborhood of each vertex of interest in the geodesic scale space to locally reconstruct the continuous DoG function. Therefore, all the analysis regarding the salient keypoints can be conducted analytically from the locally fitted quadratic function. For a 3D-space-like GSS, a trivariate function is employed for the quadratic fitting. The defined fitting problem can be solved using least squares approach.

In order to ensure that the spans of the local frame in each dimension are close in magnitude, equalizing the computational precision during the fitting process, the $D(v, \sigma)$ in the neighborhood of (v_0, σ_i) needs to be

DOI: 10.1002/cav

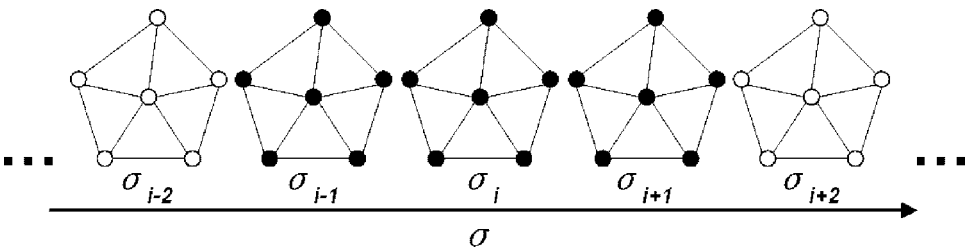


Figure 2. Local sampling frames along the σ axis.

normalized as

$$D(v,\sigma)=\begin{cases}\tilde{D}\left(\frac{v-v_0}{\overline{\|e\|}},\frac{\log\sigma-\log\sigma_i}{\log\sigma_i-\log\sigma_{i-1}}\right) & \text{if } \sigma\in[\sigma_{i-1},\sigma_i]\\ \tilde{D}\left(\frac{v-v_0}{\overline{\|e\|}},\frac{\log\sigma-\log\sigma_i}{\log\sigma_{i+1}-\log\sigma_i}\right) & \text{if } \sigma\in[\sigma_i,\sigma_{i+1}]\end{cases} \quad (4)$$

where $\overline{\|e\|}$ is the average edge length in $N_1(v)$. For the scaled function $\tilde{D}(\cdot,\cdot)$, the vertex of interest has been shifted to the origin with respect to the (x, y, z) coordinate of v. Furthermore, we approximate the neighborhood of v by its tangential plane, which can be parameterized by two independent variables (x, y). Suppose $\mathbf{x}$ and $\mathbf{y}$ are the two orthogonal unit vectors in the tangential plane, both orthogonal to the normal vector $\mathbf{n}$. Given a vertex $w \in N_1(v)$, the map,

$$w \mapsto (x, y) = (< w - v, \mathbf{x} >, < w - v, \mathbf{y} >), \quad (5)$$

projects w onto its tangential plane.

Once the above initialization is done, the fitting is computed in the neighborhood of (v, σ_i) as in a 3D volume, considering the scale σ as another independent dimension with the existence of $\mathbf{x}$ and $\mathbf{y}$. We use $\{(v_k, \sigma_k)|v_k \in N(v, \sigma_k)\}, \sigma_k \in \{\sigma_{i-1}, \sigma_i, \sigma_{i+1}\}$ as the sample points. This sampling frame is illustrated in Figure 2, where the solid dots denotes the vertices used for the local fit. For each point, the parametric coordinate in the tangential space is denoted in the form of (x, y, σ), and the function value is given by $\tilde{D}(v(x, y), \sigma)$, or simply, $\tilde{D}(x, y, \sigma)$. Empirically, the following quadratic prototype gives the best fits, in terms of robustness to the noise as well as the adequacy in capturing the local change of $\tilde{D}(x, y, \sigma)$:

$$\tilde{D}(x, y, \sigma) = Ax^2 + By^2 + C\sigma^2 + Dxy + Ex + Fy + G\sigma + H. \quad (6)$$

Estimating the parameters [A, B, C, D, E, F, G, H]^T is a linear least squares problem, which corresponds to a system of linear equations:

$$\{\tilde{D}(x_i, y_i, \sigma_i) = Ax_i^2 + By_i^2 + C\sigma_i^2 + Dx_iy_i + Ex_i + Fy_i + G\sigma_i + H\}|_{i=0}^{n-1}, \quad (7)$$

where n denotes the number of points in the neighborhood, and i is the index of each point in the neighbor set. The location of the extremum, $(\hat{x}, \hat{y}, \hat{\sigma})$, is consequently determined by taking the partial derivatives of $\tilde{D}(x, y, \sigma)$ and setting them to zero:

$$\begin{aligned}\frac{\partial\tilde{D}(x, y, \sigma)}{\partial x} &= 2Ax + Dy + E = 0\\ \frac{\partial\tilde{D}(x, y, \sigma)}{\partial y} &= Dx + 2By + F = 0\\ \frac{\partial\tilde{D}(x, y, \sigma)}{\partial\sigma} &= 2C\sigma + G = 0\end{aligned} \quad (8)$$

Accordingly, the extremal value is given by $\tilde{D}(\hat{x}, \hat{y}, \hat{\sigma})$ as well. In very rare situations, the Equation (7) may be an underdetermined system. Then, this procedure is simply skipped. The candidate is collected as a final salient keypoint directly.

In some cases, the extrema analytically derived from $\hat{D}(x, y, \sigma)$ may be located closer to the surrounding vertices, other than the originally detected vertices. In this case, the vertex that is closest to the extremum, among the vertices used as sample points, is selected for a new round of extremum location refinement. The pre-detected pseudo-keypoint is discarded. This iterative process terminates when no closer vertex is found.

Removal of Unstable Keypoints

The GSS-based saliency detection as described in Sections "GSS-Extrema Detection" and "Location Refinement of Keypoints" is capable of detecting any scale change of the signal, but not sensitive to the magnitude.

DOI: 10.1002/cav

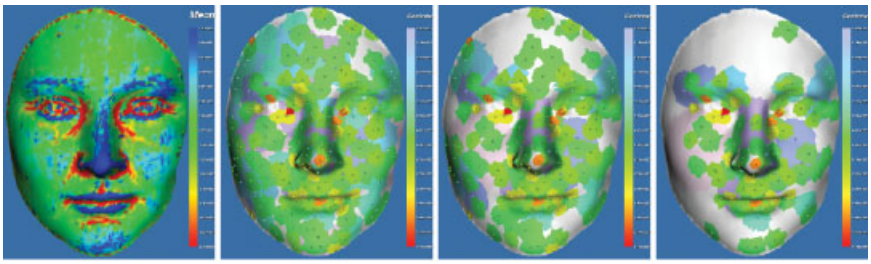


Figure 3. Geometric attributes for saliency detection. (a) Mean curvature (b) Raw (230 pts) (c) Thresholded (171 pts) (d) Refined (106 pts).

Hence, some salient regions with small contrast will be detected. These regions are not stable under noise. Even small amount of add-on noise will change the location of the local extrema. For this reason, in the first step, the local extrema with small magnitudes should be removed. Moreover, DoG has stronger response along the grooves and ridges. Although grooves and ridges can also result in important visual attentions, unfortunately, the automatic scale selection cannot identify their scales correctly.[7] Hence, it is necessary to remove those salient points as well for better surface matching.

Low contrast extrema. Besides the geometry, a number of accidental factors can also produce small perturbations of $D(v, \sigma)$, scattering unpredictably over the GSS. Besides, since the degree of saliency directly depends on the magnitude of $D(v, \sigma)$,[8] the extrema with low contrast within the adjacent scales are not as significant as other keypoints. Thus, the keypoints below a certain threshold are discarded. Suppose that the maximum of the absolute value of $D(v, \sigma)$ is $\|D\|_m$. The threshold is empirically set to $\|D\|_m \cdot 0.1$, for all the experiments in this paper. Figure 3 shows the effects of keypoint selection on a 3D face model with 10 000 mesh faces. Keypoints are rendered as uniform spheres over the associated vertices. Scale information of each keypoint is visualized by a local patch approximating the geodesic neighborhood $\{q|^{\mathcal{G}}(v, q) < \sigma_i\}$, where v is the keypoint at scale σ_i. Furthermore, color mapping is used to differentiate keypoints with different scales. The scale values in the range $[\sigma_{\min}, \sigma_{\max}]$ are logarithmically mapped.

Strong groove/ridge response. In the same manner as described in Reference [7], some keypoints detected along the grooves and ridges of surfaces are also not stable, thus need be rejected. This class of extrema in the difference-of-Gaussian function are usually poorly defined and have a large ratio between the principal curvature across the groove/ridge and the small one in the perpendicular direction. Since we have fitted the analytic form of the difference-of-Gaussian function at each keypoint, this measure can be easily computed from the 2×2 Hessian matrix, $\boldsymbol{H}$, at the specific location and scale:

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} = \begin{bmatrix} 2A & D \\ D & 2B \end{bmatrix} \tag{9}$$

The derivative of H are proportional to the principal curvatures of D. Let α and β be the larger eigenvalue and the smaller one, respectively. Then, we have:

$$\mathrm{Tr}(\mathbf{H}) = D_{xx} + D_{yy} = \alpha + \beta,$$
$$\mathrm{Det}(\mathbf{H}) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta$$

Let $\alpha = r\beta$. Keypoints with $\alpha > r\beta$ are removed. Instead of computing α and β, we can simply check if

$$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} = \frac{(\alpha+\beta)^2}{\alpha\beta} < \frac{(r+1)^2}{r} \tag{10}$$

Figure 3(a) shows the mean curvature map of the original surface; Figure 3(b) shows the all the 230 keypoints detected as extrema of the Difference-of-Gaussian function; after removal of low contrast extrema, the 171 keypoints are retained as shown in Figure 3(c). The final result after eliminating edge responses is shown in Figure 3(d).

Curvature Denoising

The aforementioned salient keypoint detection method works on arbitrary geometric properties. Maps of attribute vectors are also supported within our framework. Curvature maps such as mean curvature, gaussian curvature, and maximal/minimal curvature, and surface normal map, are widely used to describe the

DOI: 10.1002/cav

local geometry of shapes. Surface normals are generally more robust to noises than curvatures as they are the first-order partial derivatives of the raw data. However, curvature-based features are more salient, with proofs from psychophysical experiments that the human visual system decomposes complex shapes into parts based on curvature. Hence, in this paper, we mainly use curvature maps in our experiments. Moreover, since we compute local shape representation over a local neighborhood, the curvature descriptors are not sensitive to noise as demonstrated by the experiments in Experimental Results and Applications section.

There are a number of techniques available that can be used to estimate discrete curvatures on meshes.[19] Generally, techniques based on local fitted patches are more robust to the noise and meshing quality than the ones using discrete differential geometry operators, with a larger expenditure of time and computation. For all experiments here, we calculate the curvature data according to the unified framework proposed in Reference [20]. In order to handle mesh degeneracy, a preprocess of denoising needs to be performed on the surface. Based on the observation that mesh degeneracy always results in extremely high/low estimation of curvatures, we simply clamp the curvatures to 5–95% of the original range.

Surface Matching with Keypoints

After salient keypoints are detected, it is very important to construct local descriptors based on these keypoints and their associated scales.

Local Shape Descriptors

A lot of work has been focused on the search of effective local shape descriptors for determining shape similarity[14,15,22,23] since local shape properties are usually the distinctive features of a certain class. In general, these approaches first define a set of uniformly distributed points on the surface, around which a radial portion of the shape is used as a local support for computing the shape descriptors. A drawback of these approaches is that, since they do not discriminate the degree of distinctiveness (saliency) from region to region of the surface, a large portion of the computation spent on the shape descriptor and subsequent matching is unnecessary. For instance, shape descriptors locating on the flat areas will be almost always matched to multiple flat locations of the target shape and thus can not distinguish shapes since this kind of structure is trivial among most shapes. In addition, these methods are sensitive to the distribution of the points and the predefined size of radial support.

In contrast, our method detects the salient keypoints along with their intrinsic scales of the underlying features. Therefore, local shape descriptors only need to be computed at selected salient locations for the purpose of shape matching. More importantly, since the support sizes are set to be the automatically detected scales of the features, the descriptors are scale-invariant based on the accordingly scaled local frames. Therefore, better performance can be achieved at a mush less cost.

Given a mesh $M = \{V, E, F\}$ and a set of keypoints with their respective scale and normal $\{v, \sigma, \vec{n}\}$, we define a local descriptor which is adapted from the spin image[22] to the setting of our framework. First, an orthogonal local frame $(\vec{u}, \vec{v}, \vec{n})$ is defined for each keypoint. The choice of $\vec{u}$ and $\vec{v}$ is not unique, all of which, however, result in the same spin image. Within this frame, two cylindrical coordinates can be defined in the same manner as shown in Reference [22]: Suppose $\mathcal{P}$ is the tangential plane of keypoint P and $\mathcal{L}$ is the line through P in the normal direction. Given a point Q in 3D, the two coordinates are α, the perpendicular distance between Q and $\mathcal{L}$, and β, the signed perpendicular distance between Q and $\mathcal{P}$. A spin image is created as a 2D accumulator indexed by α and β. For each bin, bilinear interpolation is used to distribute the contribution of the sample point.

Next, different from the original spin image generation, the support size is determined based on the detected scale, σ, of the keypoint, because the optimal coverage of the spin image should be consistent with the intrinsic scale of a specific feature structure. Let R_s denote the radius of the sampling disc centered at P in the tangential plane $\mathcal{P}$, which is computed as

$$R_s = \max(\vec{n} \times (Q - P)),\ Q \in \{q | \mathcal{G}(q, P) < \sigma\} \qquad (11)$$

Given a sampling resolution $M \times N$, we cast rays from a set of points in the disc around P:

$$S_{kl} = P + \frac{lR_s}{M}\left(\cos\left(\frac{2\pi k}{N}\right) \cdot \vec{u} + \sin\left(\frac{2\pi k}{N}\right) \cdot \vec{v}\right) \qquad (12)$$

where $k = 1, \ldots, N$ and $l = 1, \ldots, M$. Let $\tilde{S}_{kl}$ be the intersection point of the local neighborhood and the ray casted from S_{kl}. The cylindrical coordinates $(\alpha_{kl}, \beta_{kl})$ of $\tilde{S}_{kl}$ is used to generate the spin image of P. Empirically,

DOI: 10.1002/cav

we set the sampling resolution to 32×8 and the size of the spin image to 8×8, for experiments shown in this paper. In order to maximize the descriptiveness of the spin image signature, we do not further filter $\{\tilde{S}_{kl}\}$ with the *support angle*.[22] Because the sampling frame is already adaptively scaled according to specific salient features, this descriptor is very robust in our setting as demonstrated in the experiment section.

Keypoint Matching

Since each geometric feature has a spin image representation, matching surface features is actually the matching of the image signatures. In order to establish the correspondence, we compute the correlation coefficient between spin images of the source and the target models:

$$\text{corr}(X, Y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} \tag{13}$$

where X and Y are two spin images, $\bar{x}$ and $\bar{y}$ are the mean values of the image pixels $\{x_i \in X\}$ and $\{y_i \in Y\}$, respectively. When two images are highly correlated, we consider that the corresponding features are likely to be matched.

Moreover, because we have reduced the number of points in the matching stage by only comparing salient keypoints with already-detected scales, the computational efficiency is dramatically improved. Suppose A and B are two sets of keypoints from the source and target objects, respectively. The two maps can be formulated as $f_{A\to B}(\cdot)$ and $f_{B\to A}(\cdot)$. A successful match is established if and only if, for a certain point a from the source, the following condition is met:

$$f_{B\to A}(f_{A\to B}(a)) = a \tag{14}$$

The matched two set of keypoints are bijective, thus symmetric.

Experimental Results and Applications

In this section we analyze the robustness of our proposed 3D surface matching method using salient keypoints on real face data with noise and different facial expressions. Furthermore, we demonstrate the performance of our method through an application in surface alignment and stitching. Our method is implemented on a PC with 3.8 GHz CPU and 2 GB RAM. Computing the restricted geodesic neighbors of each vertex takes most of the time spent through the whole process of salient keypoint detection, which has a complexity of $O(N \log N)$ where N is the number of vertices. For a mesh with modest 5 000 faces, the computation for the geodesic neighborhood costs about 3 minutes, while other steps almost give immediate outputs.

Repeatability of Saliency Detection

Stability of salient keypoints and associated scales is a very important factor in surface matching task. As inferred from the methodology, rotation and scaling will not change the saliency detection. We mainly evaluate the stability of saliency detection under noise. We first add random noise on each vertex of the mesh along the normal direction. Let r denote the average edge length of the mesh. The noise ranges from $-c \cdot r$ to $c \cdot r$, where we set c to be 0, 5, 15, and 25%, respectively, for each case. The percentage of detected common keypoints only decreases slightly, which depends on a pre-assigned theshold. This shows that our method is very robust to the noise in terms of the repeatability of keypoints' locations and scales. Figure 4 shows the actual repeated keypoints on surfaces with different extents of noise.

Face Matching with Expressional Changes

We present experiments in which 3D face matching is performed under expressional changes. In this section we use real faces' scans (10 subjects) to analyze the performance of our proposed 3D shape matching method. For all the subjects with different expressions, our method can correctly match the same subject with different expressions while differentiating different subjects solely based on the number of matched keypoints. Figure 5(a–b) shows a portion of the matched same subject with different expressions and Figure 5(c–d) shows the differentiation between two different subjects, where the number of matched keypoints is nearly null. Note that only matches with high confidence (high correlation coefficients of the corresponding scale-invariant spin images) are selected. A small number of keypoints are descriptive enough to differentiate models from different subjects and, meanwhile, cluster distinct expressions of the same person. The experiments indicate that our method is efficient for shape retrieval and shape recognition. Figure 6 shows the extensive experiments

DOI: 10.1002/cav

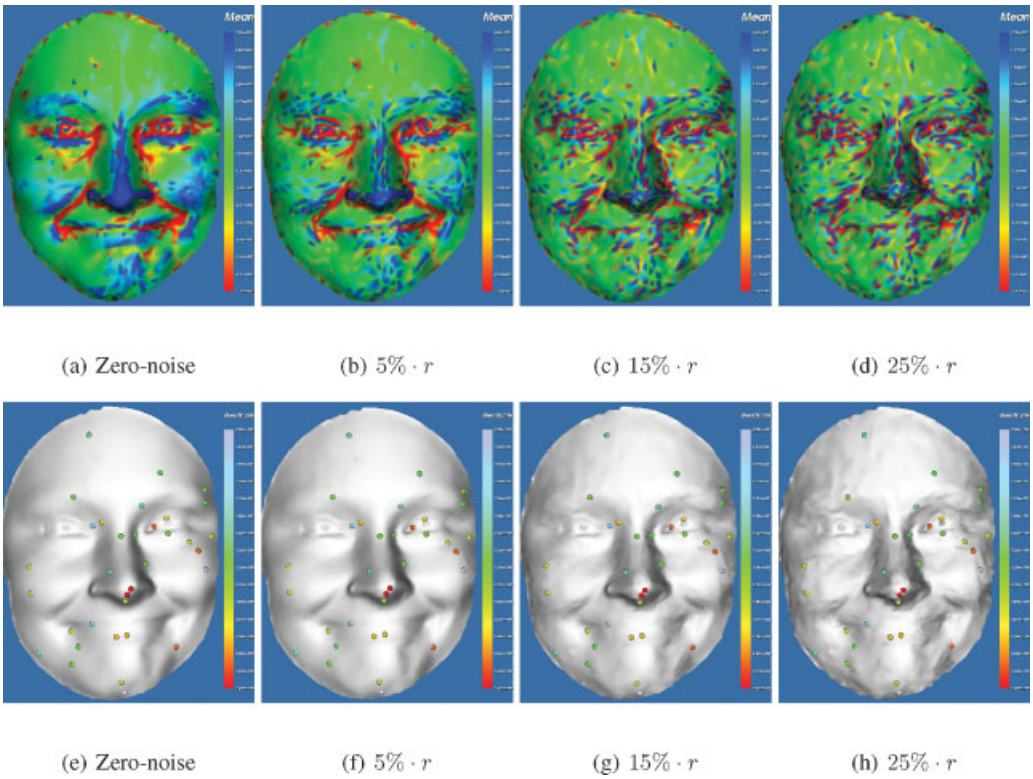


Figure 4. Examples of keypoint detection to random noise. (a)–(d) visualize the noise perturbation by the mean curvature rendering; (e)–(h) show the detected keypoints on each mesh, respectively.

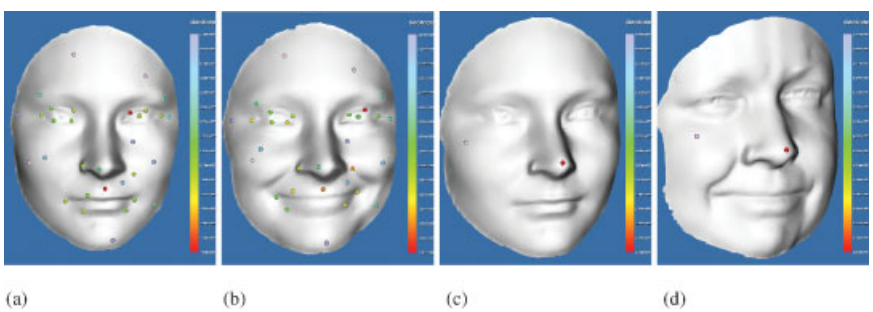


Figure 5. Face matching between inter-expressions and inter-subjects. (a) Express 1 (b) Express 2 (c) Subject 1 (d) Subject 2.

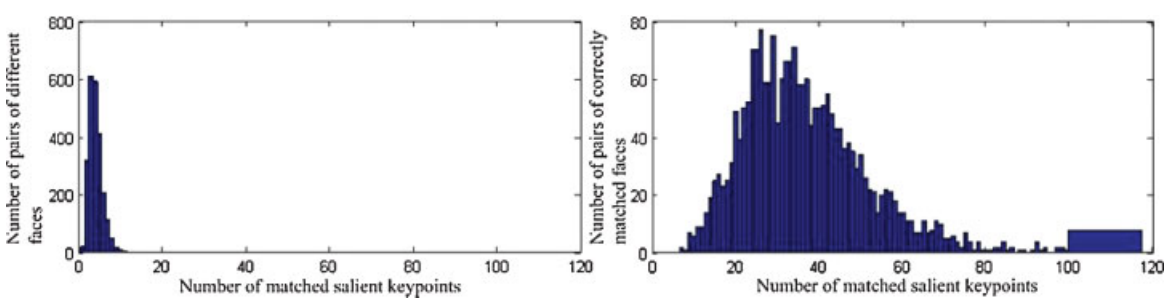


Figure 6. Statistics on face matching between inter-expressions and inter-subjects.

DOI: 10.1002/cav

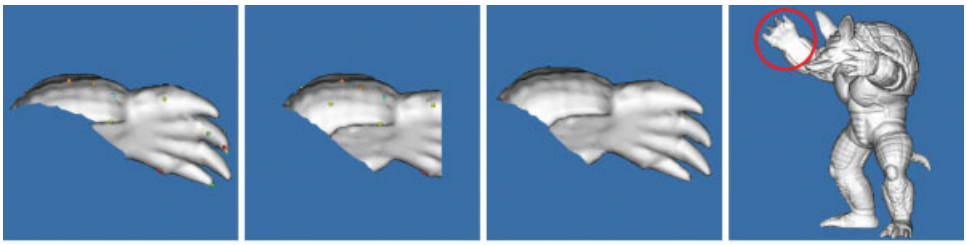


Figure 7. Surface alignment and stitching based on keypoint correspondences. (a) Piece 1 (b) Piece 2 (c) Registered piece (d) Completed model.

that we performed on over 2 000 face scans of 10 subjects with different expressions. The horizontal axis denotes the number of matched salient keypoints, while the vertical axis denotes the number of pairs of different faces on the left and the number of pairs of correctly matched faces on the right, respectively. The statistical results show that for different subjects the number of average matched salient keypoints is only around 5 but for the same subject with different expressions the number of average matched salient points is about 35. There is clear threshold that we can use for recognition under expressional changes. If the number of matched points is smaller than 10, the face scans are estimated as being from different subject. If larger, they are considered being from the same subject.

Surface Alignment and Stitching

In this section, we apply the salient surface keypoints for surface alignment and stitching. A very important property of our method, which supports surface alignment and stitching, is that the salient surface keypoints detected in GSS can directly establish point correspondences for the 3D surfaces. The duplicated regions in the original 3D surfaces can be detected and removed by removing the overlapping areas. After that, we can stitch the 3D surface patches by connecting the exclusive regions directly on the surface. After simple mesh optimization around the joint regions, a complete 3D surface model is obtained. As an example, Figure 7 demonstrates the alignment and stitching of several pieces of 3D surfaces together to form a complete model.

Conclusions

In this paper, we have presented a non-parametric method based on geodesic scale-space processing, which can directly and automatically compute representative salient features on a surface. The scale-invariant keypoints identified in the geodesic scale space make it possible to construct salient features with derived scales, leading to a scalable, compact representation of large-scale surface models for shape matching and comparison. The experiments have demonstrated our feature detection technique and salient shape descriptors are robust and stable to noisy input and insensitive to resolution change. In addition, visual features other than shape characteristics can also be overlaid to the shape representation uniformly, providing a scalable, unified abstraction of multimodality geometric data.

In the future, we plan to use the selected salient keypoints for the construction of Shape Attribute Relational Graph (SARG),[24] in the hope of lifting the local shape descriptors proposed in this paper to a global shape representation based on the geometric saliency.

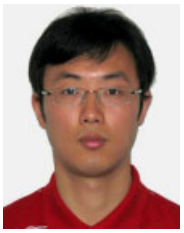
References

1. Wang S, Wang Y, Jin M, Gu XD, Samaras D. Conformal geometry and its applications on 3D shape matching, recognition, and stitching. *IEEE Transaction on Pattern Analysis Machine Intelligence* 2007; **29**(7): 1209–1220.
2. Zou G, Hua J, Muzik O. Non-rigid surface registration using spherical thin-plate splines. In *Proceedings of Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2007; pp. 367–374.
3. Shilane P, Funkhouser T. Distinctive regions of 3D surfaces. *ACM Transaction on Graphics* 2007; **26**(2): 7.
4. Gal R, Cohen-Or D. Salient geometric features for partial shape matching and similarity. *ACM Transaction on Graphics* 2006; **25**(1): 130–150.
5. Liu Y, Zha H, Qin H. Shape topics: a compact representation and new algorithms for 3D partial shape retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006; pp. 2025–2032.

DOI: 10.1002/cav

6. Gu X, Wang Y, Chan TF, Thompson PM, Yau S.-T. Genus zero surface conformal mapping and its application to brain surface mapping. *IEEE Transaction on Medical Imaging* 2004; **23**(8): 949–958.
7. Lowe DG. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 2004; **60**(2): 91–110.
8. Lee CH, Varshney A, Jacobs DW. Mesh saliency. *ACM Transaction on Graphics* 2005; **24**(3): 659–666.
9. Pauly M, Kobbelt LP, Gross M. Point-based multiscale surface representation. *ACM Transaction on Graphics* 2006; **25**(2): 177–193.
10. Lindeberg T. Scale-space theory: a basic tool for analysing structures at different scales. *Journal of Applied Statistics* 1994; **21**(2): 224–270.
11. Yamauchi H, Saleem W, Yoshizawa S, Karni Z, Belyaev A, Seidel H.-P. Towards stable and salient multi-view representation of 3D shapes. In *Proceedings of IEEE Conference on Shape Modeling and Applications*, 2006; pp. 265–270.
12. Liu Y-S, Liu M, Kihara D, Ramani K. Salient critical points for meshes. In *Proceedings of ACM Symposium on Solid and Physical Modeling*, 2007; pp. 277–282.
13. Kimmel R. Intrinsic scale space for images on surfaces: the geodesic curvature flow. *Graphical Models and Image Processing* 1997; **59**(5): 365–372.
14. Li X, Guskov I. Multi-scale features for approximate alignment of point-based surfaces. In *Proceedings of Eurographics Symposium on Geometry Processing (SGP)*, 2005; pp. 217.
15. Novatnack J, Nishino K. Scale-dependent 3D geometric features. In *Proceedings of IEEE Conference on Computer Vision (ICCV)*, 2007; pp. 1–8.
16. Castellani U, Cristani M, Fantoni S, Murino V. Sparse points matching by combining 3d mesh saliency with statistical descriptors. *Computer Graphics Forum* 2008; **27**(2): 643–652.
17. Gatzke T, Grimm C, Garland M, Zelinka S. Curvature maps for local shape comparison. In *Proceedings of International Conference on Shape Modeling and Applications (SMI)*, 2005; pp. 246–255.
18. Kimmel R, Sethian JA. Computing geodesic paths on manifolds. In *Proceedings of National Academy of Science USA*, vol. 95, 1998; pp. 8431–8435.
19. Razdan A, Bae MS. Curvature estimation scheme for triangle meshes using biquadratic bézier patches. *Computer-Aided Design* 2005; **37**(14): 1481–1491.
20. Meyer M, Desbrun M, Schröder P, Barr AH. Discrete differentialgeometry operators for triangulated 2-manifolds. In *Proceedings of Workshop Visualization and Mathematics*, 2002; pp. 35–54.
21. Shilane P, Funkhouser T. Selecting distinctive 3D shape descriptors for similarity retrieval. In *Proceedings of International Conference on Shape Modeling and Applications (SMI)*, 2006; pp. 18.
22. Johnson AE, Hebert M. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 1999; **21**(5): 433–449.
23. Gatzke T, Grimm C, Garland M, Zelinka S. Curvature maps for local shape comparison. In *Proceedings of International Conference on Shape Modeling and Applications (SMI)*, 2005; pp. 246–255.
24. Zhang D-Q, Chang S.-F. Detecting image near-duplicate by stochastic attributed relational graph matching with learning. In *Proceedings of ACM International Conference on Multimedia (MULTIMEDIA'04)*, 2004; pp. 877–884.

Authors' biographies:



Guangyu Zou is a PhD candidate in Computer Science at Wayne State University, where he is also a research assistant in the Graphics and Imaging Laboratory. He received his BE degree (2003) and ME degree (2005), both in Electrical Engineering, from Huazhong University of Science and Technology (HUST), Wuhan, China. His research interests include shape parametrization, geometric flow, shape space, and their applications to graphics, computer vision and medical imaging.

Dr. Jing Hua is an assistant professor of Computer Science at Wayne State University and the director of Graphics and Imaging Laboratory. He received his PhD (2004) in Computer Science from the State University of New York at Stony Brook. He also received his MS degree (1999) in Pattern Recognition and Artificial Intelligence from the Institute of Automation, Chinese Academy of Sciences in Beijing, China and his BS degree (1996) in Electrical Engineering from the Huazhong University of Science and Technology in Wuhan, China. His research interests include Computer Graphics, Visualization, Biomedical Imaging and Informatics. He has published over 60 journals and conference papers in the above areas. His research is currently funded by the NSF, the NIH, and the Michigan State Foundations. Dr. Hua serves as an Editorial Board *Member for Scientific Journals International* and International *Journal of Technology Enhanced Learning* and a Program Committee Member for many international conferences. He is a member of IEEE.

DOI: 10.1002/cav

Ming Dong received his BS degree from Shanghai Jiao Tong University, Shanghai, P.R. China in 1995 and his PhD from the University of Cincinnati, Ohio, in 2001, both in electrical engineering. He joined the faculty of Wayne State University, Detroit, MI in 2002. He is currently an assistant professor of Computer Science and the Director of the Machine Vision and Pattern Recognition Laboratory. His research interests include pattern recognition, data mining, and multimedia analysis. He has published over 60 technical articles, many in premium journals and conferences such as *IEEE Transactions on Neural Networks*, *IEEE Transactions on Computers*, *IEEE Transactions on Fuzzy Systems*, *IEEE ICDM*, *IEEE CVPR*, *ACM Multimedia*, and WWW. He is an associate editor of the Pattern Analysis and *Applications Journal* and was on the editorial board of *International Journal of Semantic Web and Information Systems*, 2005-2006. He also serves as a program committee member for many related conferences.

Hong Qin received the BS and MS degrees in computer science from Peking University, Beijing, and the PhD in computer science from the University of Toronto (UofT) in 1995. Currently, he is a full professor of computer science in the Department of Computer Science at the State University of NewYork at Stony Brook (Stony Brook University). During his years at UofT, he received a UofT Open Doctoral Fellowship. He was also a recipient of the US National Science Foundation (NSF) Faculty Early Career Development (CAREER) Award, the Honda Initiation Award, and the Alfred P. Sloan Research Fellowship by the Sloan Foundation. In 2005, he served as the general cochair for Computer Graphics International 2005 (CGI' 05). Currently, he is an associate editor for the *IEEE Transactions on Visualization* and *Computer Graphics*, and he is also on the editorial board of The *Visual Computer* (*International Journal of Computer Graphics*). In 2007, he is the conference cochair for the ACMSolid and Physical Modeling Symposium. His research interests include geometric and solid modeling, graphics, physics-based modeling and simulation, computer-aided geometric design, human-computer interaction, visualization, and scientific computing. He is a member of the IEEE.

DOI: 10.1002/cav