

# Invariant Surface-Based Shape Descriptor for Dynamic Surface Encoding

Tony Tung and Takashi Matsuyama

Kyoto University, Graduate School of Informatics, Japan  
tung@vision.kuee.kyoto-u.ac.jp, tm@i.kyoto-u.ac.jp

**Abstract.** This paper presents a novel approach to represent spatio-temporal visual information. We introduce a surface-based shape model whose structure is invariant to surface variations over time to describe 3D dynamic surfaces (e.g., obtained from multiview video capture). The descriptor is defined as a graph lying on object surfaces and anchored to invariant local features (e.g., extremal points). Geodesic-consistency-based priors are used as cues within a probabilistic framework to maintain the graph invariant, even though the surfaces undergo non-rigid deformations. Our contribution brings to 3D geometric data a temporally invariant structure that relies only on intrinsic surface properties, and is independent of surface parameterization (i.e., surface mesh connectivity). The proposed descriptor can therefore be used for efficient dynamic surface encoding, through transformation into 2D (geometry) images, as its structure can provide an invariant representation for 3D mesh models. Various experiments on challenging publicly available datasets are performed to assess invariant property and performance of the descriptor.

## 1 Introduction

It is one of the major goals of natural sciences to find invariant properties. In the 90s, computer vision scientists found several projectively invariant properties (e.g., viewpoint, illumination and curvature invariants) to characterize 3D object shape for recognition tasks [1, 2]. As it is difficult to find invariants on general 3D shapes that are not planar (or simple), local descriptors are used as well to model invariants and represent 3D object surface as a collection of small patches [3].

In this paper, we propose a new invariant surface-based shape descriptor for dynamic geometric objects, that is invariant to surface parameterization (e.g., surface mesh complexity or connectivity) and visual features (e.g., texture) as it relies only on intrinsic surface properties and geodesic paths. The descriptor is defined as a graph lying on object surface and anchored to invariant local features (e.g., extremal points). Positions of graph edges and nodes are optimized using a Bayesian probabilistic framework driven by two *geodesic consistency* cues: when surfaces undergo non-rigid deformations over time, the overall graph structure remains invariant to surface variations. We show that the descriptor can be applied for efficient encoding of 3D video data (or free-viewpoint video), which are becoming a popular media [4–7]. Particularly, when each 3D video

frame is reconstructed individually using multiview stereo techniques, the produced 3D surface models have no geometric consistency between each other: vertex number and mesh connectivity are different. It is then not trivial to find an optimal encoding scheme for the data structure. Moreover, as no adaptive resolution streaming mechanism exists for 3D video, communication and telepresence applications are still tedious on low-bandwidth networks. Although 3D video data can be post-processed to obtain meshes with consistent topology and connectivity [8–11], how to cope with geometry variations is still unclear (e.g., when the mesh resolution has to dynamically change). Here, inspired by [12], we propose to use the new invariant surface-based shape descriptor as cut graphs that cut open surface meshes for parameterization into a square domain. As the cut graphs are invariant regardless of mesh resolution, 3D video data can be transformed into sequences of 2D (geometry) images that are suitable for any 2D video encoding technology (e.g., MPEG-4). Related work is discussed in Sect. 2. The invariant surface-based shape descriptor is presented in Sect. 3. Section 4 introduces 3D video data encoding using the proposed model. Section 5 describes various experiments on challenging datasets. Section 6 concludes the paper with discussions.

## 2 Related work

Several multi-view video capture systems were developed in the recent years [4–7] to provide a new media that gives users free-viewpoint visualization of 3D objects in motion (namely 3D video). Unlike depth maps (2.5D data) which are unclosed surface, 3D video data represent objects in full 3D as a sequence of reconstructed closed surfaces (3D meshes). This technology has potentially several applications in medicine, culture, communication, entertainment, etc. In practice, 3D video data are reconstructed frame-by-frame using multi-view stereo techniques [13].

To encode a sequence of 3D meshes, the state-of-the-art consists mainly of: (1) methods to compress every frame independently [14]; since redundant information between frames is not managed, encoding cannot be optimal. (2) techniques designed for 3D animation sequences [15–18]; as they are dedicated to meshes sharing the same connectivity, they cannot be applied directly to 3D video data (without post-processing with a surface alignment method [8–11]) and for adaptive bitrate streaming purpose.

On the other hand, the literature has provided numerous 3D shape models based on volume, surface, global or local properties (e.g., medial axis [19], skeleton-curve [20], Reeb graphs [21]). Although most of descriptors can capture intrinsic shape property, they are not suited for dynamic representation as their structure is usually too noisy. Similarly, skeleton fitting approaches can capture intrinsic information of shape based on surface or volume [22–24], but are not particularly invariant in time and often need prior knowledge on the shape to be described (e.g., a human skeleton). Moreover, once the structure (e.g., topology) is found, its relationship to surface variations is usually lost [25].

We propose a new surface-based shape descriptor that has invariant property to surface variations, and can be used as a cut graph [16, 26] to encode 3D video data using a transformation into 2D video, by cutting and parameterizing 3D surface meshes on image planes (see Sect. 4 and [12, 27]). To our knowledge, no similar model has been designed [1, 2, 28].

### 3 Invariant surface-based shape representation

#### 3.1 Local feature extraction

Let us assume that dynamic surfaces representing real-world objects in motion can be approximated by compact 2-manifold meshes. We consider geodesic distances to characterize surface intrinsic properties, as geodesic distances are invariant to pose, and robust to shape variations when normalized [21, 9]. Let  $\mu : \mathcal{S} \rightarrow \mathbb{R}$  denote the continuous function defined on the object surface  $\mathcal{S}$ :

$$\mu(v) = \int_{\mathcal{S}} g(v, s) dS, \quad (1)$$

where  $g : \mathcal{S}^2 \rightarrow \mathbb{R}$  is the geodesic distance between two points on  $\mathcal{S}$ . Eq. 1 is the geodesic integral function whose critical points can be used to characterize shape (see Morse theory [29, 30]). For example, local maxima usually correspond to limb extremities of humans or animals while the global minimum corresponds to the body center. We use a Reeb graph to robustly identify and match critical points over time using geometry and topology information [21, 31–33] (see Fig. 1).

#### 3.2 Temporal geodesic consistency

**Definition 1.** Assuming a set of  $N$  points  $\mathcal{B} = \{b_1, \dots, b_N\}$  defined on a 2-manifold  $\mathcal{S}$ , the points  $v_1$  and  $v_2$  on  $\mathcal{S}$  are said geodesically consistent with respect to  $\mathcal{B}$  if and only if:

$$\forall i \in [1, N], \quad |g(v_1, b_i) - g(v_2, b_i)| \leq \epsilon, \quad (2)$$

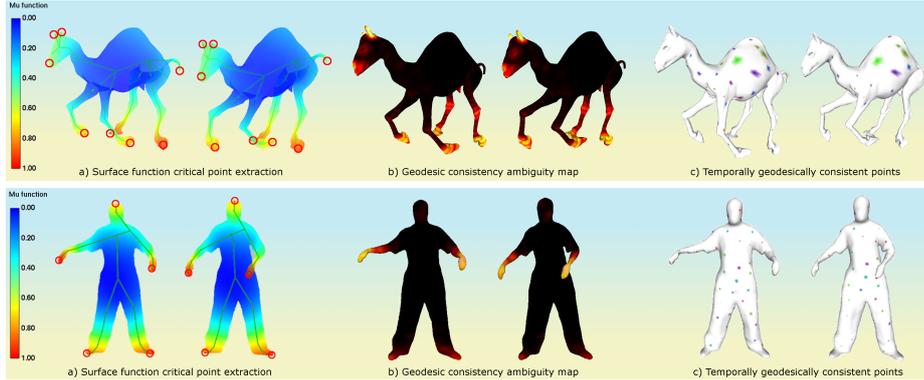
where  $\epsilon \rightarrow 0$ . If the points in  $\mathcal{B}$  do not have any particular configuration of alignment or symmetry, the geodesic consistency property can be used to uniquely locate points on  $\mathcal{S}$  when  $N > 2$ . In practice, the unicity is verified by checking the number of intersections of isovalue lines from  $\mathcal{B}$ , and ambiguities are solved by increasing  $N$  or adding geometric constraints (e.g., Euclidean distance).

**Definition 2.** Assuming a set of  $N$  points  $\mathcal{B}^t = \{b_1^t, \dots, b_N^t\}$  defined on a deformable 2-manifold  $\mathcal{S}^t$  at time  $t \in [t_b, t_e]$ , the points  $v_1^t$  and  $v_2^t$  on  $\mathcal{S}^t$  are said temporally geodesically consistent with respect to  $\mathcal{B}^t$  in  $[t_b, t_e]$  if and only if:

$$\forall t \in [t_b, t_e], \forall i \in [1, N], \quad |g(v_1^t, b_i^t) - g(v_2^{t+\delta}, b_i^{t+\delta})| \leq \epsilon, \quad (3)$$

where  $t_b < t_e$ ,  $t + \delta \in [t_b, t_e]$  and  $\epsilon \rightarrow 0$ .  $g$  is normalized using the maximum geodesic distance over all pairs of points on  $\mathcal{S}^t$  to preserve geodesic consistency

when surfaces undergo non-rigid deformations (e.g., scale changes). Figure 1 illustrates temporal geodesic consistency with respect to critical points (top: 8, bottom: 5) extracted automatically using local geometry and topology properties (see [33]). Ambiguity maps are obtained by counting the number of candidate pairs  $(v_1^t, v_2^{t+\delta})$  when  $\epsilon > 0$ . We observe that the regions located around object centers have very low ambiguity (i.e., numerical approximation is not an issue). In practice, we can search for  $v_2^{t+\delta} = \arg \min_{v \in \mathcal{S}^t} \sum_{i=1}^N |g(v_1^t, b_i^t) - g(v, b_i^{t+\delta})|$ .



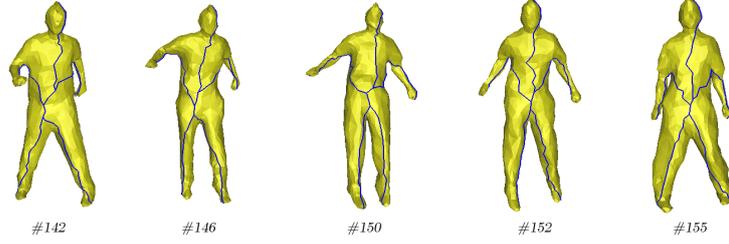
**Fig. 1. Temporal geodesic consistency.** a) Critical points extracted on surface mesh using Reeb graphs [33]. b) Geodesic consistency ambiguity map (darker means less position ambiguity). c) 50 temporally consistent points chosen randomly.

### 3.3 Invariant surface-based graph construction

**Definition 3.** Let  $\mathcal{C}^t = \{c_1^t, \dots, c_N^t\}$  denote a set of invariant local features (e.g., local extrema) on  $\mathcal{S}^t$  that are tracked over time in  $[t_b, t_e]$ . The surface-based shape descriptor  $\mathcal{T}(\mathcal{V}^t, \mathcal{P}^t)$  is a graph on  $\mathcal{S}^t$  whose nodes  $\mathcal{V}^t$  are temporally geodesically consistent with respect to  $\mathcal{C}^t$  in  $[t_b, t_e]$ . Every edge in  $\mathcal{P}^t$  of  $\mathcal{T}$  is linked to a feature in  $\mathcal{C}^t$ , and nodes of  $\mathcal{T}$  represent edge junctions. Here, an edge consists of a path<sup>1</sup> on  $\mathcal{S}^t$ . To maintain the graph structure invariant over time independently from the parameterization of  $\mathcal{S}^t$ , we develop a probabilistic framework where edge positions are optimized using two geodesic consistency cues (see Fig. 2), while being located in regions of low ambiguity (see Def. 2).

**Construction.** First, we define an initial graph structure  $\rho_0$  on  $\mathcal{S}^{t_b}$  at  $t_b$  as either the global minimum (i.e., one point) given by Eq. 1 if  $\mathcal{S}^{t_b}$  is genus-0, or as a graph cutting handles if the genus is higher (see [34], [35] and Sect. 4). Second, we initialize the graph:  $\rho \leftarrow \rho_0$ . The graph  $\mathcal{T}$  is then built by iteratively adding

<sup>1</sup> A path on a surface is a set of points linked two-by-two by a line.



**Fig. 2. Invariant surface-based shape descriptor.** The descriptor is a graph (in blue) defined on the object surfaces. Graph nodes are maintained geodesically consistent over time, whereas edges vary adaptively to surface deformations. (Bouncing sequence.)

the shortest edge linking a local feature (e.g., local maxima) in  $\mathcal{C}^{t_b}$  to the current graph structure  $\rho$  until all elements in  $\mathcal{C}^{t_b}$  are linked. At each step, the path  $\rho_j$  given by the pair of points  $(c_j^{t_b}, v_j^{t_b}) \in \mathcal{C}^{t_b} \times \rho$  verifies:

$$(c_j^{t_b}, v_j^{t_b}) = \arg \min_{(c,v) \in \mathcal{C}^{t_b} \times \rho} g(c, v). \quad (4)$$

$\rho_j$  is linked to  $\rho$ :  $\rho \leftarrow \rho \cup \rho_j$ , and  $v_j^{t_b}$  is inserted into the set  $\mathcal{V}^{t_b}$  (initially empty). When every feature in  $\mathcal{C}^{t_b}$  is linked to  $\rho$ , we obtain  $\rho = (\bigcup \rho_j) \cup \rho_0$  and we finally set:  $\mathcal{T} \leftarrow \rho$  at  $t_b$ .

For all  $t > t_b$ , the invariant model is obtained by building a graph whose nodes have temporal geodesic consistency with the prior graph nodes and are located at local maxima or in non-ambiguous regions.. The problem is formulated as an MRF to find the optimal paths linking the graph nodes using intrinsic surface properties, so that graph constructions across time are independent from surface parameterization. The algorithm to construct a graph at  $t$  is the following:

1. Extract local features  $\mathcal{C}^t = \{c_1^t, \dots, c_N^t\}$  on  $\mathcal{S}^t$  using Eq. 1 and match them to prior ones in  $\mathcal{C}^{t-1}$  (e.g., using geometry and topology information [33]).
2. Derive an initial structure  $\rho_0^t$  on  $\mathcal{S}^t$  geodesically consistent to the prior one. Note that for genus-0 surface,  $\rho_0^t$  is usually a point located around the object center. Set the graph at  $t$ :  $\rho^t \leftarrow \rho_0^t$ .
3. Edges that link the features  $\mathcal{C}^t$  to the current graph structure  $\rho^t$  are added iteratively, and in the same order as the prior steps. Let  $\mathcal{P}^t = \{p_i^t\}$  denote the set of points forming a path (a graph edge) linking a feature  $c^t$  to a node  $v^t$  at  $t$ , and  $\mathcal{D}^t = \{d_i^t\}$  denote the set of points forming the shortest path linking  $c^t$  to  $v^t$  (e.g., using Dijkstra's algorithm). To obtain the optimal path  $\mathcal{P}^t = \{p_i^t\}$ , the problem is expressed as a MAP-MRF where the surface mesh vertices at  $t$  serve as sites. Probabilities of  $p_i^t$  to be at some positions at  $t$  are computed given known priors  $\mathcal{P}^{t-1}$  and  $\mathcal{D}^t$ . The posterior probability to

maximize is:

$$\Pr(\mathcal{P}^t | \mathcal{D}^t, \mathcal{P}^{t-1}) \propto \prod_i E_d(p_i^t, d_i^t) E_p(p_i^t, p_i^{t-1}) \prod_i \prod_{j \in \mathcal{N}(i)} V(p_i^t, p_j^t), \quad (5)$$

where  $E_d$  and  $E_p$  are the local evidence terms for a point  $p_i^t$  to be at positions inferred from  $d_i^t$  and  $p_i^{t-1}$  respectively,  $\mathcal{N}(i)$  is the neighborhood of  $i$ , and  $V$  is a pair-wise smoothness assumption (so that  $\mathcal{P}^t$  forms a path on  $\mathcal{S}^t$ ).  $E_d$  and  $E_p$  are defined as what follows:

$$E_d(p_i^t, d_i^t) = f_d \left( \sum_{k \in [1, N]} \|g(p_i^t, c_k^t) - g(d_i^t, c_k^t)\| \right), \quad (6)$$

$$E_p(p_i^t, p_i^{t-1}) = f_p \left( \sum_{k \in [1, N]} \|g(p_i^t, c_k^t) - g(p_i^{t-1}, c_k^{t-1})\| \right), \quad (7)$$

where  $f_d$  and  $f_p$  are Gaussian distributions centered on  $d_i^t$  and  $p_i^{t-1}$  respectively,  $g$  is the normalized geodesic distance,  $c_k^t \in \mathcal{C}^t$  and  $c_k^{t-1} \in \mathcal{C}^{t-1}$ . Note that indices were simplified for clarity:  $\mathcal{P}^{t-1}$ ,  $\mathcal{P}^t$  and  $\mathcal{D}^t$  may not have the same number of elements, and  $d_i^t$  and  $p_i^{t-1}$  are the closest point to  $p_i^t$  on  $\mathcal{D}^t$  and  $\mathcal{P}^{t-1}$ . Hence, Eq. 5 estimates the probability of  $\mathcal{P}^t$  to be geodesically consistent to the previous edge  $\mathcal{P}^{t-1}$ , while being influenced by the shortest path  $\mathcal{D}^t$ . Let  $\mathcal{P}^*$  denote the optimal path linking the feature  $c^t$  to the node  $v^t$ . Thus, we have to estimate:

$$\mathcal{P}^* = \arg \max_{\{\mathcal{P}^t\}} \Pr(\mathcal{P}^t | \mathcal{D}^t, \mathcal{P}^{t-1}), \quad (8)$$

where  $\{\mathcal{P}^t\}$  denotes all the possible paths linking  $c^t$  to  $v^t$ . Shortest paths are added one-by-one to avoid edge overlapping when linking local features.  $E_d$  acts as a force that attracts the path to a state where the stress is lower (see Fig. 2) when an elastic deformation occurs or in case of surface noise (e.g., 3D reconstruction artifact). As well,  $E_d$  prevents the model to be subject to error accumulation over time, causing drift effects. On the other hand,  $E_p$  maintains the graph structure consistent over time, which can be crucial for some applications (see Sect. 4).

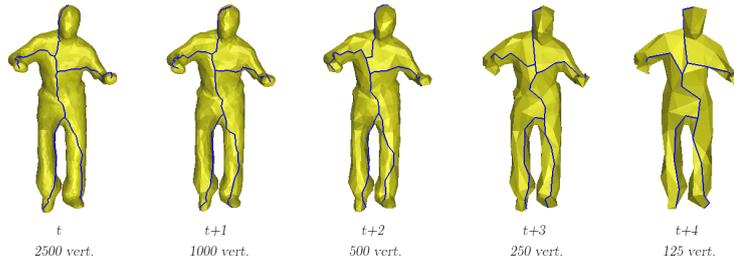
$\rho^t$  is obtained by iteratively adding paths  $\mathcal{P}_j^*$  linking  $c_j^t \in \mathcal{C}^t$  to the current graph at node  $v_j^t$ .  $v_j^t$  is the closest point on the current graph to:

$$\bar{v}_j^t = \arg \min_{v \in \rho^t} [\lambda \cdot g(\hat{v}_j^t, v) + (1 - \lambda) \cdot g(\check{v}_j^t, v)], \quad (9)$$

where  $\hat{v}_j^t$  is the point in  $\mathcal{S}^t$  geodesically consistent to  $v_j^{t-1}$  in  $\mathcal{S}^{t-1}$  with respect to  $\mathcal{C}^t$ ,  $\check{v}_j^t$  is the intersection point given by the shortest path from  $c_j^t$  to  $\rho^t$ , and  $\lambda = 0.5$  is a weight. (Temporal priors are discarded if  $\lambda = 0$ .) In addition,  $v_j^t$  is constrained to belong to the edge derived from the edge containing  $v_j^{t-1}$ . The structure of  $\mathcal{T}$  is therefore maintained invariant over time. Note that priors can be extended to  $\{\mathcal{P}^{t-k}\}_{t_b < k < t}$ .

4. Repeat Step 3. until every feature in  $\mathcal{C}^t$  is linked to  $\rho^t$ . Finally, the graph  $\mathcal{T}$  at  $t$  is given by  $\rho^t \leftarrow (\bigcup_j \mathcal{P}_j^{*,t}) \cup \rho_0^t$ .
5. Set  $t \leftarrow t + 1$  and repeat Step 1. to 4. for all  $t < t_e$ .

Note that the optimization problem in Step 3. can be effectively solved by dynamic programming. Finally, as illustrated in Fig. 3 we obtain a graph that is invariant over time regardless of the surface parameterization (i.e., mesh complexity and connectivity).



**Fig. 3. Invariant property against surface parameterization.** The graph structure is maintained invariant even though the surface mesh complexity and connectivity change. Here, the number of vertices varies from 2500 to 125 vertices. (Lock sequence.)

## 4 3D video data encoding

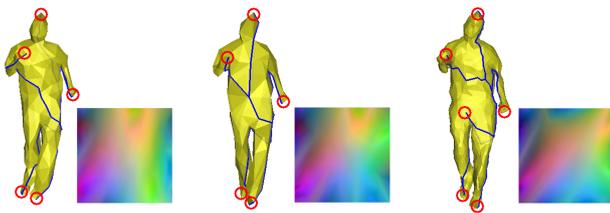
The surface-based shape descriptor  $\mathcal{T}$  introduced in the previous section can provide an invariant structure to 3D data obtained independently, such as a sequence of 3D meshes obtained from multiple view stereo [4, 13, 6]. Hence, we propose to apply the descriptor for 3D video data encoding using a strategy inspired by the geometry images [12].

For each frame, the graph structure of  $\mathcal{T}$  is used as a cut graph  $\rho$  that cuts and opens the 3D surface mesh  $M$  into a disk (a genus-0 chart).  $M$  is then mapped onto a flat parameter domain, which will be used as an image plane. Finally,  $M$  is resampled on a regular grid, where the 3D coordinates XYZ are scaled and stored as RGB pixel components to form a 2D (geometry) image  $\mathcal{I}$ . To retrieve  $M$  from  $\mathcal{I}$ , RGB values are simply reconverted to 3D coordinates. When applying the transformation on a sequence of 3D meshes, the process returns a sequence of images (i.e., a video). As the graph structure is invariant over the sequence, consecutive frames vary smoothly and can therefore be efficiently encoded using any popular codec for 2D video. Note that if a lossy compression method is used for encoding and alters the border of  $\mathcal{I}$ , cracks may be observed on the reconstructed surface around the cut  $\rho$ . In that case, a post-processing step (e.g., mesh joining or hole filling) may be necessary to preserve the topology

of the initial mesh. The advantage of the proposed invariant surface-based shape descriptor for 3D video encoding is at least twofold:

1. The shape descriptor can be used as cut graphs to produce smoothly varying geometry images from real-world 3D video data independently from the surface parameterization, i.e., even though the mesh resolution or connectivity is inconsistent between consecutive frames. Hence the model allows for adaptive bitrate streaming application, whereas state-of-the-art methods cannot be applied [16, 27].
2. In standard parameterization approaches [36, 12, 27], the computation of the cut graph  $\rho$  is obtained iteratively and requires several parameterization steps to detect all the local extrema one-by-one (e.g., using triangle geometric stretch). On the other hand, the proposed strategy is one-shot, and still guarantees that the generated cut path passes through all local extrema of  $M$  (i.e., surface protrusions), which is a crucial condition to preserve the geometry accuracy after transformation. When the cut graphs are well defined, the transformation can be used for lossless compression of 3D meshes.

**Topology change.** As the cut graph passes through all extrema, critical points usually lie at the boundaries of the parameter domain. When surface topology changes, the number of critical points may vary, and the graph structure can locally change. This results in a discontinuity between consecutive geometry images that cannot be avoided. On the other hand, it guarantees that the original surface topology is preserved and can be reconstructed from a single chart. Otherwise a surface alignment method could be applied as preprocessing [9–11], but large resolution variations as shown in Fig. 3 would not be handled and original topology would be lost. Methods that estimate global geodesic distortions for shape matching are usually robust to local surface deformation [37, 9]. However, the measures can be strongly affected by surface topology changes, as opposed to the proposed descriptor which is only locally affected. Figure 4 shows geometry image discontinuities when altering geodesic consistency of nodes and adding an arbitrary critical point. As critical points are matched across time, image regions with no perturbation remain aligned (see left part of images).



**Fig. 4.** (Left and Center) Geometry images show discontinuities when graph nodes are not geodesically consistent. (Right) Adding an arbitrary critical point alters locally the image boundaries. (Bouncing sequence.)

## 5 Experimental results

**Datasets.** For experimental validations, we have tested the algorithm on publicly available datasets of 3D video reconstructed from multi-view images [6, 10]. They consist of real subjects wearing loose clothing and performing various actions, such as dancing or jumping. Surfaces can therefore vary a lot between two consecutive frames when the motion is fast. Our experiments aim to assess the invariant property of the proposed descriptor regardless of surface parameterization, and its performance (e.g., reconstruction accuracy) when applied for 3D video adaptive bitrate streaming. We use 3D mesh sequences processed by [11] as the surface genus is theoretically consistent over the sequences. (In practice, 3D video data can be post-processed using a surface alignment method to prevent surface topology changes [8–11].) In addition, we remeshed the sequences to cancel all mesh connectivity consistency, and produced mixed resolution 3D video data containing alternatively 3D meshes of 1000, 500, 250 and 125 vertices.

We perform comparisons to a state-of-the-art parameterization technique [27], where cut graphs are obtained by iterative parameterizations. The approach, here named Geometry Image Sequence (GIS), is known to optimally encode closed 3D surface meshes. Results obtained with our proposed technique on mesh sequences having same resolution are denoted ‘fixed’, whereas results obtained on sequences with meshes having various resolutions are denoted ‘mixed’. Computation time for a 1000 vertex mesh is in the order of seconds on dual-core PC.

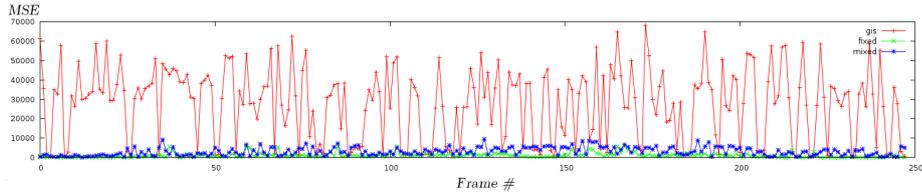
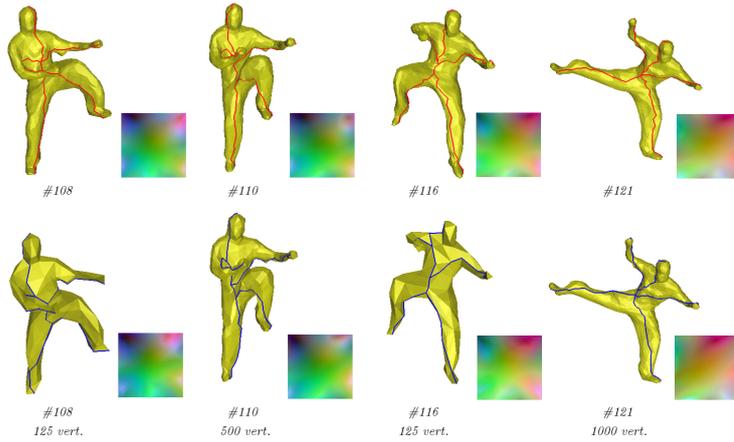
**Invariant property evaluation.** To assess the invariant property of the descriptor to surface variations and its ability to produce consistent geometry images that varies smoothly, the mean square error of pixel values (MSE) between consecutive geometry images is computed (smaller MSE is better). It allows us to estimate how much the geometry images vary over a sequence. In our experiments, the size of geometry images has been fixed to  $128 \times 128$  pixels (encoded in RGB with 8bit per pixel component) for the sake of consistent comparison. (To achieve optimal streaming, the geometry images should indeed be resized with respect to the mesh resolution.) Table. 1 shows average MSE obtained on various sequences. The proposed descriptor shows remarkable invariant property between consecutive frames: average MSE(fixed) values are very low. Moreover, the resolution changes do not affect the performance: MSE(mixed) are low as well. Note that as GIS does not contain any stabilization mechanism: average MSE(GIS) values are high and are given for comparison.

Fig. 5 shows MSE for the Lock sequence. The other sequences return similar results. Figure 6 illustrates invariant graphs obtained with our approach with fixed and mixed mesh resolution.

**Reconstruction accuracy.** To assess the reconstruction accuracy of geometry images obtained from the invariant surface-based shape descriptor used as cut graphs, Hausdorff distances are computed between original meshes and reconstructed meshes [38]. Average Hausdorff distances  $\Delta$  between ground truth

**Table 1.** Average MSE of pixel values between consecutive geometry images.

	MSE(GIS)	MSE(fixed)	MSE(mixed)
Bouncing	35302	2886	3224
Crane	28485	1670	2025
Handstand	30671	1261	3125
Kickup	27700	1938	3753
Lock	22466	1700	3037
Samba	35302	2886	3282

**Fig. 5. MSE of Lock sequence.** Our approach produces geometry images that vary very smoothly regardless of surface mesh complexity and connectivity.**Fig. 6. Graph invariant property regardless of surface mesh complexity and connectivity.** (top) shows a mesh sequence with 1000 vertices. (bottom) shows the same sequence with meshes at different resolutions. Although surface parameterizations are different, the proposed surface-based shape descriptor computed on the Lock sequence shows invariant property and adaptivity: graphs and geometry images remain similar.

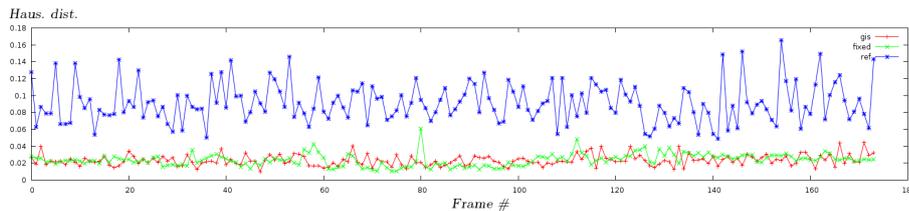
sequences and reconstructed surfaces by GIS and our proposed method (with fixed resolution) are reported in Table. 2. We can observe similar performances

between the proposed approach and GIS as  $\Delta$  is very low for both methods. Results between original data and simplified meshes (125 vertices) are given for comparison (see  $\Delta(\text{ref})$ ).

**Table 2.** Average Hausdorff distances  $\Delta$  to ground truth.

	$\Delta(\text{GIS})$	$\Delta(\text{proposed})$	$\Delta(\text{ref})$
Bouncing	0.0122	0.0126	0.0885
Crane	0.0126	0.0132	0.0729
Handstand	0.0119	0.0122	0.0920
Kickup	0.0118	0.0120	0.0862
Lock	0.0079	0.088	0.0783
Samba	0.0223	0.0237	0.0913

As shown in Figure 7, our method can achieve accurate reconstruction (comparable to GIS) while using an original one-shot processing, as opposed to standard iterative parameterizations employed by GIS. Additional examples are given in Fig. 8 showing the descriptor invariant property to surface undergoing large deformations.

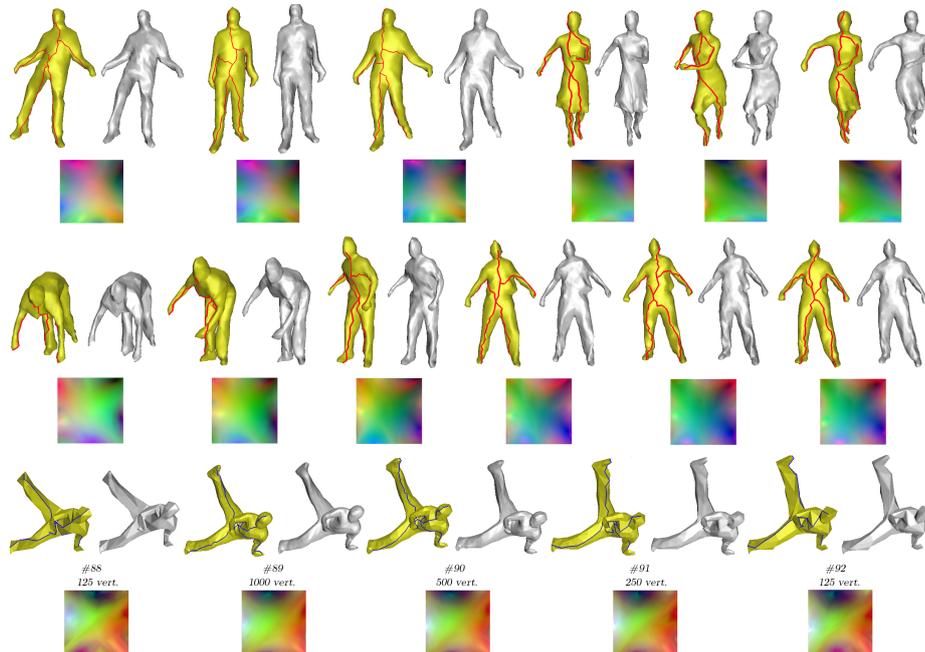


**Fig. 7.** Hausdorff distances  $\Delta$  for Samba sequence. Our approach allows accurate reconstruction of mesh sequences comparable to state-of-the-art implementation [27]).

**Encoding performance.** Table. 3 shows 3D video encoding performance with respect to different strategies. Our method clearly performs better.

## 6 Conclusion

We present a novel invariant shape descriptor to represent spatio-temporal visual information that varies over time, such as 3D dynamic surfaces. The proposed descriptor consists in a surface-based graph that lies on object surfaces, and is anchored to local features. The overall graph structure is made invariant to surface variations using surface intrinsic geometric properties while surfaces undergo



**Fig. 8. Encoding and reconstruction.** Surface-based graphs are shown in red and blue (for mixed resolutions), reconstructions are shown in grey. Although surfaces undergo strong variations, the invariant surface-based shape descriptor produces smoothly varying geometry images and accurate surface reconstruction. Sequences are: (Top) Crane and Samba, (Middle) Handstand, (Bottom) Kickup.

**Table 3.** 3D video encoding. For each format, the size of each sequence is given in KB. Standard H.264/MPEG-4 is used for compression of geometry images ( $128 \times 128$ p).

	#fr.	OFF(zip)	GIS	Proposed
Bouncing	174	16,300	304.4	169.9
Crane	173	14,100	283.7	162.7
Handstand	173	24,700	283.4	154.7
Kickup	219	29,900	365.1	197.0
Lock	249	32,400	388.2	204.0
Samba	174	22,200	304.0	173.2

non-rigid deformation. In particular, the graph is defined within a probabilistic framework using temporal geodesic consistency cues as priors, and is independent to surface parameterization. Hence, the descriptor can be used to bring an invariant structure to 3D geometric data that are produced independently, such as 3D video obtained from multiple view stereo.

We show that the proposed shape descriptor can be employed as surface cut graphs, which enables 3D surface models to be transformed into 2D (geometry) images using a one-shot strategy while geometry is accurately preserved. Moreover, the invariant property of the representation allows the production of smoothly varying images, regardless of the 3D surface mesh complexity and connectivity. Therefore, the approach is suitable for adaptive bitrate streaming of 3D video data, which was a challenging issue as state-of-the-art techniques are only designed to optimally encode 3D animated mesh sequences sharing a same mesh connectivity. For further research, additional surface features such as color (when available) may be exploited.

**Acknowledgments.** This work was supported in part by the JST-CREST project “Creation of Human-Harmonized Information Technology for Convivial Society”, and the Japan Society for the Promotion of Science (Kakenhi Wakate-B No. 23700170). The authors thank Dr. Lyndon Hill for his preliminary work on this project.

## References

1. Forsyth, D.A., Mundy, J.L., Zisserman, A., Coelho, C., Heller, A., Rothwell, C.: Invariant descriptors for 3d object recognition and pose. *IEEE PAMI* **13** (1991)
2. Mundy, J., Zisserman, A.: *Geometric invariance in computer vision*. MIT Press (1992)
3. Rothganger, F., Lazebnik, S., Schmid, C., Ponce, J.: 3d object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints. *IJCV* **66** (2006) 231–259
4. Matsuyama, T., Wu, X., Takai, T., Nobuhara, S.: Real-time 3d shape reconstruction, dynamic 3d mesh deformation, and high fidelity visualization for 3d video. *CVIU* **96** (2004) 393–434
5. Allard, J., M enier, C., Raffin, B., Boyer, E., Faure, F.: *Grimage: Markerless 3d interactions*. *ACM SIGGRAPH - Emerging Technologies* (2007)
6. Starck, J., Hilton, A.: *Surface capture for performance-based animation*. *IEEE CGA* (2007)
7. de Aguiar, E., Stoll, C., Theobalt, C., Ahmed, N., Seidel, H.P., Thrun, S.: *Performance capture from sparse multi-view video*. *ACM Trans. Graphics* **27** (2008)
8. Starck, J., Hilton, A.: *Spherical matching for temporal correspondence of non-rigid surfaces*. *IEEE ICCV* (2005)
9. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: *Calculus of non-rigid surfaces for geometry and texture manipulation*. *IEEE Trans. VCG* (2007) 902–913
10. Vlastic, D., Baran, I., Matusik, W., Popovic, J.: *Articulated mesh animation from multi-view silhouettes*. *ACM Trans. Graphics* **27** (2008)
11. Cagniard, C., Boyer, E., Ilic, S.: *Probabilistic deformable surface tracking from multiple videos*. *ECCV* (2010)
12. Gu, X., Gortler, S., Hoppe, H.: *Geometry images*. *ACM SIGGRAPH* (2002)
13. Seitz, S., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: *A comparison and evaluation of multi-view stereo reconstruction algorithms*. *IEEE CVPR* (2006)
14. Alliez, P., Gotsman, C.: *Recent advances in compression of 3d meshes*. *Advances in Multiresolution for Geometric Modelling*. Springer-Verlag (2005)

15. Alexa, M., Müllen, W.: Representing animations by principal components. *Computer Graphics Forum* **19** (2000)
16. Briceno, H., Sandler, P., McMillian, L., Gortler, S., Hoppe, H.: Geometry videos: A new representation for 3d animations. *Eurographics/SIGGRAPH Symp. Computer Animation* (2003) 136–146
17. Karni, Z., Gotsman, C.: Compression of soft-body animation sequence. *Computers & Graphics* **28** (2004) 25–34
18. Mamou, K., Zaharia, T., Preteux, F., Stefanoski, N., Ostermann, J.: Frame-based compression of animated meshes in mpeg-4. *IEEE ICME* (2008)
19. Blum, H.: A transformation for extracting new descriptors of shape. *Models for the perception of speech and visual form* (1967)
20. N.Cornea, D.Silver, X.Yuan, R.Balasubramanian: Computing hierarchical curve skeletons of 3d objects. *The Visual Computer* **21** (2005) 945–955
21. Hilaga, M., Shinagawa, Y., Kohmura, T., Kunii, T.L.: Topology matching for fully automatic similarity estimation of 3d shapes. *ACM SIGGRAPH* (2001) 203–212
22. J.Carranza, C.Theobalt, M.Magnor, H.-P.Seidel: Free-viewpoint video of human actors. *ACM Trans. Graphics* **22** (2003) 569–577
23. K.Palagyi, A.Kuba: A parallel 3d 12-subiteration thinning algorithm. *Graph. Models and Image Proc.* **61** (1999) 199–221
24. Baran, I., Popovic, J.: Automatic rigging and animation of 3d characters. *ACM Trans. Graphics* **26** (2007) 27
25. T.Tung, T.Matsuyama: Topology dictionary for 3d video understanding. *IEEE Trans. PAMI* **34** (2012) 1645–1657
26. Habe, H., Katsura, Y., Matsuyama, T.: Skin-off: Representation and compression scheme for 3d video. *Picture Coding Symposium* (2004)
27. Saboret, L., Alliez, P., Lévy, B.: Planar parameterization of triangulated surface meshes. In *CGAL Reference Manual*. CGAL Editorial Board, 4.0 edition (2012)
28. Mortara, M., Patanè, G.: Affine-invariant skeleton of 3d shapes. *Shape Modeling International* (2002)
29. Morse, M.: *The calculus of variations in the large*. American Mathematical Society, Colloquium Publication 18, New York (1934)
30. Pascucci, V., Scorzelli, G., Bremer, P.T., Mascarenhas, A.: Robust on-line computation of reeb graphs: Simplicity and speed. *ACM Trans. Graphics* **26** (2007)
31. Edelsbrunner, H., Harer, J., Mascarenhas, A., Pascucci, V.: Time-varying reeb graphs for continuous space-time data. *Symp. Computational Geometry* (2004)
32. Klein, T., Ertl, T.: Scale-space tracking of critical points in 3d vector fields. *Proc. Topology-Based Methods in Visualization* (2005)
33. Tung, T., Schmitt, F.: The augmented multiresolution reeb graph approach for content-based retrieval of 3d shapes (code on webpage). *Int’l J. Shape Modeling* **11** (2005) 91–120
34. Taubin, G., Rossignac, J.: Geometric compression through topological surgery. *ACM Trans. Graphics* **17** (1998) 84–115
35. Erickson, J., Har-Peled, S.: Optimally cutting a surface into a disk. *Discrete & Computational Geometry* **31** (2004) 37–59
36. Floater, M.: Parametrization and smooth approximation of surface triangulations. *CADG* **14** (1997) 231–250
37. Mémoli, F., Sapiro, G.: A theoretical and computational framework for isometry invariant recognition of point cloud data. *Found. Comput. Math.* **5** (2005) 313–347
38. P. Cignoni, C.R., Scopigno, R.: Metro: measuring error on simplified surfaces. *Computer Graphics Forum* **17** (1998) 167–174