Pattern Recognition 79 (2018) 1–11

Contents lists available at ScienceDirect

Pattern Recognition

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

A graph-based approach for detecting common actions in motion capture data and videos

Costas Panagiotakis [a,c,*], Konstantinos Papoutsakis [b,c], Antonis Argyros [b,c]

[a] Department of Business Administration, Technological Educational Institute of Crete, Crete, Agios Nikolaos 72100 Greece
[b] Computer Science Department, University of Crete, Crete, Heraklion 70013 Greece
[c] Computational Vision and Robotics Laboratory, Institute of Computer Science, FORTH, Crete, Heraklion 70013 Greece

ARTICLE INFO

Article history:
Received 4 August 2017
Revised 25 January 2018
Accepted 2 February 2018

Keywords:
Common action detection
Video co-segmentation
Temporal action co-segmentation
Dynamic Time Warping

ABSTRACT

We present a novel solution to the problem of detecting common actions in time series of motion capture data and videos. Given two action sequences, our method discovers all pairs of common subsequences, i.e. subsequences that represent the same or similar action. This is achieved in a completely unsupervised manner, i.e., without any prior knowledge of the type of actions, their number and their duration. These common subsequences (commonalities) may be located anywhere in the original sequences, may differ in duration and may be performed under different conditions e.g., by a different actor. The proposed method performs a very efficient graph-based search on the matrix of pairwise distances of frames of the two sequences. This search is supported by an objective function that captures the trade off between the similarity of the common subsequences and their lengths. The proposed method has been evaluated quantitatively on challenging datasets and in comparison to state of the art approaches. The obtained results demonstrate that the proposed method outperforms the state of the art methods both in the quality of the obtained solutions and in computational performance.

1. Introduction

The unsupervised discovery of similar temporal patterns (e.g. similar actions) in time series is considered as an important problem in pattern recognition and computer vision. In this research, we are interested in the detection of common subsequences (commonalities) in two sequences of motion capture data or videos of continuous actions, without any prior knowledge on the type of actions, their number, or their duration. The problem was introduced by Chu et al. [1] as Temporal Commonalities Discovery (TCD), applied to pairs of image sequences containing facial expressions or motion capture data. It has also been tackled in [2] as co-action discovery in multiple image sequences, in [3] as video co-segmentation for action extraction and more recently in [4] as temporal action co-segmentation in pairs of videos. A similar problem appearing in the knowledge discovery and data mining community considers the discovery of multiple common patterns within the same signal [5,6], time series [7,8], or string [9]. In this setting, the discovered commonalities are called motifs [7,8]. This work is also motivated by the task of unsupervised discovery of common human actions in this type of input [10]. Other relevant problems include image co-segmentation [11], image/video co-localization [12] and video co-summarization [13].

The discovery of commonalities in time series is a challenging problem with applications in several domains, including but not limited to data mining and content retrieval, audio and natural language processing, image/video analysis, bio-informatics, economics, physics and more. Both the supervised and the unsupervised versions of the problem are of great importance and interest [14]. For example, the detection of the longest common subsequence has been successfully used for dynamic hand gesture classification [15]. The problem of periodicity detection [16,17] can also be seen as an instance of the problem of finding commonalities/common subsequences between two different videos. In addition, the detected commonalities between video pairs can be used in video retrieval [18,19] which is the task of finding the most similar video based on a query video. Action co-segmentation can also be used to detect recurring combinations of actions without knowledge of what the common events are, how many there are, or when they begin and end. When the action labels of one of the two sequences are given, the detection of commonalities can be used for human action recognition [20–27], which constitutes a central problem in computer vision and pattern recognition with a huge range of potential applications [21,26,27], including the fields of surveil-

* Corresponding author.
E-mail addresses: cpanag@ics.forth.gr (C. Panagiotakis), papoutsa@ics.forth.gr (K. Papoutsakis), argyros@ics.forth.gr (A. Argyros).

https://doi.org/10.1016/j.patcog.2018.02.001
0031-3203/

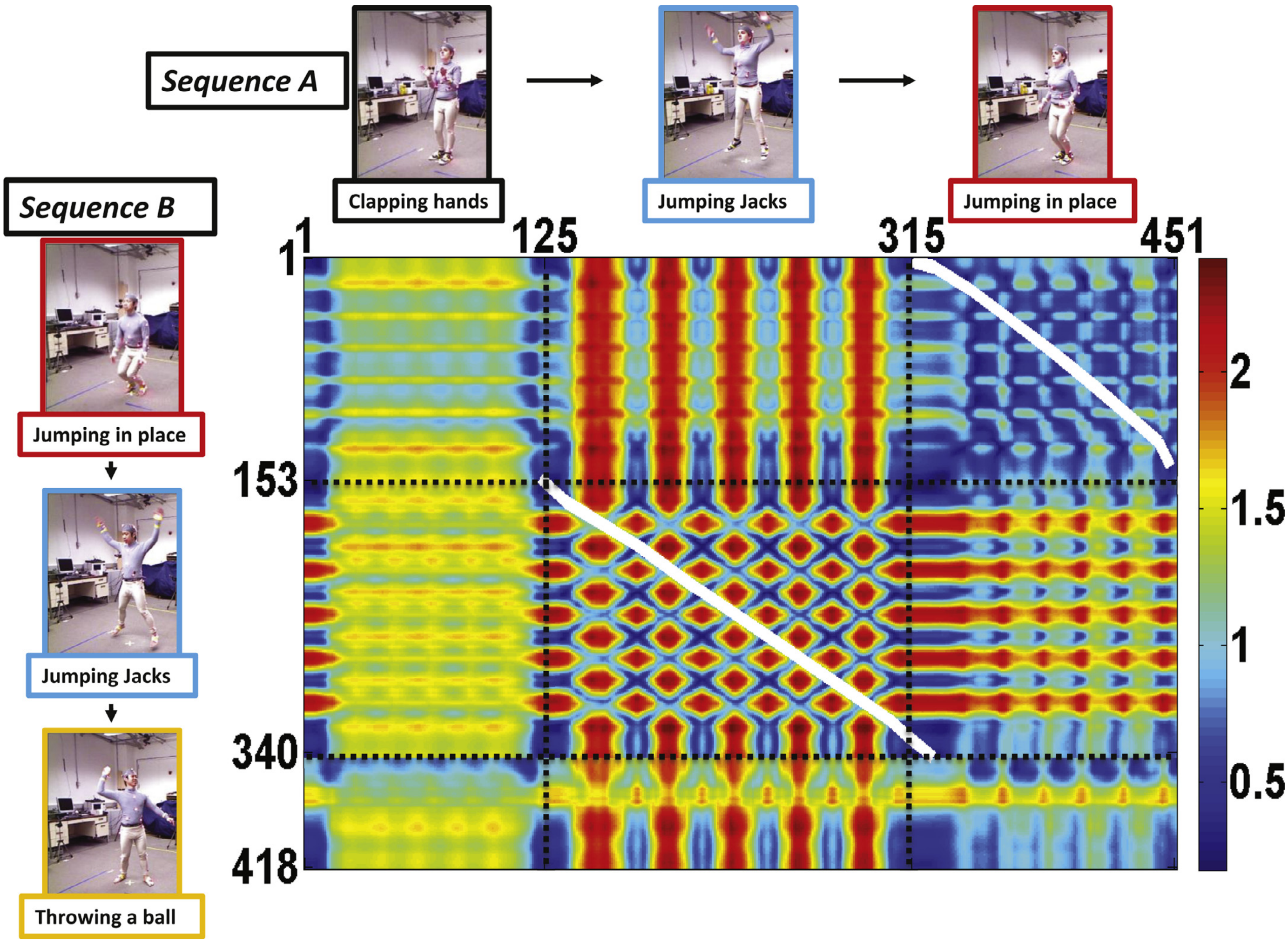

Fig. 1. Two detected commonalities (white curves) projected on the corresponding distance matrix, which was computed based on the pair-wise Euclidean distances between the frames of the image sequences A and B.

lance, advanced human computer interaction, content-based video retrieval, abnormal or suspicious activities detection, health monitoring and athletic performance analysis.

In this work, we consider commonalities between two multi-dimensional time series A and B, representing video or motion capture data. In such a setting, a candidate commonality is a pair of subsequences of A and B which can also be viewed as a path on the distance matrix D of all pairwise distances between the elements (frames) of A and B. Fig. 1 visualizes such a matrix in the form of a heat map, where two commonalities (white curves) are projected. Warm (cold) colors represent large (small) pairwise distances, respectively. The total cost of a path reflects how dissimilar the subsequences of a candidate commonality are. Low (high) cost paths correspond to similar (dissimilar) sub-sequences. A path of small length will tend to have low cost. However, it corresponds to a commonality of short subsequences and is probably not that interesting. As paths increase in length, their cost also increases. Thus, the trade-off between the length of the path (the duration of the commonality) and its cost should be balanced. Detecting multiple commonalities amounts to finding all paths in D that correspond to really common actions in A and B. The lack of supervision in this process has a twofold meaning: (a) no prior model or information on the actions is assumed to be known and (b) the number of commonalities is not assumed to be known a priori.

Given the potential commonality of two subsequences s_A and s_B of two sequences A and B, the corresponding commonality path and its associated cost can be estimated by employing Dynamic Time Warping (DTW) [28]. DTW is a widely-used algorithm for the optimal, non-linear temporal alignment of two sequences and has been extensively used for the alignment of time series [29] like human motion sequences [30] and speech/audio signals [31]. Recently, DTW has been successfully combined with canonical correlation analysis for temporal alignment of multi-modal data, such as acoustic and visual information [32]. The DTW cost is linear to the product of the lengths of the compared sequences, that is, $O(|s_A||s_B|)$. Thus, the naive approach to solve the multiple commonalities discovery problem would be to enumerate all possible paths, evaluate them and keep the best ones. Since the number of comparisons to be performed is $O(|A|^2|B|^2)$, it turns out that this exhaustive scheme has a complexity of $O(|A|^3|B|^3)$. This is prohibitive even for input sequences with a handful of frames. To deal with this problem, the proposed algorithm takes advantage of the properties of the distance matrix D and achieves state of the art performance with a computational complexity of $O(|A||B|)$. This makes possible the discovery of multiple commonalities even for input sequences of many thousands of frames.

In summary, the main contributions of this paper is (a) the formulation of the problem of unsupervised discovery of multiple commonalities in two time series as a search problem on a graph