Signal Processing 110 (2015) 108–122

Contents lists available at ScienceDirect

Signal Processing

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

# Sparse motion bases selection for human motion denoising

Jun Xiao [a], Yinfu Feng [a,*], Mingming Ji [a], Xiaosong Yang [b], Jian J. Zhang [b], Yueting Zhuang [a]

[a] School of Computer Science, Zhejiang University, Hangzhou 310027, PR China
[b] Bournemouth University, Poole Dorset BH12, United Kingdom

ARTICLE INFO

*Article history:*
Received 4 May 2014
Received in revised form
5 August 2014
Accepted 12 August 2014
Available online 20 August 2014

*Keywords:*
Human motion denoising
Data-driven
Poselet model
$\ell_1$-minimization

ABSTRACT

Human motion denoising is an indispensable step of data preprocessing for many motion data based applications. In this paper, we propose a data-driven based human motion denoising method that sparsely selects the most correlated subset of motion bases for clean motion reconstruction. Meanwhile, it takes the statistic property of two common noises, i.e., Gaussian noise and outliers, into account in deriving the objective functions. In particular, our method firstly divides each human pose into five partitions termed as poselets to gain a much fine-grained pose representation. Then, these poselets are reorganized into multiple overlapped poselet groups using a lagged window moving across the entire motion sequence to preserve the embedded spatial–temporal motion patterns. Afterward, five compacted and representative motion dictionaries are constructed in parallel by means of fast K-SVD in the training phase; they are used to remove the noise and outliers from noisy motion sequences in the testing phase by solving $\ell_1$-minimization problems. Extensive experiments show that our method outperforms its competitors. More importantly, compared with other data-driven based method, our method does not need to specifically choose the training data, it can be more easily applied to real-world applications.


## 1. Introduction

Motion capture is a powerful and mature technique for creating realistic computer character animation. It has been widely adopted in a large variety of applications such as animation production, computer games, human–computer interaction and medical rehabilitation [1–3]. In these applications, high-quality motion data are demanded for the purpose of accurate motion analysis and generation. However, even with a highly professional motion capture system there are many instances where some markers are occluded or mismatched [4–7]. As a result, it is necessary to fill the missing markers, while it may result in a certain percentage of noise. On the other hand, if some markers are mismatched when the tacking algorithm confuses the trajectory of one marker with that of another in some cases, the captured motion data contain serious error which can be regarded as bad noise or outliers. In order to clean the noisy data, most of the commercial motion capture systems provide various post process softwares for editing motion data including filling missing values and removing noise. To undertake the task of motion editing, the user must be patient and have professional knowledge of human motion capture. The underlying denoising/smoothing methods of these softwares mainly derive from linear and/or nonlinear interpretation methods, which suggests that they are only

* Corresponding author.
*E-mail addresses:* junx@zjuem.zju.edu.cn (J. Xiao), fyf200502@hotmail.com (Y. Feng), mjcs@zju.edu.cn (M. Ji), xyang@bournemouth.ac.uk (X. Yang), jzhang@bournemouth.ac.uk (J.J. Zhang), yzhuang@zju.edu.cn (Y. Zhuang).

http://dx.doi.org/10.1016/j.sigpro.2014.08.017
0165-1684 

efficient for dealing with simple or short-term noise cases. When these methods are used to handle some complex or long-term noise cases, the filtered motion will be distorted and unrealistic. That is to say, they may fail under these circumstances. Moreover, it is time-consuming and error-prone to process the noisy motion data in manual [8].

Meanwhile, more and more low-cost depth sensors (e.g., the Microsoft Kinect and SoftKinect) that can acquire the depth stream with acceptable accuracy have been released in recent years. With the aid of these new-fashioned products, many classic difficult computer vision problems like background subtraction and human detection become tractable. It also provides new opportunities for developing accessible motion capture. In light of this, some new algorithms have been proposed to recover human motion from the depth stream in real-time [9,10]. Compared with the traditional motion capture techniques such as the optical-based motion capture, the motion data generated from the depth stream are more likely to contain noise and outliers. For instance, if an actor performs the freestyle swimming action in front of a Microsoft Kinect, the recovered motion of the actor's two hands will be distorted due to the reason of self-occlusion of human body parts. In fact, researchers still have an uphill journey in improving the quality of these newly generated motion data.

To improve the aforementioned issues, a lot of researchers have plunged into the topic of motion data denoising over the years. Through a great deal of effort, a number of human motion denoising methods and techniques have been proposed. However, some intrinsic shortcomings of these methods hinder them from being widely applied in real-world applications. Take the popular signal-based denoising methods (e.g., Gaussian low-pass filter and discrete cosine transform (DCT)) for example, although they are easy to implement and only require a little of computational cost, they ignore the underlying structure correlation between different human joints and cannot preserve the embedded spatial–temporal motion patterns [11–15]. Indeed, human motion involves highly coordinated movement and the movement between different human joints are not independently [8]. As an improvement, the dynamic system based methods represented by Kalman filter and linear dynamic system (LDS) have been developed to discover hidden variables and learn their dynamics [16,17]. But a little of time delay will appear after motion denoising with the dynamic system based methods [18].

On the other hand, with the explosive growth of the available motion capture data in recent years, data-driven based motion denoising methods have attracted much attention [8,19]. Lou and Chai [8] proposed an example-based data-driven method to learn a series of filter bases, which hold some spatial–temporal patterns embedded in precaptured motion data, and then use them along with the robust statistics technique to filter noisy motion data [8]. Their method received encouraging results both on the real and simulated motion data. However, they use all of the learned filter bases to reconstruct the clean motion sequences indiscriminately, so their training database must be behavior-specific and typically only contains motion data selected from the same action with different style variants. Otherwise, the performance of their method will decline significantly since the bases learned from motion data with different action contain significantly different spatial–temporal patterns.

To overcome the shortcoming of [8], we propose a new data-driven based human motion denoising method in this paper. The key ideas of our paper are in twofold: sparsely selecting the most correlated subset of motion bases for clean motion reconstruction and taking the statistic property of motion noise into account in deriving our objective functions. The flowchart of our proposed method is illustrated in Fig. 1. And, the major contributions of our proposed method are summarized as follows.

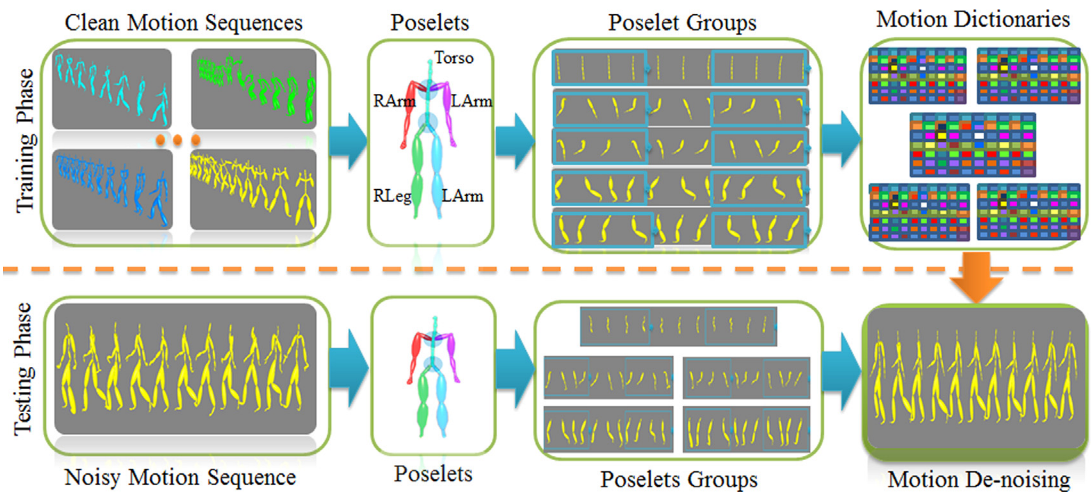


**Fig. 1.** The illustration of our proposed human motion data denoising framework. For the input motion sequences, we first divide each human pose into five partitions, which are termed as poselets. These poselets are then grouped together using a lagged window moving through the entire motion sequences to generate poselet groups. In the training phase, we use these poselet groups to learn five motion dictionaries and adopt these learned motion dictionaries to remove the noise and outliers from noisy poselet groups in the testing phase. Finally, we reorganize the filtered poselet groups to reconstruct the clean motion sequences.