Information Sciences xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Human motion recovery jointly utilizing statistical and kinematic information

Guiyu Xia, Huaijiang Sun*, Guoqing Zhang, Lei Feng

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, PR China

A R T I C L E I N F O

Article history:
Received 22 April 2015
Revised 18 November 2015
Accepted 25 December 2015
Available online xxx

Keywords:
Motion capture data
Sparse coding
Motion recovery
Kinematic constraint

A B S T R A C T

Human motion data that are captured by the markers attached to an actor's body have been widely used in many areas. However, occlusion caused by the actor's body or clothing might make several markers missing for a period of time during the capture process, which highlights the need for motion recovery in the human motion capture process. In recent years, low-rank matrix completion and sparse coding have been used in many data-driven motion recovery methods. However, applying them directly to recover missing data is not effective because low rank is only a basic statistical property of human motion. In addition, the dictionary is usually learned and used in a complete feature space, while human motion must be recovered from an incomplete feature space. Moreover, low-rank matrix completion and sparse coding take advantage only of the statistical property and ignore another important property, i.e., the kinematic property of human motion. Inspired by coupled dictionary learning, we modify the traditional dictionary learning process and propose a new process for the special task of motion recovery. The new recovery process jointly utilizes statistical and kinematic information. Within the proposed method, we first learn a dictionary from a large number of complete–incomplete training frame pairs, to preserve the statistical information of motion data. Then, with the smoothness constraint and the bone-length constraint which take the kinematic information into recovery process, we recover motions using sparse representations of incomplete frames and a learned dictionary through an optimization model. Additionally, we employ two gradient-based optimization algorithms for dictionary learning and motion recovery. Extensive experiment results and comparisons with four other state-of-the-art methods demonstrate the effectiveness of the proposed method.

1. Introduction

With the rapid development of virtual reality technology, human motion capture data are widely used in many areas, such as the movie industry, computer games and sports training. It is a new type of multi-media data that mainly describe human motions. Although traditional multi-media data such as videos can also record human motion with image sequences, human motion related studies based on videos, such as human motion analysis [25] and pose estimation [12,14,31], are more difficult because image data are very complicated and may include a considerable amount of useless or redundant

* Corresponding author. Tel.: +8613905172533.
E-mail addresses: xiaguiyu1989@sina.com (G. Xia), sunhuaijiang@njust.edu.cn (H. Sun), xiayang14551@163.com (G. Zhang), fenglei492327278@126.com (L. Feng).

http://dx.doi.org/10.1016/j.ins.2015.12.041
0020-0255/© 2016 Published by Elsevier Inc.

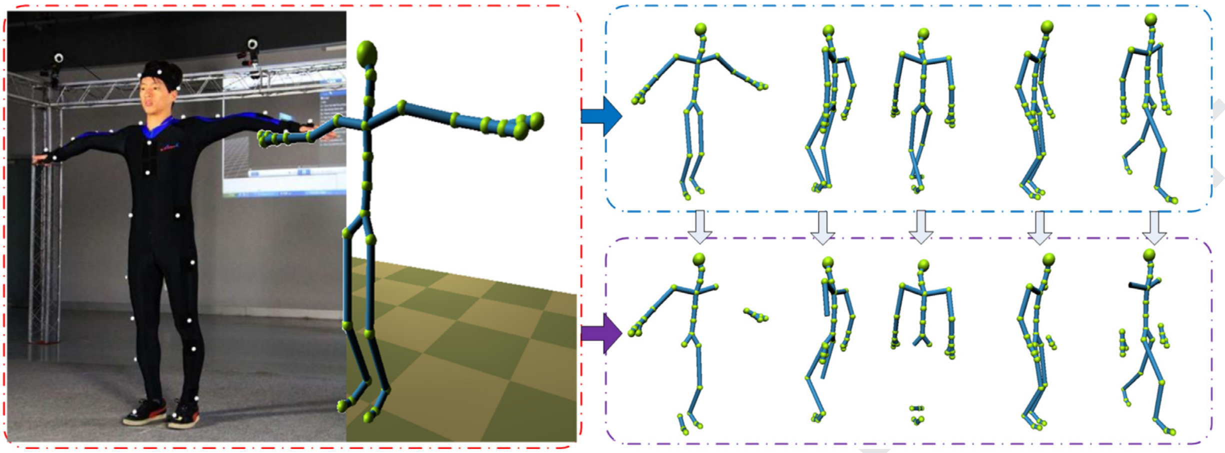

Fig. 1. Sketch of human motion capture system and motion occlusion.

information. However, human motion capture data are very clean and only contain positions or orientations of a tree structure, which corresponds to a human skeleton. A recent notable application of motion capture data is the movie *Avatar*, where the exciting applications start with an accurate acquisition of high-quality human motion data. Motion capture is a prevalent technique for capturing and analyzing human articulation. Corresponding to a temporal sequence, human motion is captured as motion frames that record the positions or orientations of all of the joints at every capturing time point. As shown in the left subfigure of Fig. 1, an optical motion capture system such as Vicon [1], utilizes video cameras to track the movements of a set of reflective markers that are strategically attached to the actor's body. However, even with the costly professional motion capture equipment, motion capture data may still be inaccurate or incomplete.

When the actor is performing, some markers may not be visible to certain cameras because of the occlusion by the actor's body or clothing, so that the captured motion is similar to the right subfigure of Fig. 1. To ensure the high-quality performance in every application, the corrupted data should be pre-processed. Therefore, an important branch of motion capture research is to handle two highly correlated and frequently co-occurred sub-problems: one is to predict missing values in motion data, and the other is to remove both the noises and outliers. The former sub-problem is called motion recovery, and the latter sub-problem is called motion denoising. Although their respective tasks are slightly different, the technology used to implement them is very similar. In this paper, we mainly focus on motion recovery; however, we do not distinguish motion recovery and motion denoising because of their similarity of the technology used.

Human motion recovery is challenging because human motion involves highly coordinated movements. Standard signal denoising technologies (if we take the missing values as noise), such as the Gaussian low-pass filter and the Kalman filter [8,42,43], often process each DOF independently so that the filtered motions often appear uncoordinated or unnatural because the spatial-temporal characteristics are undermined. Interpolation methods [13,26,34,45] can preserve the spatial-temporal characteristics embedded in natural human motion and can effectively estimate the missing marker. Although motion data are recorded in a linear space, they are essentially nonlinear. When we use interpolation methods, we should assume that the motion data are linear. Thus, these methods are not theoretically reasonable. On the other hand, assuming that motion data are linear can make recovery easy and fast in a linear space. Inspired by locally linear embedding [35], we assume motion data are locally linear even though they are a manifold structure globally, so that a locally linear method can address the motion recovery problem.

Sparse coding [30,36] has become a hot research topic in recent years, and has been used to solve many practical problems in signal processing, statistic recognition and pattern recognition [32]. Additionally, locality-constrained Linear Coding as a variant of sparse coding has even been successfully used to perform human pose estimation [37] and colorization for a gray-scale facial image [24]. The core idea of sparse coding is to find a sparse representation for the target signal using a overcomplete dictionary. Its essence is a locally linear interpolation so that atoms used in the dictionary are scaled neighbors of the target signal in fact. This characteristic of sparse coding matches our assumption perfectly. In this paper, we present a human motion recovery method that uses a dictionary to represent and recover incomplete frames. Moreover, we borrow the idea of compressive sensing [2] and image super resolution [50,51], which are very successful applications of sparse coding. Their dictionary learning processes aim to enable the dictionary to reconstruct complete signals from incomplete signals. We take the idea and propose a new dictionary learning process according to the special task of motion recovery in practice, which will be discussed in detail in Section 4.1.

Another challenge of human motion recovery is that the recovered motion must satisfy certain kinematic constraints. Therefore, the proposed method exploits not only the statistical characteristics but also the kinematic characteristics, while other data-driven methods mostly focus on the former and seldom utilize the latter. Bone length is a rigid kinematic constraint that is often ignored in many data-driven methods [10,16,26,27,46,47]. It provides some useful information that can

Please cite this article as: G. Xia et al., Human motion recovery jointly utilizing statistical and kinematic information, Information Sciences (2016), http://dx.doi.org/10.1016/j.ins.2015.12.041