Computer Vision and Image Understanding 132 (2015) 75–86

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Efficient tracking of human poses using a manifold hierarchy ☆

Alexandros Moutzouris [a,*], Jesus Martinez-del-Rincon [b], Jean-Christophe Nebel [a], Dimitrios Makris [a]

[a] Digital Imaging Research Centre, Kingston University, United Kingdom
[b] Queen's University, Belfast, United Kingdom

ARTICLE INFO

Article history:
Available online 25 October 2014

Keywords:
Human body pose estimation
Pose tracking
Dimensionality reduction

ABSTRACT

In this paper a 3D human pose tracking framework is presented. A new dimensionality reduction method (Hierarchical Temporal Laplacian Eigenmaps) is introduced to represent activities in hierarchies of low dimensional spaces. Such a hierarchy provides increasing independence between limbs, allowing higher flexibility and adaptability that result in improved accuracy. Moreover, a novel deterministic optimisation method (Hierarchical Manifold Search) is applied to estimate efficiently the position of the corresponding body parts. Finally, evaluation on public datasets such as HumanEva demonstrates that our approach achieves a 62.5–65 mm average joint error for the walking activity and outperforms state-of-the-art methods in terms of accuracy and computational cost.

1. Introduction

In computer vision, human motion analysis is a rapidly growing area, because of not only the methodological challenge that motion analysis implies but also its many applications such as surveillance, human computer interaction, sports analysis and computer graphics. Human motion analysis approaches have been relying on activity recognition, pose estimation and pose tracking methodologies. Whereas activity recognition aims at classifying the type of the activity performed by a human being, pose estimation deals with estimating the skeletal position of a person for one or more frames. Although this makes the solution independent of the previous poses, this is also more sensitive to errors. Consequently, pose estimation is often integrated within a pose tracking framework where past information is exploited to estimate the current pose in a more efficient way. Human pose tracking methods that rely on markerless approaches are generally desirable because of their non-invasive nature that widens significantly their potential application. Multi-camera systems are able to mitigate the complexity of markerless approaches and to deal with the inevitable limb occlusions. Still, the variety of human postures and activity styles, and the high complexity of modelling the human body make this problem a technically demanding and computationally expensive task.

In this paper, we introduce a hierarchical dimensionality reduction method, namely Hierarchical Temporal Laplacian Eigenmaps (HTLE). It goes beyond the hierarchical structure of human body parts, represented as pairwise relationships as in Yang and Lee [38], Urtasun and Darrell [33], Amin et al. [1] by considering divisions at different hierarchical levels, similarly to Han et al. [14], Darby et al. [10], Raskin and Rudzsky [25], Wang et al. [37], Tian et al. [32]. The HTLE approach allows searching each level of a posture hierarchy separately, thus modelling new, unseen poses. Furthermore, we propose a markerless, hierarchical pose tracking method, namely Hierarchical Manifold Searching (HMS), designed for multi-camera scenarios. Our framework operates in a two-phase approach; first, a training set is used in order to generate a hierarchy of low dimensional manifolds using HTLE and second, pose tracking is performed in a hierarchical manner using HMS. Therefore, unlike conventional dimensionality reduction methods which are restricted to the set of poses present in a training set, our framework is capable of moving beyond the training set and generating new poses that have never been seen before. In addition, instead of searching the whole hierarchy as performed in previous studies using computationally expensive particle filtering [10,25], the computational cost of the proposed method is reduced by using deterministic optimisation applied to a subset of manifolds in the hierarchy. Our approach also deals with style variability, i.e. pose differences for a given activity resulting from either individual's personality or distinct conditions, by allowing an extra final level of hierarchy where each body part is individually adjusted in an unconstrained manner [24]. We show that our methodology improves computational efficiency and accuracy.

☆ This paper has been recommended for acceptance by Edwin R. Hancock.

* Corresponding author.
E-mail addresses: moutzouris.a@gmail.com (A. Moutzouris), j.martinez-del-rincon@qub.ac.uk (J. Martinez-del-Rincon), J.Nebel@kingston.ac.uk (J.-C. Nebel), D.Makris@kingston.ac.uk (D. Makris).

http://dx.doi.org/10.1016/j.cviu.2014.10.005
1077-3142/

1.1. Related work

First, different 3D human pose tracking techniques are discussed. Then, methodologies based on dimensionality reduction methods are presented. Finally, hierarchical approaches are described.

Early human pose tracking methods were based on gradient descent [16]. However, they suffer from finding local optima thus giving poor tracking results. In order to search the complex space of human postures, the particle filter (PF) method [2] has been used. However, the high dimensionality of this space makes it difficult to sample the solution space efficiently [28] and prevents divergence. Deutscher and Reid [11] proposed an Annealed Particle Filter (APF) that improves the efficiency of the particle filter search; however it is still computational expensive due to the high dimensionality. Moreover, in an unconstrained searching environment, when a particle filter-based tracker diverges, convergence in the following frames becomes problematic [26]. Gall et al. [13] introduce a multi-layer framework based on simulated annealing that combines stochastic optimisation, filtering, and local optimisation which led to results slightly more accurate than APF. However, as all particle based estimation methods, a large number of particles is required which increases complexity and computational cost [28,4].

In order to deal with the high complexity of modelling articulated human motion, nonlinear dimensionality reduction methods have been used in tracking pipelines, exploiting available training sequences for known actions. They are grouped into two categories: mapping-based and embedded-based approaches. Mapping approaches, such as Gaussian Process Latent Variable Model (GP-LVM) [18,15], employ probabilistic nonlinear functions in order to map the embedded space to the data space. Consequently, their training is time-consuming and convergence is not guaranteed [35], especially for applications which are based on large training sets. Embedded approaches provide an estimate of the structure of the underlying manifold by means of approximating each data point according to their local neighbours on the manifold. The main drawback of these methods is the lack of mapping functions between high and low dimensional spaces, although Radial Basis Function Networks are usually used to resolve this issue [20]. This category of techniques includes Local, Linear Embedding [27] (LLE), Isometric Feature Mapping (Isomap) [31], Laplacian Eigenmaps (LE) [5] and Local tangent space alignment (LTSA) [39].

Since human motion may be described by time series, the temporal dependencies between consecutive poses can be used to improve human tracking applications. These temporal constrains ensure that points that are close in time will be close in the low dimensional space. Spatio-Temporal Isomap (ST-Isomap), [17] an extension of Isomap, changes the original weights in the graph of local neighbours in order to emphasise the similarity between temporally related points. Gaussian Process Dynamical Models (GPDM) [36] integrates time information using Gaussian Process priors to create dynamics in the low dimensional space. Urtasun et al. [34] use GPDM for learning human poses and motion priors for 3D people tracking. However, most of these methods suffer from the fact they are person dependent: they are not able to efficient track people with their corresponding style who do not belong to the training set, which reduces their application. Alternatively, Temporal Laplacian Eigenmaps (TLE) [22] was specifically designed to address the issue of modelling activities of different people by suppressing their stylistic differences and producing a coherent manifold. The resulting manifold has a 1D dimensionality, which is suitable for fast exploration. Nonetheless, none of the above approaches allows the recovery of unseen poses. This is because dimensionality reduction methods are activity dependent, that is, they can only represent those activities that they have learned during training, usually a single activity.

Hierarchical methodologies that consider divisions of human body parts at different hierarchical levels have been proposed to extend the pose space by decoupling the motion of individual limbs which allows dealing with unseen activities. Such methodologies were proposed for 2D pose estimation in Wang et al. [37], Tian et al. [32]. The Hierarchical Gaussian Process Latent Variable Model (H-GPLVM) [19] has been applied to activity recognition [14] and pose estimation [25,10], based on a hierarchy of manifolds trained using different activities. Han et al. [14] and Darby et al. [10] used H-GPLVM for training two different activities and the APF method to search for poses that result from combinations of these activities. Using this learnt hierarchical model for multiple activities they can recover novel poses which are not present in the training dataset. For example, training data for a person walking and a person standing and waving allow detecting a person who is walking whilst waving. The hierarchy is able to recognise the posture of the upper body from the first training activity and that of the lower body from the other one. Similarly, Raskin and Rudzsky [25] presented an extension of the Gaussian Process Annealed Particle Filter (GPAPF) method [26] called Hierarchical Annealing Particle Filter (H-APF). This method also uses H-GPLVM to generate a hierarchy of manifolds in the low-dimensional space, and the APF method to generate particles in the latent space. H-APF is tested in a multi-activity scenario combining walking and jogging activities, where the activity of every frame is estimated before pose estimation. Although all these hierarchical approaches allow the generation of unseen poses where individual body part postures originally belonged to different activities, their main drawback is their high computational cost, since APF is used to search through the whole hierarchy.

1.2. Overview

The pipeline of our approach is presented in Fig. 1. More specifically, the training set comes from MoCap data describing the activity of interest as a sequence of human poses. Activity Manifolds are learned by the proposed Hierarchical Temporal Laplacian Eigenmaps (HTLE) (Fig. 1a), as described in Section 2. The pose

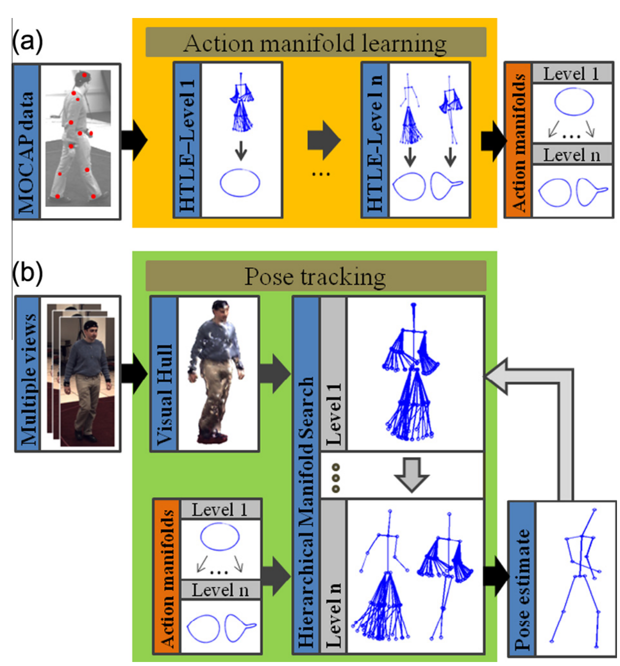

Fig. 1. (a) Training and (b) pose tracking pipelines.