Image and Vision Computing 32 (2014) 131–154

Contents lists available at ScienceDirect

Image and Vision Computing

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

3D human motion analysis framework for shape similarity and retrieval ☆

Rim Slama [a,b], Hazem Wannous [a,b,*], Mohamed Daoudi [a,c]

[a] LIFL laboratory (UMR CNRS 8022), Villeneuve d'Ascq, France
[b] University of Lille 1, Villeneuve d'Ascq, France
[c] Institut Mines-Télécom/Télécom Lille, Villeneuve d'Ascq, France

ARTICLE INFO

Article history:
Received 30 March 2013
Received in revised form 3 December 2013
Accepted 30 December 2013

Keywords:
Motion analysis
Shape similarity
3D video retrieval
3D human action

ABSTRACT

3D shape similarity from video is a challenging problem lying at the heart of many primary research areas in computer graphics and computer vision applications. In this paper, we address within a new framework the problem of 3D shape representation and shape similarity in human video sequences. Our shape representation is formulated using extremal human curve (EHC) descriptor extracted from the body surface. It allows taking benefits from Riemannian geometry in the open curve shape space and therefore computing statistics on it. It also allows subject pose comparison regardless of geometrical transformations and elastic surface change. Shape similarity is performed by an efficient method which takes advantage of a compact EHC representation in open curve shape space and an elastic distance measure. Thanks to these main assets, several important exploitations of the human action analysis are performed: shape similarity computation, video sequence comparison, video segmentation, video clustering, summarization and motion retrieval.

Experiments on both synthetic and real 3D human video sequences show that our approach provides an accurate static and temporal shape similarity for pose retrieval in video, compared with the state-of-the-art approaches. Moreover, local 3D video retrieval is performed using motion segmentation and dynamic time warping (DTW) algorithm in the feature vector space. The obtained results are promising and show the potential of this approach.

1. Introduction

While human analysis in 2D image and video has received a great interest during the last two decades, 3D human body is still a little explored field. Relatively few authors have so far reported works on static analysis of 3D human body, but even fewer on 3D human video analysis.

Parallel to this, 3D video sequences of human motion are more and more available. In fact, their acquisition by multiple view reconstruction systems or animation and synthesis approaches [1,2] received a considerable interest over the past decade, following the pioneering work of Kanade [3].

Most of the recent research topics on 3D video focus mainly on performance, quality improvements and compression methods [4,2,5]. Consequently, 3D videos are yet mainly used for display. However, the acquisition of long sequences produces massive amounts of data which necessitates efficient schemes for navigating, browsing, searching, and viewing video data. Hence, we need to develop an efficient and effective descriptor to represent body shape and pose for shape retrieval and video clustering. We also need a motion retrieval system to look for relevant information quickly.

3D Human body shape similarity is an important area, recently attracted more attention in the field of human–computer interface (HCI) and computer graphics, with many related research studies. Among these, research started with 3D features have been applied for body pose estimation and 3D video analysis.

In this paper, a unified framework providing several processing modules is presented. All viewed within a duality pose/motion approach as summarized in Fig. 1 below.

We first focus on the analysis of human pose and we propose a novel 3D human curve-based shape descriptor called extremal human curves (EHC). This descriptor, extracted on body surface, is based on extremal features and geodesics between them. Every 3D mesh is represented by a collection of these open curves. The mesh to mesh comparison is then performed in a Riemannian shape space using an elastic metric between each two correspondent human curves.

At this level, our ultimate goal is to be able to perform reliable reduced representation based on geodesic curves for shape and pose similarity metric. Invariant to pose changes, our EHC descriptor allows pose (and motion) comparison of subjects regardless of translation, rotation and scaling. Such descriptor can be employed not only in pose retrieval for video annotation and concatenation but also in motion retrieval, clustering and activity analysis.

☆ This paper has been recommended for acceptance by Matthew Turk.

* Corresponding author at. Rue Marconi, Cité Scientifique, 59650 Villeneuve-d'Ascq, France.
E-mail addresses: rim.slama@telecom-lille.fr (R. Slama), hazem.wannous@telecom-lille.fr (H. Wannous), mohamed.daoudi@telecom-lille.fr (M. Daoudi).

0262-8856/$ – see front matter.
http://dx.doi.org/10.1016/j.imavis.2013.12.011

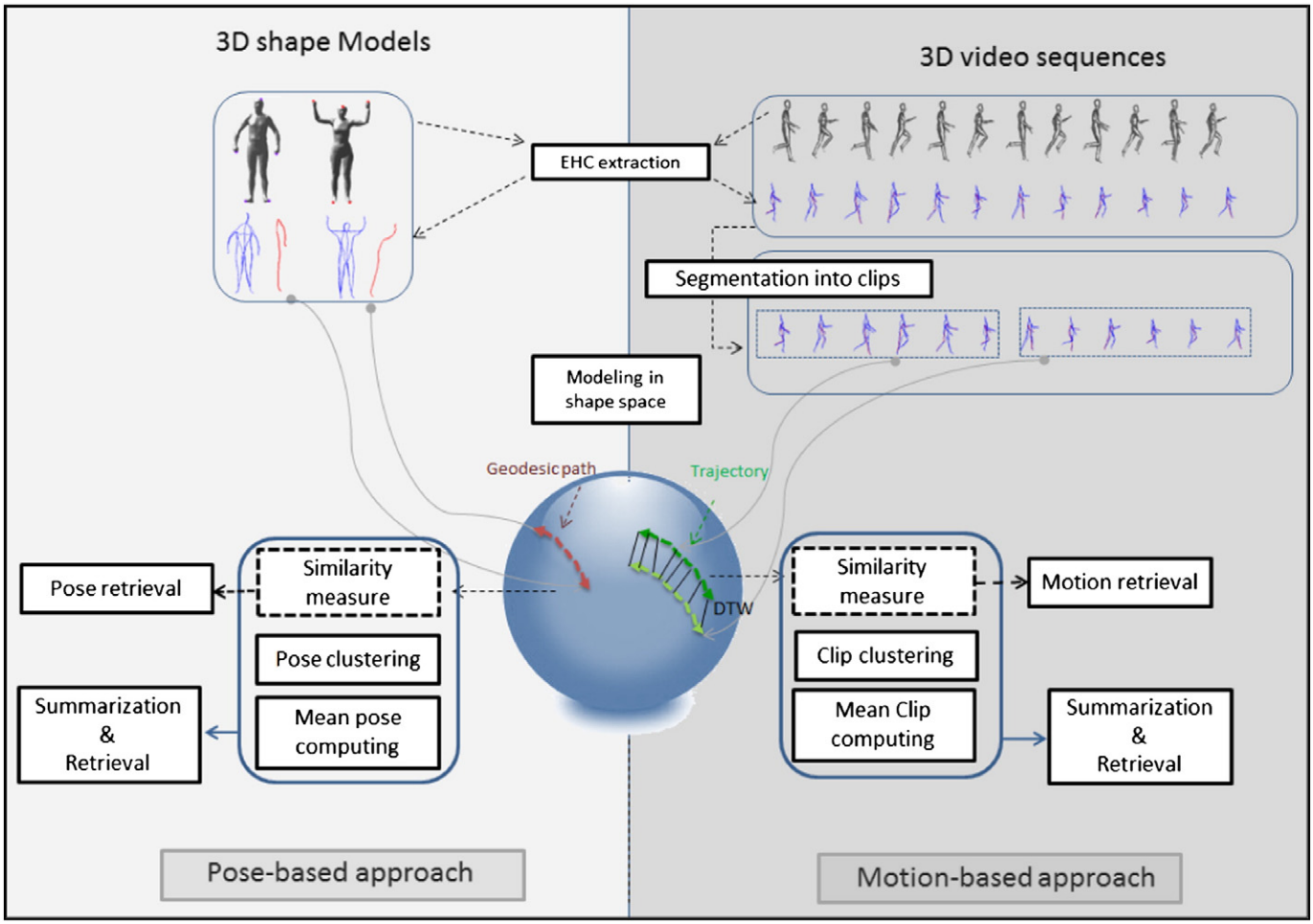

Fig. 1. Overview of 3D human motion framework.

Second, we are interested in the task of video segmentation and comparison between motion segments for video retrieval. As a 3D video of human motion consists of a stream of 3D models, we assume that EHC features are extracted from all 3D shape frames of the sequence, which is further segmented. For direct comparison of video sequences, the motion segmentation can play an important role in the dynamic matching by segmenting automatically the continuous 3D video data into small units describing basic movements, called clips.

For the segmentation of these units, an analysis of minima on motion vector is performed using the metric employed to compare EHC representations. Finally, the motion retrieval is achieved thanks to the dynamic time warping (DTW) algorithm in the feature vector space.

The contributions of this paper are:

- The proposed surface-based shape descriptor called EHC provides a compact representation of the shape. Thereby, reducing both the required space for storage and the time for comparison. As our descriptor is composed of a collection of local human curves, the EHC can find a number of useful applications lying on body part analysis.
- The use of video segmentation allows a semantic analysis of the human motion, within a hierarchical structure of three level "video-clip-pose".
- The modeling of curves in the shape space manifold allows calculating statistics on shape models and motion clips. Thanks to this latter, templates for the pose/clip are computed as average of a collection of poses/clips. The matching with such templates which represents a class, reduces retrieval complexity algorithm from n to $log(n)$.
- The development of a unified framework, viewed as a duality pose/motion, for several processing modules on video retrieval and understanding, where all use the same features and similarity metric.

The outline of this paper is as follows: Section 3 discusses related works in the area of static and temporal shape similarity and video retrieval. The extremal curve extraction is presented in Section 4. Section 5 describes the pose modeling in shape space and the elastic metric used for curve comparison. In Section 6, our approach used for motion segmentation and retrieval is presented. Section 7 describes video clustering and summarization for motion understanding. In Section 8, evaluation of our framework and experimental results for shape similarity, video segmentation and retrieval is performed. Finally, we conclude by a discussion of the limitations of the approach in Section 9 and a summarization of our result issues for future works in the Conclusion section.

2. Related works

3D shape representation and similarity have been under investigation for a long time in various research fields (computer vision, computer graphics, robotics) and for various applications (3D object recognition, classification, retrieval). We address below, the most relevant works related to our approach, which only utilize the full-reconstructed 3D data for shape similarity in 3D human video.

Most works which address this problem evaluate a similarity metric on static shape descriptors based on the surface or on the volume. Others propose to extend the static approaches to temporal shape descriptors.

2.1. Static descriptors

Some of widely used 3D object representation approaches include: spin images, spherical harmonics, shape context and shape distribution.