

12-2011

Modeling 3D articulated motions with conformal geometry videos (CGVs)

Dao T. P. QUYNH

Nanyang Technological University

Ying HE

Nanyang Technological University

Xiaoming CHEN

Nanyang Technological University

Jiazhi XIA

Nanyang Technological University


Qian SUN

Nanyang Technological University

See next page for additional authors

DOI: <https://doi.org/10.1145/2072298.2072349>

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

 Part of the [Databases and Information Systems Commons](#), and the [Numerical Analysis and Computation Commons](#)

Citation

QUYNH, Dao T. P.; HE, Ying; CHEN, Xiaoming; XIA, Jiazhi; SUN, Qian; and HOI, Steven C. H.. Modeling 3D articulated motions with conformal geometry videos (CGVs). (2011). *MM '11: Proceedings of the 2011 ACM Multimedia Conference: November 28 - December 1, Scottsdale, AZ*. 383-392. Research Collection School Of Information Systems.

Available at: https://ink.library.smu.edu.sg/sis_research/4178

This Conference Proceeding Article is brought to you for free and open access by the School of Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Author

Dao T. P. QUYNH, Ying HE, Xiaoming CHEN, Jiazhi XIA, Qian SUN, and Steven C. H. HOI

Modeling 3D Articulated Motions with Conformal Geometry Videos (CGVs) *

Dao T. P. Quynh Ying He Xiaoming Chen Jiazhi Xia Qian Sun Steven C. H. Hoi
School of Computer Engineering
Nanyang Technological University
Singapore
daot0006|yhe|xmchen|xiaj0002|sunq0004|chhoi@ntu.edu.sg

ABSTRACT

3D articulated motions are widely used in entertainment, sports, military, and medical applications. Among various techniques for modeling 3D motions, geometry videos (GVs) are a compact representation in that each frame is parameterized to a 2D domain, which captures the 3D geometry (x, y, z) to a pixel (r, g, b) in the image domain. As a result, the widely studied image/video processing techniques can be directly borrowed for 3D motion. This paper presents conformal geometry videos (CGVs), a novel extension of the traditional geometry videos by taking into the consideration of the isometric nature of 3D articulated motions. We prove that the 3D articulated motion can be uniquely (up to rigid motion) represented by (λ, H) , where λ is the conformal factor characterizing the intrinsic property of the 3D motion, and H the mean curvature characterizing the extrinsic feature (i.e., embedding or appearance). Furthermore, the conformal factor λ is pose-invariant. Thus, in sharp contrast to the GVs which capture 3D motion by three channels, CGVs take only one channel of mean curvature H and the first frame of the conformal factor λ , i.e., approximately 1/3 the storage of the GVs. In addition, CGVs have strong spatial and temporal coherence, which favors various well studied video compression techniques. Thus, CGVs can be highly compressed by using the state-of-the-art video compression techniques, such as H.264/AVC. Our experimental results on real-world 3D motions show that CGVs are a highly compact representation for 3D articulated motions, i.e., given CGVs and GVs of the same file size, CGVs show much better visual quality than GVs.

Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems]: Video

General Terms

Algorithms, Design

*Area chair: Wei Tsang Ooi

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.
Copyright 2011 ACM 978-1-4503-0616-4/11/11 ...\$10.00.

Keywords

Geometry videos, conformal geometry videos, 3D articulated motion, deformable objects, video compression, H.264/AVC, conformal parameterization, isometric transformation.

1. INTRODUCTION

The media and entertainment industry has experienced explosive growth in the last decade. Computer animated characters are now necessary components of many applications, including video games, movies, social virtual world, human computer interface designs, mass communication and psychological studies. In order to make these animated characters convincing, they require sophisticated facial expressions and body motions. Thus, 3D motion/video is an emerging research topic. Recent advances in 3D scanning technology, such as multi-view photometric stereo [25], template based space-time registration [15], fringe projection [38], etc. allow the acquisition of 3D motions in real-time. However, the captured 3D motions are usually represented by bulky, irregular polygonal meshes, which poses substantial challenges for modeling and compression.

Geometry image [3] is an emerging technique that intelligently encodes the 3D geometry into an image format, in which each pixel r, g, b represents a 3D vertex x, y, z . Geometry images naturally bridge two research fields, image processing and geometric processing, thus, one can borrow the widely studied image processing techniques (e.g., compression [9] [19]) to 3D geometry. Geometry videos (GVs) extend geometry images to capture 3D motion in a video format. It has been shown that geometry video is effective in modeling 3D facial expressions [34] and 3D animation [2].

With a widespread belief within the media community that a 3D surface has three functional freedoms (i.e., x, y , and z of each vertex), the existing research of geometry compression relies on the ambient space \mathbb{R}^3 in which the surface is embedded and focuses on the extrinsic properties. However, the classical result in differential geometry shows that a surface can be uniquely determined up to rigid motion by its first fundamental form and mean curvature H . With conformal parameterization, the first fundamental form is completely encoded into conformal factor λ , which is the scaling factor of infinitesimal patches of surface. In other words, a 3D surface has only two degrees of freedom, one characterizing its intrinsic feature (i.e., metric) and the other for the extrinsic feature (i.e., embedding), see Fig 1. Inspired by this, this paper presents *conformal geometry videos (CGVs)*, a novel extension of the traditional GVs for modeling general 3D motions of articulated models, which are approximate isometric transformation. By parameterizing 3D motions using conformal parameterizations, we partition the 3D motion into two video sequences λ and H , where the conformal factor λ encodes the intrinsic feature, the mean cur-

vature H encodes the extrinsic feature. Observing that 3D articulated motions are approximate isometric transformations, conformal factor λ , the intrinsic feature, remains unchanged, i.e., it is invariant to 3D poses. Thus, different frames distinguish only by their extrinsic feature H . As a result, we only need to handle the pose-dependent mean curvature. This leads to a highly compact representation that takes only approximately 1/3 the space of conventional geometry videos. In addition, CGVs have strong spatial and temporal coherence, and can be heavily compressed further by using the state-of-the-art video compression techniques, such as H.264/AVC. Our experimental results on real-world 3D motions demonstrate that CGVs outperforms the conventional GVs by providing a more compact and effective modeling framework for 3D articulated motions.

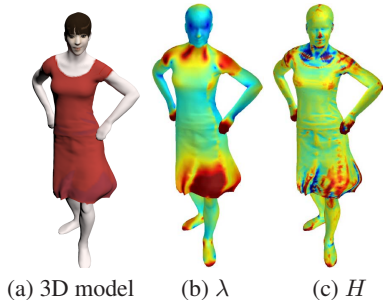


Figure 1: With conformal parameterization, the 3D geometry can be determined (uniquely up to rigid motion) by its conformal factor λ and mean curvature H . λ and H are rendered in color for illustration purpose.

So as presented above, CGV naturally bridges 3D motion modeling and 2D video processing, allowing us to borrow the well-studied video processing technique (e.g., H.264) to model and compress 3D motion data. The specific contributions of this paper include:

- We present CGV, which naturally extend the conventional GV framework by taking into the consideration of the approximate isometric nature of articulated 3D motion.
- We present a set of algorithms to construct CGVs, including geometric feature detection and tracking, constrained polycube conformal parameterization, polycube flattening, and optimization-based 3D motion reconstruction.
- We show that CGVs have stronger spatial and temporal coherence than that of GVs and natural videos, and such redundancy can be significantly removed by the powerful compression tools available in H.264/AVC.

2. PREVIOUS WORK

This section reviews the related work in three different research fields, 3D motion processing, video compression, and global conformal parameterization, which are naturally integrated by the proposed conformal geometry video framework.

2.1 3D motion processing

Recent advances in 3D scanning technology [25] [13] [15] [38] allow us to capture the 3D human motion in real time, which also pose significant challenges in data processing. Each frame of the scanned motion is given in its own scanner space other than in the object space. Therefore, from the analysis and processing point of view, it's highly desirable to align or register the captured data in the object space. Feature tracking is the key technique to find the

correspondence among different frames of 3D motion. Wang *et al.* presented a data-driven approach for video speed precise facial tracking and expression retargeting [31]. To handle surface matching with noise, occlusion and clutter, Wang *et al.* adopted least square conformal parameterization simplifying the 3D human face registration problem to a 2D image matching problem [29]. Mitra *et al.* proposed an algorithm to register large sets of unstructured point clouds of moving and deforming objects without computing correspondences [17]. Wang *et al.* developed an efficient non-rigid 3D motion tracking algorithm to establish inter-frame correspondences that facilitate the temporal study of subtle motions in facial expressions [30].

Since the scanned 3D motion is usually bulky, compression becomes an emerging problem in 3D motion data transmission and storage. Han *et al.* proposed time-varying meshes (TVM) by extending the block matching algorithm from 2D video to 3D meshes [6]. By considering both spatial and temporal redundancies, the compression ratio of TVM is between 6 : 1 and 50 : 1 [7]. Yamasaki *et al.* proposed algorithms of intra-frame and inter-frame coding for both geometry and color texture of TVM [36].

Another approach to compress 3D motions is GVs. Briceño *et al.* parameterized the synthetic animated mesh sequence onto a rectangle domain and then formed GVs [2]. They classified the frames as I-Frames and P-Frames and then used 2D wavelet-based techniques to compress the geometry video. Taking the advantage of approximate isometric feature of facial expressions, Xia *et al.* [34] developed an expression-invariant 3D face parameterization algorithm that can guarantee the exact feature correspondence, which favors the geometry video compression by using the state-of-the-art H.264/AVC.

2.2 Video compression

Video compression has been extensively studied in the past several decades. The traditional 2D video compression techniques can be categorized as prediction, transformation, quantization and entropy coding, see [21] for a comprehensive survey. H.264/AVC is the state-of-the-art video coding standard [10], which is able to achieve very low bitrates by employing many advanced compression techniques such as block-size adaptive intra-frame prediction, quarter-pixel-precision multi-frame motion estimation, block-size adaptive transformation and quantization, high-performance CABAC entropy coding [32]. CGVs are fundamentally different than natural videos and traditional GVs in that CGVs show strong coherence in both temporal and spatial domains, which is particularly suitable for H.264/AVC compression (see Section 5.2 for more details).

2.3 Global conformal parameterization

The research area of global surface parameterization has a long and fruitful tradition in computer graphics community [20]. Conformal parameterization has drawn particular attention due to its many promising properties, such as angle preserving, intrinsic to the geometry, insensitive to resolution/tessellation, etc.

Gu and Yau [5] pioneered a global conformal parameterization algorithm for surfaces with arbitrary genus. They analyzed the structure of the space of all global conformal parameterizations of a given surface and found all possible solutions by constructing a basis of the underlying linear solution space. The resulting parameter lines minimize angle distortion but may have a rather large metric distortion. Furthermore, the space of global conformal parameterizations is too rigid to allow a local alignment of the parameter lines at given surface features. Many follow-up works have been proposed to improve the conformal parameterization quality by reduc-

ing the area distortion and allowing user control on the singularities and/or parameter lines, such as [11] [14] [37], just name a few.

Polycube map is a variant of the global parameterization in that the parametric domain (i.e., a polycube) can be embedded in \mathbb{R}^3 rather than the abstract charts in \mathbb{R}^2 . Polycube parameterization has several unique advantages over the chart based parameterization, which are particularly useful for conformal geometry video constructions: 1) The parametric domain mimics the geometry of the 3D shape, thus, the area distortion can be significantly reduced. 2) The user can easily control the location of singularities and direction of parameter lines. 3) Polycube can be flattened to a 2D rectangular domain, such that the boundary is axis aligned. Thus, one can sample the input 3D model into a completely regular geometry image.

There are several approaches to parameterize 3D models to polycube domain. Tarini *et al.* [23] pioneered the concept of polycube mapping. However, as direct projection from 3D model to polycube is used, their method is not bijective, since two vertices may share the same projected image on the polycube. Wang *et al.* [27] introduced an intrinsic approach that is guaranteed to be a diffeomorphism. However, in this scheme it is difficult to control the polycube map, i.e., a feature on the 3D model may not be mapped to a desired location on the polycube. In their follow-up work, Wang *et al.* [28] proposed the user-controllable polycube map where the users can explicitly specify the pre-images of the polycube corners. Lin *et al.* [16] proposed an automatic approach to construct the polycube map of non-trivial topology. He *et al.* [8] proposed a divide-and-conquer approach to compute polycube map of large-scale models. Xia *et al.* [33] presented an editable polycube map framework that allows the users to easily specify the mapping constraints and post-edit the parameterization. Wan *et al.* [26] presented an optimization framework to find the optimal polycube mapping by trading off the area distortion and the complexity of the polycube.

Although there are extensive research in surface parameterization, little progress has been reported on motion data parameterization, in which the key challenge is to keep the temporal and spatial coherence. Xia *et al.* proposed an algorithm to expression-invariant parameterization algorithm for 3D human faces [34]. Their method computes the harmonic field on a topological annulus and then builds the bijective parameterization by tracing the integral curves of the harmonic field. This method works well for 3D faces which have simple topology and geometry. However, it is difficult to extend their method for parameterization of general 3D articulated motions, which may have much more complicated geometry and topology than human faces.

3. CONFORMAL GEOMETRY VIDEOS

3.1 Conformal representation of 3D surfaces

We use the following notations throughout this paper:

λ	Conformal factor
H	Mean curvature
K	Gaussian curvature
\mathbf{r}	an oriented 2-manifold surface
Δ	Laplace-Beltrami operator
M	Motion sequence, M_i is the i th frame
$k_t^{M_i}(x, x)$	Heat Kernel Signature of point $x \in M_i$

A classical question in differential theory is which data are sufficient to describe a surface in space up to rigid motions.

Let $\mathbf{r}(u, v)$ be the surface embedded in \mathbb{R}^3 and (\cdot, \cdot) be the dot

product. Let $E = (\mathbf{r}_u, \mathbf{r}_u)$, $F = (\mathbf{r}_u, \mathbf{r}_v)$, and $G = (\mathbf{r}_v, \mathbf{r}_v)$ be the coefficients of the first fundamental form, where \mathbf{r}_u and \mathbf{r}_v are the tangent vectors.

$$I = (d\mathbf{r}, d\mathbf{r}) = Edu^2 + 2Fdudv + Gdv^2$$

Let $L = (\mathbf{r}_{uu}, \mathbf{n})$, $M = (\mathbf{r}_{uv}, \mathbf{n})$, and $N = (\mathbf{r}_{vv}, \mathbf{n})$ be the coefficients of second fundamental form, where \mathbf{n} is the normal.

$$II = (d\mathbf{r}, d\mathbf{n}) = Ldu^2 + 2Mdudv + Ndv^2$$

According to the fundamental theorem of surfaces, if E, F, G, L, M, N are given functions of u and v , sufficiently differentiable, which satisfy the Gauss-Codazzi equations and the added conditions that $EG - F^2 > 0$, $E > 0$, $G > 0$, there exists a surface, uniquely determined except for its position in space, which has respectively as its first and second fundamental forms the quadratic forms $Edu^2 + 2Fdudv + Gdv^2$ and $Ldu^2 + 2Mdudv + Ndv^2$. This theorem guarantees not only the existence of a surface, provided the conditions are met, but also that if two surfaces have the same fundamental forms, they are congruent.

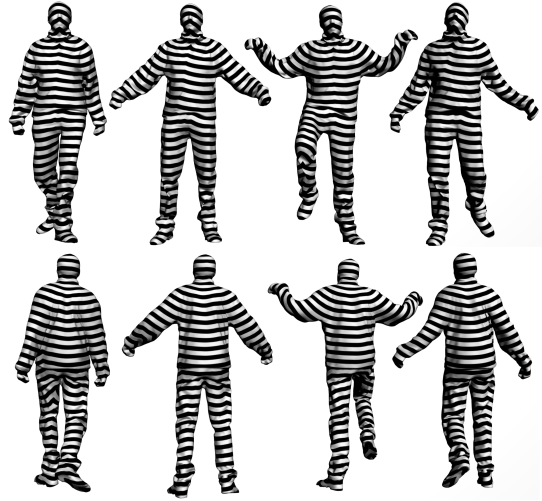


Figure 3: 3D human motions are approximate isometric (i.e., metric preserving). Given various poses of a subject, we compute the geodesic distance with the source point at his head and render the isolines of the resultant geodesic distance fields. Clearly, the geodesic distances (or the metrics) are intrinsic features, that are highly consistent and invariant to the poses.

Note that under conformal parameterization, the u - and v -parameter lines are orthogonal, so $F = (\mathbf{r}_u, \mathbf{r}_v) = 0$. Considering Gauss-Codazzi equations contain three equations, there are only two degrees of freedoms among the six coefficients, E, F, G, L, M and N . In general, a 3D surface can be uniquely (up to rigid motion) determined by the first fundamental form and mean curvature, which is stated in the following theorem [4].

Theorem 1. A closed surface $\mathbf{r} \in \mathbb{R}^3$ with conformal parametrization is determined by its conformal factor λ and mean curvature H uniquely up to rigid motion, where

$$\begin{aligned} \Delta \ln \lambda &= \lambda^2 K \\ H\mathbf{n} &= \Delta \mathbf{r}, \end{aligned}$$

Δ is the Laplace-Beltrami operator and K the Gaussian curvature.

Intuitively speaking, conformal factor can be thought of as scaling infinitesimal patches of the surface, while mean curvature characterizes how the shape is embedded in \mathbb{R}^3 . Theorem 1 states that

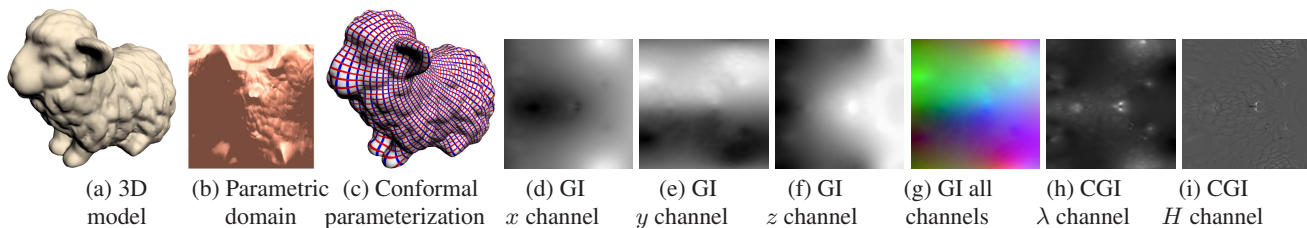


Figure 2: Encoding 3D geometry into conformal factor λ and mean curvature H . The 3D model (a) is conformally parameterized to a rectangular domain (b). The texture mapping in (c) illustrates the angle preserving property of conformal parameterization, i.e., the u - and v -parameter lines are orthogonal everywhere. The traditional geometry image (GI) captures 3D geometry into three grayscale images (d)-(f) or a color image (g), in which each vertex coordinate (x, y, z) is represented by the color of a pixel (r, g, b) . In contrast, conformal geometry image (CGI) encodes the 3D shape (uniquely up to rigid motion) by its conformal factor (h) and mean curvature (i), which is more storage efficient than the traditional geometry image.

a surface can be uniquely represented (up to rigid motion) by its conformal factor λ and mean curvature H rather than the absolute coordinate (x, y, z) . The traditional geometry images have to save $3n$ data, i.e., each pixel has three channels, where n is the number of pixels in the geometry image. In sharp contrast, using conformal parameterization, we can encode 3D shape into a conformal geometry image which only takes $2n$ space without loss of any information. Figure 2 shows an example of representing the 3D model using the conformal (λ, H) representation.

3.2 Conformal representation of 3D motions

With conformal parameterization, the traditional geometry image can be converted into conformal geometry image, in which $1/3$ space can be saved immediately. Thus, it is very natural to ask how can we extend geometry video to conformal geometry video. Observe that 3D motions of articulated objects (e.g., human, animals) are approximate isometric, i.e., the geodesic distances are preserved as show in Figure 3. Since the conformal factor λ is an intrinsic feature that depends on the metric only, it is invariant to the poses as well. As a result, we only need to capture the mean curvature H in the conformal geometry video. This is characterized in the following theorem:

Theorem 2. Consider a conformally parameterized *deformable* model $s(u, v, t)$, where t is the time and (u, v) are the isothermal parameters. If the deformation $s(u, v, \cdot)$ is isometric, then the conformal factor $\lambda(u, v, \cdot)$ is independent of time t . Furthermore, the 3D motion $s(\cdot, \cdot, t)$ can be completely reconstructed by using $\lambda(\cdot, \cdot, 0)$ and $H(\cdot, \cdot, t)$ for $\forall t$.

We omit the proof as it is straightforward. The general framework for constructing conformal geometry videos is shown in Algorithm 1.

Remark 1. The above framework is very general in that various conformal parameterization algorithm can be applied depending on the geometry and/or topology of the input 3D motion,. Furthermore, algorithms may also differ in the strategies of encoding, compression and decoding. For example, different strategies of discretization of λ and H will lead to different encoding/decoding algorithms; one has many choices on the video compression techniques, such as H.264/AVC, MPEG, etc. This paper focuses on the 3D articulated motions, which can be parameterized to polycube domain. We compress CGVs using H.264/AVC and reconstruct 3D motions by an optimization approach. The details of our algorithm are presented in Section 4 and 5.

Remark 2. As the 3D shape is uniquely determined up to rigid

Algorithm 1: Constructing Conformal Geometry Videos

Parameterization (Sec 4):

Given a 3D isometric motion M with k frames, conformally parameterize M such that all the salient features are consistently mapped to the parametric domain, which guarantees the intrinsic features (e.g., metric, conformal factor, Gaussian curvature, etc) are pose invariant.

Encoding (Sec 5.1):

Compute the conformal factor λ_1 for the first frame M_1 . Compute the mean curvature $H_i, i = 1, \dots, k$ for every frame M_i .

For each frame M_i , choose four non-coplanar anchor points, and record their locations, denoted by l_i , which is used to eliminate the reconstruction ambiguity.

Output the conformal geometry video $\lambda_1, H_1, \dots, H_k$ and l_1, \dots, l_k .

Compression (Sec 5.2):

Apply the video compression technique to the mean curvature sequence $H_i, i = 1, \dots, k$.

Decoding (Sec 5.3):

To recovery the i -th frame M_i , compute the 3D shape from λ_1 and H_i .

Eliminate the rigid motion ambiguity by using the locations of four anchor points l_i .

motion (i.e., rotation and translation) by conformal factor λ and mean curvature H , we need to eliminate the rigid motion ambiguity. Note that a 3D rigid motion can be uniquely determined by given the coordinates of four non-coplanar points (a.k.a. anchor points). Comparing to the number of vertices in each frame, the storage of these four anchor points can be ignored.

Remark 3. A conformal geometry video with k frames takes $nk + n + 4k$ space, where n is the number of pixels/vertices in each frame. In contrast, the traditional geometry video takes $3nk$ space. Thus, CGVs take approximately $1/3$ space of GVs , thus, CGVs are more compact and space efficient than GVs .

Remark 4. The proposed conformal geometry videos are fundamentally different than the natural videos (NVs) and traditional geometry videos. First, CGVs contain only one channel, i.e., mean curvature H , while both GVs and NVs have three channels, i.e., (x, y, z) or (r, g, b) . Second, the conformal factor λ and mean curvature H are independent, while the three channels of GVs are not independent. Third, CGVs show stronger spatial and temporal coherence than GVs and NVs (see Section 5.2 for discussion).

Figure 4 demonstrated the CGV and GV of the deformed cloth.

Since the cloth deformation is perfectly isometric, the λ channel characterizing the intrinsic property is pose-invariant (see Row 2). Thus, CGVs only need to keep one frame of λ and all frames of H . GV, however, differ significantly for all channels for every frame.

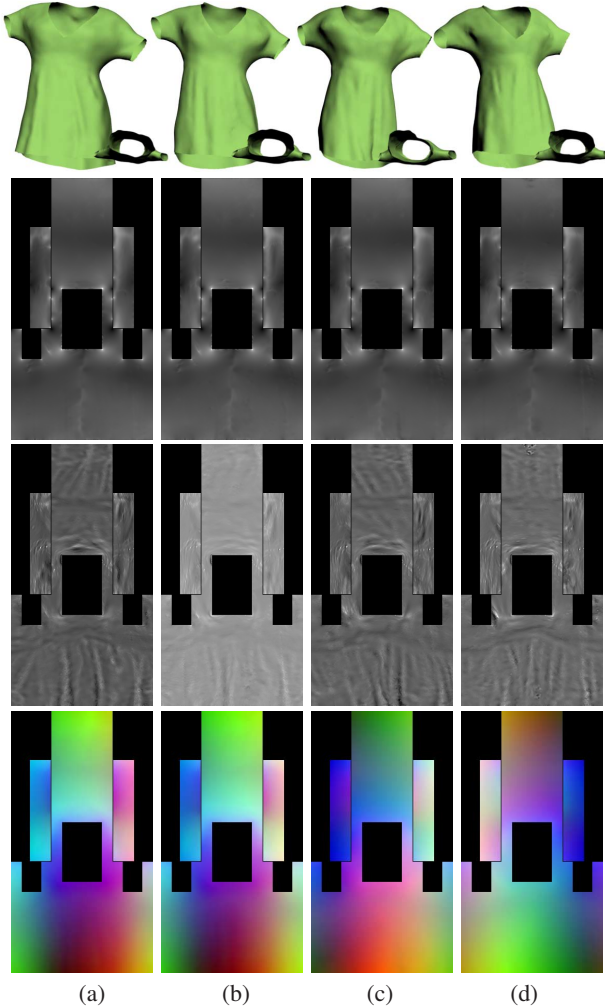


Figure 4: Geometry videos vs conformal geometry videos. The given cloth deformation (row 1) is perfectly isometric, so the conformal factor λ (row 2) is pose-invariant. As a result, CGVs only need to capture the mean curvature H (row 3), which characterizes the extrinsic feature. The traditional geometry video (row 4), in sharp contrast, differs significantly in all channels.

4. CONFORMAL PARAMETERIZATION OF 3D ARTICULATED MOTIONS

As mentioned in Section 2.2, there are many conformal parameterization algorithms available. However, most of them parameterize the 3D models into charts with irregular boundaries, which may cause numerical issues when reconstructing the curved boundary. In this paper, we use conformal polycube parameterization for two reasons: First, the parametric domain is a 3D shape that mimics the geometry of the input model, thus, the area distortion can be controlled in an easy and intuitive manner. Second, polycube has regular atlas in that each polycube face is the union of squares. Thus, each chart has axis-aligned boundaries, which facilitates the encoding and decoding.

The input of our algorithm is a sequence of captured 3D motion, each frame is represented by a triangle mesh. Note that the scanned 3D motion data are usually given in the reference system of the scanner without registration in object space, thus, they do not have correspondence between two frames. To construct conformal geometry videos, all the salient features must be *consistently* parameterized to the desired locations on the parametric domain, otherwise, the conformal factor may not be pose-invariant.

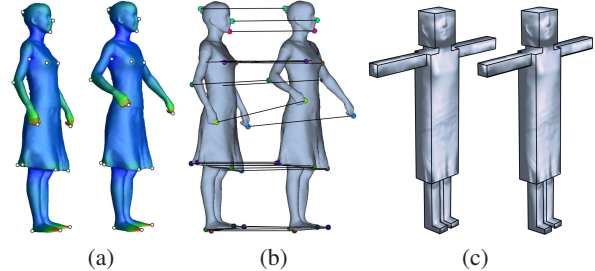


Figure 5: Feature detection and tracking. (a) The heat kernel signature (HKS), where blue (resp. red) represents low (resp. high) HKS value, and the detected feature points. (b) Feature points tracking between two adjacent frames. (c) By our constrained polycube mapping algorithm, different poses can be parameterized to the polycube in a consistent manner.

Feature detection For each frame M_i , we compute the Heat Kernel Signature (HKS) [22] $k_t^{M_i}(x, x)$ for each point $x \in M$, with a short diffusion time ($t = 0.007$ in our experiments). Then we use the local maxima of the HKS to identify the feature points [18]. In practice, we consider a point p to be a feature point on M_i if $k_t^{M_i}(p, p) > k_t^{M_i}(x, x)$ for all x in the two-ring neighborhood of point p . Figure 5(a) shows the computed HKS with relative large time and the detected geometric feature points.

Feature tracking Give two adjacent frames M_i and M_{i+1} , we detect the set of feature points $P \subseteq M_i$ and $Q \subseteq M_{i+1}$ by geometric feature detection. For each point $p \in P$, we find the correspondence $q \in Q$ using the method [18] and record the distance of the candidate pair (p, q) . We then choose the pair with the shortest distance as the ground truth matching, and find the correspondence of feature points between M_i and M_{i+1} by isometric matching [18]. Note that not every feature point in P is mapped to Q and vice versa. In practice, we discard the pairs with distance 4 times larger than that of the pair with the shortest distance. Figure 5(b) shows the correspondence among the feature points between M_i and M_{i+1} .

Constrained conformal polycube parameterization Once we find the correspondence of the salient features in the 3D motion, we can parameterize it to the polycube domain by using the editable polycube map approach [33]. This method is able to parameterize a complex 3D model to the polycube such that the user-specified constraints on the 3D model can be accurately mapped to the desired location on the polycube. Given the first frame M_i with the detected features (e.g., head, hands, feet, etc), we sketch their desired locations on the polycube, and compute the polycube map using [33]. Then we parameterize $M_i, i > 1$, to the same polycube domain with the feature constraints tracked by the heat kernel signature. Next, we apply a Laplacian smoothing to the computed polycube maps, which leads to high quality conformal parameterization (see Figure 6(d)). Finally, we cut the 3D polycube open and flatten it to a rectangle domain in \mathbb{R}^2 (see Figure 6(c)).

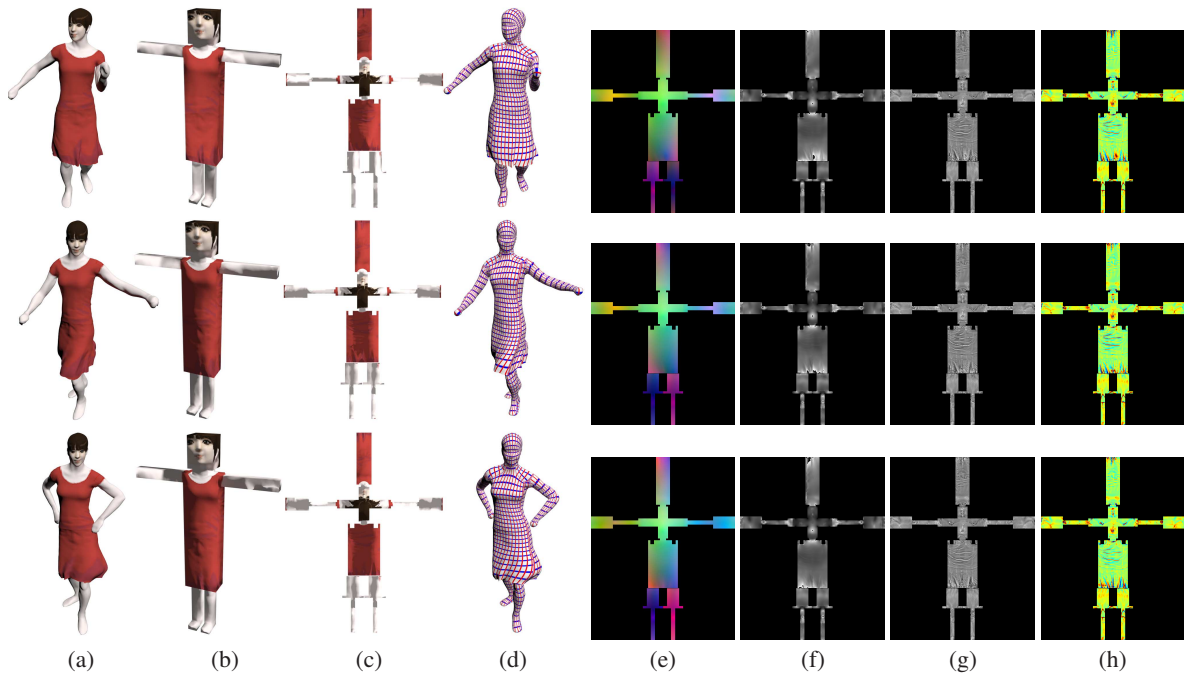


Figure 6: The samba sequence. (a) 3D motion. (b) Conformal polycube parameterization. (c) Flattened polycube to \mathbb{R}^2 . (d) Checkerboard texture mapping induced by the polycube parameterization. (e) Geometry videos (x , y and z channels). (f) and (g) show the λ and H channels of conformal geometry videos. To better visualize the pose-dependent feature of mean curvature, we render H with color in (h).

5. COMPRESSION WITH CGVS

5.1 Encoding

In the continuous setting, the conformal factor measures the scaling of the infinitesimal patches of the surface. In discrete setting, each 3D pose M_i is represented by a quadrilateral mesh. Let $Nb(v)$ denote the four neighbors of vertex v (if v is a polycube corner, $Nb(v)$ contains 3 or 5 elements). We approximate conformal factor of a vertex v as the average lengths of the edges adjacent to v , i.e.,

$$\lambda(v) = \frac{1}{|Nb(v)|} \sum_{p \in Nb(v)} l(v, p),$$

where $l(v, p)$ is the length of edge $\{v, p\}$. The discrete mean curvature of vertex v is approximated as

$$H(v) = (\Delta v, \mathbf{n}),$$

where the discrete Laplace-Beltrami operator is

$$\Delta v = v - \frac{1}{|Nb(v)|} \sum_{p \in Nb(v)} p.$$

As mentioned before, the conformal factor λ is pose-invariant if the 3D motion is perfectly isometric. However, such isometric motion is quite rare in real-world applications. In fact, most of the 3D articulated motions are approximate isometric, thus, the conformal factor λ is not a constant. To accurately reconstruct M_i ($i > 1$), we choose a set of key frames in the 3D motion, and then define the conformal factor image for each key frame. In practice, $k/10$ key frames (i.e., one key frame in every 10 frames) of conformal factor lead to satisfactory results in our experiments (see Section 6 for the details).

5.2 Compression

The conformal geometry video contains k frames of mean curvature H and m key frames of conformal factor λ , where $m (\ll k)$ is specified by the user. We partition conformal geometry video into two independent grayscale video sequences, each of them is compressed by H.264/AVC in 4:0:0 YUV format at various bitrates. For comparison, the compression of traditional geometry videos with three channels is compressed similarly except in 4:4:4 YUV format.

We have conducted a statistical analysis on the samba sequence in the format of CGVs, GVVs and NVs, respectively. As shown in Figure 7(a)-(c), the conformal factor λ and mean curvature H of CGVs are more concentrated into a smaller number of buckets. We also compared CGV to GV and NV in Figure 7(d) by using the Prewitt based histograms [35]. It has been observed that CGVs also provide more concentrated results in the histogram distributions. Furthermore, Figure 7(e) compares the mean square errors (MSEs) for the conformal factor (only applied to the key frames) and mean curvature (all frames) of CGVs, GVVs, and the corresponding NVs after a 32x32 full search motion estimation at a given QP(=8). Clearly, the MSEs of CGVs are smaller than that of the NVs and GVVs. This proves that the redundancy in CGVs can be significantly removed by motion estimation, which is the most powerful and effective tool in H.264/AVC. The above statistical results show that CGV pixels are highly correlated in both spatial and temporal domain. The results are also consistent with our subjective observations - we found that the neighboring frames of CGV are very similar in appearance, and the neighboring pixels share similar pixel values. On the other hand, the H.264/AVC standard mainly takes advantage of removing the spatial and temporal redundancies of video sequences to achieve compression. Consequently, our CGV is very suitable to be compressed with H.264/AVC.

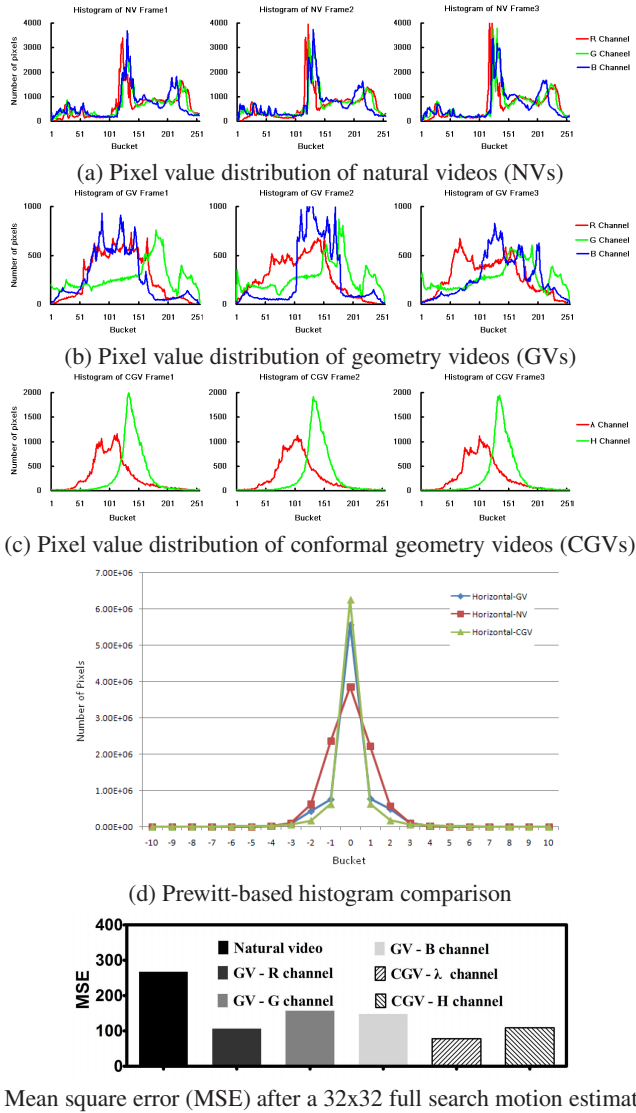


Figure 7: CGVs are fundamentally different than GV and NVs in that CGVs have stronger spatial and temporal coherence as shown by the smaller MSE. Also, the pixel values of CGVs are concentrated on a small number of buckets.

5.3 Decoding

To reconstruct the 3D geometry from the conformal geometry video, we solve the following optimization problem:

$$\min_{v_i} \sum (\|\Delta v_i\|^2 - H_i^2)^2 + \sum |\lambda(v_i)^2 - \lambda_i^2|^2, \quad (1)$$

where λ_i is the decoded conformal factor of v_i , and Δv_i gives the mean curvature vector of vertex v_i . Note that one can reconstruct the geometry uniquely up to rigid motion (i.e., rotation and translation) from (λ, H) . We eliminate the ambiguity of rigid motion by using the locations of four non-degenerate anchor points.

Note that Eqn. 1 is a non-linear programming, thus, it is difficult to obtain the global optimal in general. In our implementation, we take advantage of the spatial and temporal coherence of conformal geometry videos. As the location of four anchor points are already known, we compute the rigid motion between the anchor points of the frame M_{i-1} and frame M_i . Next we apply this rigid motion to

the reconstructed geometry M_{i-1} and then use it as the initial guess for the current frame. This strategy works very well in practice.

6. EXPERIMENTAL RESULTS

This section presents the experimental results of modeling 3D motions using our conformal geometry video framework. We also compare our CGVs with the traditional GVs.

Test sequences: We conducted experiments on 5 real-world human motion datasets (samba, march, squat, handstand and bouncing) and 1 synthetic dataset (cloth deformation). The human motion is approximately isometric transformation where the motion is composed of a few non-rigidly moving parts. The samba sequence consists of 200 frames, the march and squat 250 frames, the handstand and bouncing 174 frames. Each of frame in human motion datasets has 250k vertices. The synthetic cloth deformation is isometric, and consists of 450 frames with 160k vertices in each frame.

H.264/AVC configuration: We have used the H.264/AVC reference software (JM14.2 version) to encode the test sequences. The compression procedure is as we described in Section 5.2. To have different video qualities, we encoded each sequence at different bitrates by changing the quantization parameters (QPs) in the range of -24 to -4 (please note that we have adopted the negative QPs in order to encode high-quality videos). Each test sequence is encoded at 30 frames per second with IPBPB GOP structure, 32x32 full search motion estimation and CABAC entropy coding. The intra-frame period was set to 30 and the Rate Distortion Optimization (RDO) was turned on.

3D motion compression results: To make a fair comparison between GVs and CGVs, we chose the compressed GVs and the compressed CGVs with similar file sizes (i.e., similar bitrates). Table 1 shows the low, medium and high bitrates of GVs and CGVs that we chose for the comparison on 5 human motion sequences and the cloth sequence. Then we reconstructed the 3D motion from the compressed GVs and CGVs. For the synthetic cloth deformation which is isometric, we only need one frame of the conformal factor. For the 3D human motions which are approximate isometric, we set $k/10$ key frames (i.e., one key frame every 10 frames) for the conformal factor.

Table 1: The bitrates of GVs and CGVs of 6 tested sequences

	Low Bitrate	Medium Bitrate	High Bitrate
GV samba	1526.86	2616.42	5594.50
CGV samba	1463.30	2433.94	5013.71
GV march	1104.69	1886.72	3060.75
CGV march	965.73	1852.06	2975.82
GV squat	1148.10	1960.72	3308.69
CGV squat	1097.78	1870.92	3290.72
GV handstand	1087.69	1939.89	3241.26
CGV handstand	940.54	1606.60	2820.94
GV bouncing	1270.20	2247.14	3745.29
CGV bouncing	1184.47	2223.89	3826.06
GV cloth	1249.14	2173.03	3763.88
CGV cloth	1212.08	2137.01	3004.56

Figures 8 and 9 show the experimental results on two real datasets and one synthetic dataset. We observed that both CGVs and GVs perform well in high bitrates. However, GVs perform poorly in low bitrates, see the artifacts (non-smoothness, self-intersection, fold-over, etc) in the reconstructed 3D motion. In contrast, CGVs generate 3D motions with much better quality even in low bitrates. For quantitative comparison between GVs and CGVs, we measured the

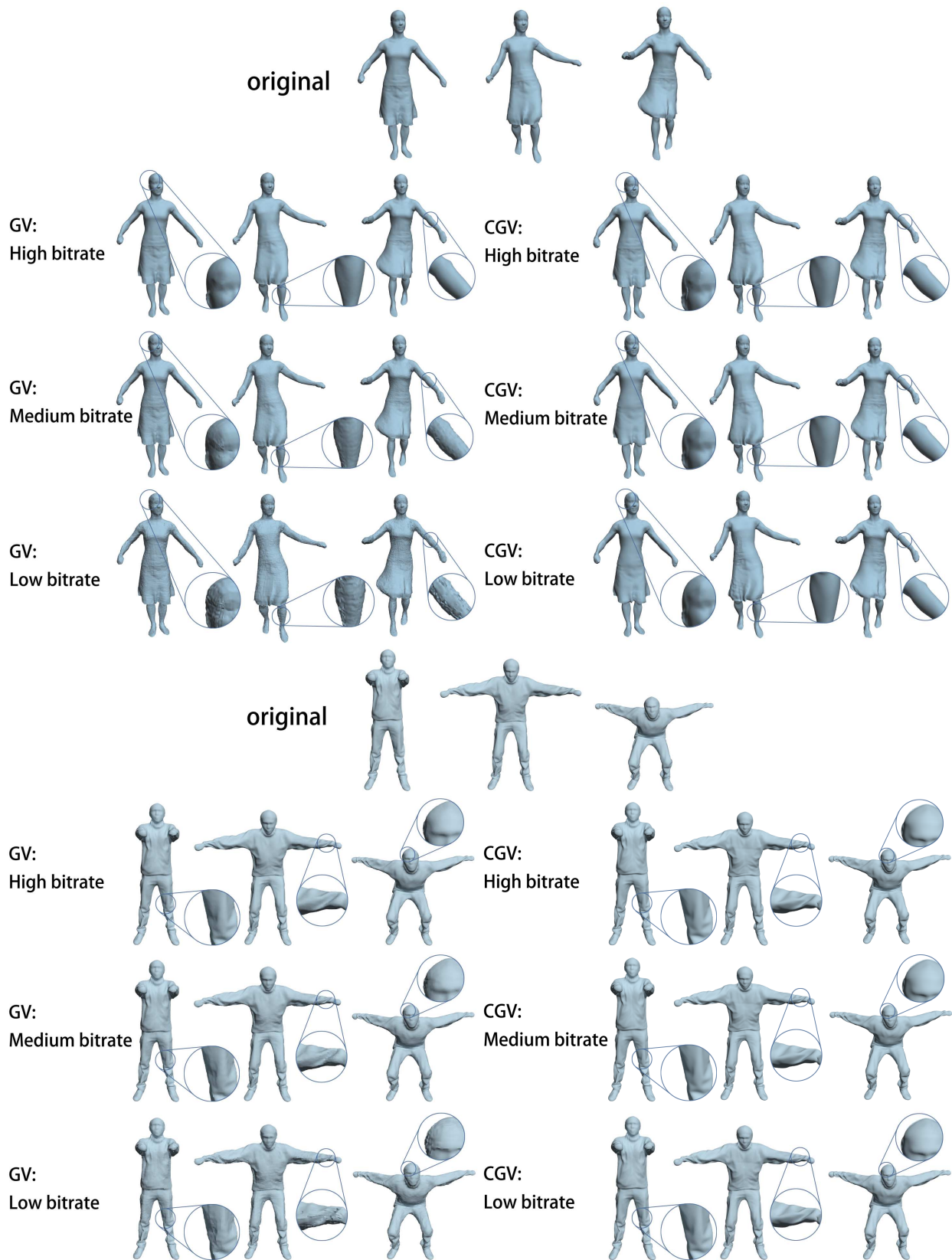


Figure 8: Experimental results on the samba (top) and squat (bottom) sequences with high, medium and low bitrates for GVs and CGVs. Note that the quality of CGVs generated 3D motion does not drop too much when bitrate is reduced. In contrast, GVs perform poorly on low bitrates. Images are rendered in high resolution that allows close-up evaluation.

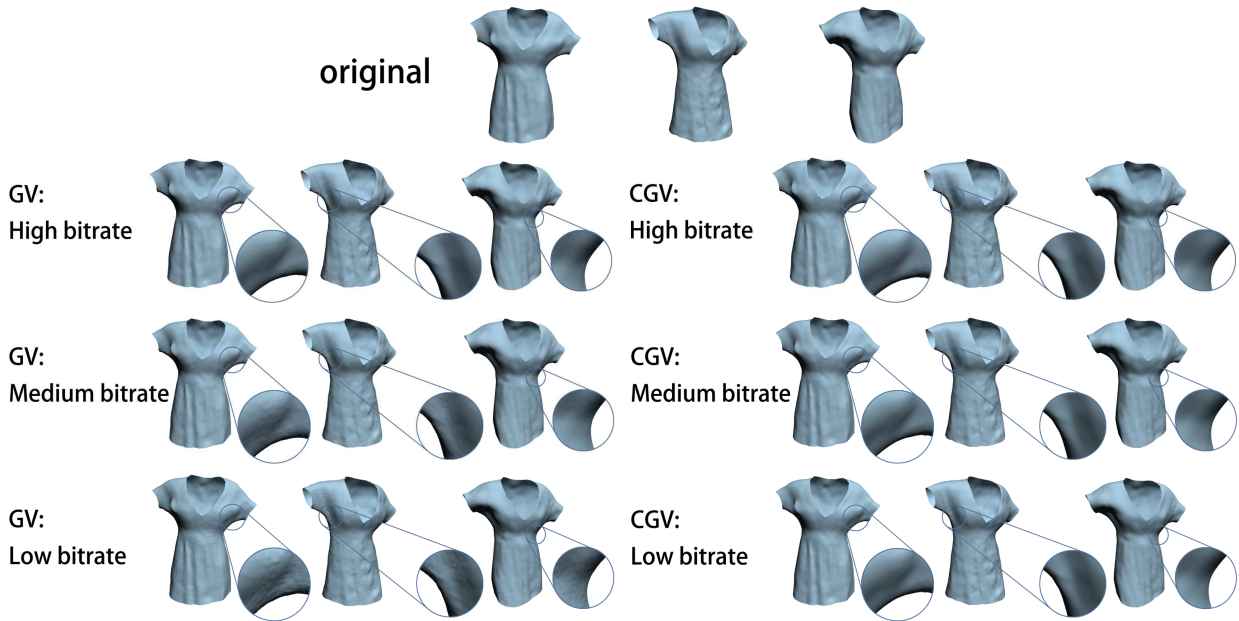


Figure 9: The cloth sequence with high, medium and low bitrates for GVs and CGVs. Note the artifacts (such as non-smoothness, self-intersection, fold-over, etc) on the GVs with low bitrates, while CGVs with similar bitrates lead to visually pleasing results without obvious artifacts.

compression quality by computing the smoothness metric [12],

$$\sqrt{\frac{1}{|V|} \sum_{i=1}^{|V|} \|\Delta v_i - \Delta v'_i\|^2},$$

where v_i and v'_i are the vertices in the original and reconstructed meshes respectively. This smoothness metric characterizes the local change of differential coordinates and is used for measurement of shape quality. The smaller the value of the smoothness metric, the closer the reconstructed model to the original model and visually more pleasing. Conversely, the larger the value of the smoothness metric, the less smooth the reconstructed data (e.g. blocking artifacts) and visually less pleasing. As we show in the Table 2 the normalized smoothness metric of all the datasets, CGVs consistently outperform GVs since GVs encode $x/y/z$ directly which is very sensitive to the block artifacts in low and medium bitrates. In sharp contrast, CGVs encode the 3D geometry into conformal factor and mean curvature, and reconstruct $x/y/z$ by solving an optimization problem that naturally preserves the smoothness of the 3D motion (based on the assumption that 3D articulated motions are spatially smooth, which is true for 3D human and animals). Thus, CGVs are visually more pleasing than GVs.

The time statistics were measured in seconds on a PC with 2.66 GHz CPU and 4GB RAM: Feature detection takes 9s on the first frame and tracking in subsequent frames is done in real time, parameterization 1s/frame, compression 12s/frame, decompression real time, 3D reconstruction 23s/frame. In reconstruction part, our current solver is conjugate gradient with linear convergence rate. Other advanced solvers (e.g. BFGS) can significantly improve the performance.

7. CONCLUSIONS

This paper presented conformal geometry videos (CGVs), a novel extension of the traditional geometry videos by taking into the consideration of the isometric nature of 3D articulated motions. We showed that the 3D articulated motion can be uniquely determined

Table 2: The normalized smoothness metric of GVs and CGVs of 6 tested sequences, with unit 10^{-4}

	Low Bitrate	Medium Bitrate	High Bitrate
GV samba	6.51	2.69	1.79
CGV samba	2.97	1.96	1.72
GV march	8.68	4.91	3.20
CGV march	3.46	2.01	1.85
GV squat	6.86	4.01	2.75
CGV squat	2.21	1.17	1.16
GV handstand	9.82	5.55	3.53
CGV handstand	3.71	2.89	2.68
GV bouncing	9.54	5.36	3.40
CGV bouncing	5.63	4.49	4.24
GV cloth	2.20	1.19	0.73
CGV cloth	0.76	0.56	0.46

(up to rigid motion) by (λ, H) , where λ is the conformal factor characterizing the intrinsic property of the 3D motion, and H the mean curvature characterizing the extrinsic feature (i.e., embedding or appearance). The rigid motion ambiguity can be easily eliminated by the locations of four non-degenerate anchor points. Since the conformal factor λ is pose-invariant, CGVs take only 1/3 the storage requirement of the traditional GVs. Our statistical tests showed that CGVs have strong spatial and temporal coherence, which is highly desirable for H.264/AVC compression. We applied conformal geometry videos to model real-world 3D human motions and synthetic cloth deformation and demonstrated that that CGVs outperform GVs with better quality and less space complexity.

There are several research directions that are worth to investigate in the future. First, as the real-world 3D motions are only approximate isometric, we need to define the conformal factor for a certain number of key frames. In our current implementation, we set one key frame among every 10 frames. Although simple, this strategy is obviously far from optimal, as the locations of key frames are data dependent. We will develop an optimization approach to determine the locations of the key frames. Second, decoding 3D motion

from λ and H is to solve a degree 4 nonlinear programming. We use the conjugate gradient approach in our current implementation. Adopting advanced solvers or GPU based solvers can improve the performance. Third, we assume that the input 3D motion data are closed and error-free. However, the captured 3D motions always have various defects, such as holes, noise, etc. It would be interesting to investigate techniques for parameterizing and modeling such incomplete data.

Acknowledgement

This work was supported by National Research Foundation (NRF) under grant NRF2008IDM-IDM004-006. The 3D human motion and cloth datasets are courtesy of [24] and [1] respectively. We would like to thank the anonymous reviewers for their constructive comments.

8. REFERENCES

- [1] D. Bradley, T. Popa, A. Sheffer, W. Heidrich, and T. Boubekeur. Markerless garment capture. *ACM Trans. Graph.*, 27(3), 2008.
- [2] H. M. Briceño, P. V. Sander, L. McMillan, S. Gortler, and H. Hoppe. Geometry videos: a new representation for 3d animations. In *SCA*, pages 136–146, 2003.
- [3] X. Gu, S. J. Gortler, and H. Hoppe. Geometry images. *ACM Trans. Graph.*, 21:355–361, July 2002.
- [4] X. Gu, Y. Wang, and S.-T. Yau. Geometric compression using Riemann surface structure. *Communications in Information and Systems*, 3(3):171–182, 2003.
- [5] X. Gu and S.-T. Yau. Global conformal surface parameterization. In *SGP*, pages 127–137, 2003.
- [6] S.-R. Han, T. Yamasaki, and K. Aizawa. Time-varying mesh compression using an extended block matching algorithm. *TCSVT*, 17(11):1506–1518, 2007.
- [7] S.-R. Han, T. Yamasaki, and K. Aizawa. Geometry compression for time-varying meshes using coarse and fine levels of quantization and run-length encoding. In *ICIP*, pages 1045–1048, 2008.
- [8] Y. He, H. Wang, C.-W. Fu, and H. Qin. A divide-and-conquer approach for automatic polycube map construction. *Comput. Graph.*, 33:369–380, 2009.
- [9] H. Hoppe and E. Praun. *Advances in Multiresolution for Geometric Modelling*, chapter Shape compression using spherical geometry images., pages 27–46. Springer-Verlag, 2003.
- [10] ITU-T. The h.264/avc standard. *ITU-T Rec. H.264*, 2008.
- [11] F. Kälberer, M. Nieser, and K. Polthier. Quadcover - surface parameterization using branched coverings. *Comput. Graph. Forum*, 26(3):375–384, 2007.
- [12] Z. Karni and C. Gotsman. Spectral compression of mesh geometry. In *SIGGRAPH*, pages 279–286, 2000.
- [13] T. P. Koninckx and L. Van Gool. Real-time range acquisition by adaptive structured light. *TPAMI*, 28:432–445, 2006.
- [14] Y.-K. Lai, M. Jin, X. Xie, Y. He, J. Palacios, E. Zhang, S.-M. Hu, and X. Gu. Metric-driven rosy field design and remeshing. *TVCG*, 16:95–108, 2010.
- [15] H. Li, B. Adams, L. J. Guibas, and M. Pauly. Robust single-view geometry and motion reconstruction. *ACM Trans. Graph.*, 28(5), 2009.
- [16] J. Lin, X. Jin, Z. Fan, and C. C. L. Wang. Automatic polycube-maps. In *GMP*, pages 3–16, 2008.
- [17] N. J. Mitra, S. Flöry, M. Ovsjanikov, N. Gelfand, L. Guibas, and H. Pottmann. Dynamic geometry registration. In *SGP*, pages 173–182, 2007.
- [18] M. Ovsjanikov, Q. Mérigot, F. Mèvoli, and L. Guibas. One point isometric matching with the heat kernel. *Comput. Graph. Forum*, 29:1555–1564, 2010.
- [19] G. Peyré and S. Mallat. Surface compression with geometric bandelets. *ACM Trans. Graph.*, 24:601–608, 2005.
- [20] A. Sheffer, E. Praun, and K. Rose. Mesh parameterization methods and their applications. *Found. Trends. Comput. Graph. Vis.*, 2:105–171, 2006.
- [21] G. Sullivan and T. Wiegand. Video compression - from concepts to the H.264/AVC standard. In *Proceedings of the IEEE*, pages 18–31, 2005.
- [22] J. Sun, M. Ovsjanikov, and L. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *SGP*, pages 1383–1392, 2009.
- [23] M. Tarini, K. Hormann, P. Cignoni, and C. Montani. Polycube-maps. *ACM Trans. Graph.*, 23(3):853–860, 2004.
- [24] D. Vlastic, I. Baran, W. Matusik, and J. Popovic. Articulated mesh animation from multi-view silhouettes. *ACM Trans. Graph.*, 27(3), 2008.
- [25] D. Vlastic, P. Peers, I. Baran, P. Debevec, J. Popović, S. Rusinkiewicz, and W. Matusik. Dynamic shape capture using multi-view photometric stereo. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 28(5), Dec. 2009.
- [26] S. Wan, Z. Yin, K. Zhang, H. Zhang, and X. Li. A topology-preserved optimization algorithm for polycube mapping. *Computers & Graphics*, 2011.
- [27] H. Wang, Y. He, X. Li, X. Gu, and H. Qin. Polycube splines. *Computer-Aided Design*, 40(6):721–733, 2008.
- [28] H. Wang, M. Jin, Y. He, X. Gu, and H. Qin. User-controllable polycube map for manifold spline construction. In *SPM*, pages 397–404, 2008.
- [29] S. Wang, Y. Wang, M. Jin, X. Gu, and D. Samaras. 3d surface matching and recognition using conformal geometry. In *CVPR (2)*, pages 2453–2460, 2006.
- [30] Y. Wang, M. Gupta, S. Zhang, S. Wang, X. Gu, D. Samaras, and P. Huang. High resolution tracking of non-rigid motion of densely sampled 3d data using harmonic maps. *IJCV*, 76(3):283–300, 2008.
- [31] Y. Wang, X. Huang, C. Lee, S. Zhang, Z. Li, D. Samaras, D. Metaxas, A. Elgammal, and P. Huang. High resolution acquisition, learning and transfer of dynamic 3D facial expressions. In *CGF*, pages 677–686, 2004.
- [32] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra. Overview of the H.264/AVC video coding standard. *TCSVT*, 13:560–576, 2003.
- [33] J. Xia, I. Garcia, Y. He, S.-Q. Xin, and G. Patow. Editable polycube map for gpu-based subdivision surfaces. In *I3D, I3D '11*, pages 151–158, 2011.
- [34] J. Xia, Y. He, D. T. P. Quynh, X. Chen, and S. C. H. Hoi. Modeling 3d facial expressions using geometry videos. In *ACM Multimedia*, pages 591–600, 2010.
- [35] J. Xia, D. T. P. Quynh, Y. He, X. Chen, and S. C. H. Hoi. Modeling and compressing 3d facial expressions using geometry videos. *TCSVT*, accepted, 2011.
- [36] T. Yamasaki and K. Aizawa. Patch-based compression for time-varying meshes. In *ICIP*, pages 3433–3436, 2010.
- [37] Y.-L. Yang, J. Kim, F. Luo, S.-M. Hu, and X. Gu. Optimal surface parameterization using inverse curvature map. *TVCG*, 14:1054–1066, 2008.
- [38] S. Zhang and P. Huang. High-resolution, real-time 3d shape acquisition. In *CVPRW'04 Vol. 3*, page 28, 2004.