Computer Aided Geometric Design 43 (2016) 211–225

Contents lists available at ScienceDirect

Computer Aided Geometric Design

www.elsevier.com/locate/cagd

Low-latency compression of mocap data using learned spatial decorrelation transform

Junhui Hou [a], Lap-Pui Chau [a], Nadia Magnenat-Thalmann [b], Ying He [c,*]

[a] School of Electrical and Electronics Engineering, Nanyang Technological University, 639798, Singapore
[b] Institute for Media Innovation, Nanyang Technological University, 639798, Singapore
[c] School of Computer Engineering, Nanyang Technological University, 639798, Singapore

ARTICLE INFO

Article history:
Available online 13 February 2016

Keywords:
Motion capture
Data compression
Transform coding
Low latency
Optimization

ABSTRACT

Due to the growing needs of motion capture (mocap) in movie, video games, sports, etc., it is highly desired to compress mocap data for efficient storage and transmission. Unfortunately, the existing compression methods have either high latency or poor compression performance, making them less appealing for time-critical applications and/or network with limited bandwidth. This paper presents two efficient methods to compress mocap data with low latency. The first method processes the data in a frame-by-frame manner so that it is ideal for mocap data streaming. The second one is clip-oriented and provides a flexible trade-off between latency and compression performance. It can achieve higher compression performance while keeping the latency fairly low and controllable. Observing that mocap data exhibits some unique spatial characteristics, we learn an orthogonal transform to reduce the spatial redundancy. We formulate the learning problem as the least square of reconstruction error regularized by orthogonality and sparsity, and solve it via alternating iteration. We also adopt a predictive coding and temporal DCT for temporal decorrelation in the frame- and clip-oriented methods, respectively. Experimental results show that the proposed methods can produce higher compression performance at lower computational cost and latency than the state-of-the-art methods. Moreover, our methods are general and applicable to various types of mocap data.

1. Introduction

As a highly successful technique, motion capture (mocap) has been widely used to animate virtual characters in distributed virtual reality applications and networked games (Capin et al., 1999; Gutierrez et al., 2003). Due to the large amount of data and the limited bandwidth of communication network, congestion, packet loss, and delay often occur in mocap data transmission. Therefore, mocap data compression, specially lossy compression, is necessary to facilitate storage and transmission.

Thanks to its smooth and coherent nature, mocap data exhibits high degree of temporal and spatial redundancy, making compression possible. To date, many mocap compression algorithms have been proposed (see Section 2). Among these approaches, most are *sequence-based* (e.g., Chattopadhyay et al., 2007; Gu et al., 2009; Tournier et al., 2009; Lin et al., 2011;

* Corresponding author.
E-mail addresses: jhhou@ntu.edu.sg (J. Hou), elpchau@ntu.edu.sg (L.-P. Chau), nadiathalmann@ntu.edu.sg (N. Magnenat-Thalmann), yhe@ntu.edu.sg (Y. He).

http://dx.doi.org/10.1016/j.cagd.2016.02.002
0167-8396/

Váša and Brunnett, 2014; Hou et al., 2014a, 2015a) in that they process all the frames of a mocap sequence at a time. These methods are able to achieve high compression performance. However, such a good compression performance comes at a price of high latency, i.e., a large number of frames have to be captured and stored before compression, making them more suitable for efficient storage. On the other hand, the *frame-based* (e.g., Kwak and Bajic, 2011) approaches aim at time-critical applications (e.g., interactive applications) due to their no-latency nature. Unfortunately, the existing frame-based methods have poor compressing performance compared with the sequence-based methods, since they cannot explore spatial and temporal correlation well. As none of the sequence- and frame-based methods is perfect, it is natural to consider the *clip-based* (e.g., Arikan, 2006; Liu and McMillan, 2006; Chew et al., 2011) methods which segment mocap data into short clips, providing a trade-off between latency and compression performance.

In this paper, we present two efficient methods for compressing mocap data with low latency. The first method processes the data in a frame-by-frame manner, hereby compressing the data without any inherent latency at all. The second one is clip-based and can achieve higher compression performance while keeping the latency fairly low and controllable. Since mocap data exhibits some unique spatial characteristics, we propose a learned spatial decorrelation transform (LSDT) to explore the spatial redundancy. Taking the data content into account, the LSDT learns an orthogonal matrix via an ℓ_0-norm regularized optimization. Due to its data adapted nature, the proposed LSDT outperforms the commonly used data-independent transforms, such as discrete cosine transform (DCT) and discrete wavelet transform (DWT), in terms of compression performance. We also adopt a predictive coding and temporal DCT for temporal decorrelation in the frame- and clip-based methods, respectively. We observe promising experimental results and demonstrate that our methods can produce higher compression performance at lower computational cost and latency than state-of-the-art.

The rest of this paper is organized as follows: Section 2 comprehensively reviews previous work on mocap data compression. Section 3 gives the proposed frame- and clip-based methods. Section 4 shows the key component of the proposed methods, i.e., the learned spatial decorrelation transform, followed by the experimental results and discussion in Section 5. Finally, Section 6 concludes this paper.

2. Related work

All compression schemes aim at exploiting correlations among the data, so does mocap data compression. In terms of decorrelation techniques, the existing mocap data compression algorithms can be roughly classified into four groups, which are reviewed and analyzed as follows.

2.1. Principal component analysis (PCA)

As a very popular technique, principal component analysis projects the data onto few principal orthogonal bases to convert data into a smaller set of values of linearly uncorrelated data.

Breaking the mocap database into short clips that are approximated by Bézier curves, Arikan (2006) performed clustered PCA to reduce their dimensionality. Liu and McMillan (2006) projected only the keyframes on the PCA bases and interpolated the other frames via spline functions. Motivated by the repeated characteristics of human motions, Lin et al. (2011) projected similar motion clips into PCA space and approximated them by interpolating functions with range-aware adaptive quantization. Observing that distortion to each of the joints causes a different overall distortion, Váša and Brunnett (2014) proposed perception-driven error metric so that important joints have a higher precision than that of joints with small impact. They presented a Lagrange multiplier-based preprocessing for adjusting the joint precision. After Lagrangian equalization, the entire mocap sequence is projected into PCA pose space. Then, PCA is applied to short clips for further reducing the temporal coherence.

Principal geodesic analysis (PGA) is a generalization of PCA for handling the case where the data is sampled from curved manifolds. Tournier et al. (2009) presented a PGA-based method for the poses manifold in the configuration space of a skeleton, leading to a reduced, data-driven pose parameterization. Compression is then obtained by storing only the approximate parameterization along with the end-joints and root-joints trajectories.

Although PCA can decorrelate mocap data very well, its bases are data-dependent and usually difficult to compress. Therefore, one has to explicitly store the orthogonal bases, which reduces the overall compression performance. Furthermore, PCA is usually applied to the whole mocap sequence (e.g., Karni and Gotsman, 2004; Váša and Brunnett, 2014), resulting in a high latency.

2.2. Discrete wavelet and cosine transforms

DCT and DWT are commonly used techniques for converting correlated data into frequency domain, in which energy mainly concentrates on sparse frequencies (or most transform coefficients tend to zero). DCT and DWT have been widely adopted in some video/image coding standards (Wiegand et al., 2003; Sullivan et al., 2012). Moreover, they also have been exploited in the compression of 3D geometric data, e.g., static/dynamic meshes (Gu et al., 2002; Hou et al., 2014b, 2015b) and mocap data (Preda et al., 2007; Chew et al., 2011; Kwak and Bajic, 2011).

Kwak and Bajic (2011) applied 1D DCT to the predictive residuals between consecutive frames for exploiting the spatial coherence. In contrast, Preda et al. (2007) applied 1D DCT/DWT to the residuals of motion compensation along the temporal