Image and Vision Computing 54 (2016) 1–11

Contents lists available at ScienceDirect

Image and Vision Computing

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

Randomized time warping for motion recognition ☆

Chendra Hadi Suryanto[a,*], Jing-Hao Xue[b], Kazuhiro Fukui[a]

[a] Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
[b] Department of Statistical Science, University College London, London WC1E 6BT, United Kingdom

ARTICLE INFO

Article history:
Received 8 May 2015
Received in revised form 30 May 2016
Accepted 13 July 2016
Available online 25 July 2016

Keywords:
Feature extraction
Dynamic time warping
Subspace method
Hankel matrix
Motion recognition

ABSTRACT

Dynamic time warping (DTW) has been widely used for the alignment and comparison of two sequential patterns. In DTW, dynamic programming is used to avoid an exhaustive search for the alignment. In this paper, we propose a randomized extension of the DTW concept, termed randomized time warping (RTW), for motion recognition. RTW generates time elastic (TE) features by randomly sampling the sequential data while retaining the temporal information. A set of TE features is represented by a low-dimensional subspace, called the sequence hypothesis (Hypo) subspace, and the similarity between two sequential patterns is defined by the canonical angles between the two corresponding Hypo subspaces. In essence, RTW simultaneously computes multiple degrees of similarities between a number of warped patterns' pair candidates, while in practice, RTW generalizes the Hankel matrix commonly used in modeling of system dynamics. We demonstrate the applicability of RTW through experiments on gesture recognition using three public datasets, namely, the Cambridge gesture database, a subset of the one-shot-learning dataset from the ChaLearn Gesture Challenge, and the KTH action dataset.

1. Introduction

Dynamic time warping (DTW), which is also termed dynamic programming-matching, has been widely used for sequential data analysis. Early uses of DTW range from the comparison of amino acids sequences in bioinformatics [1], through speech recognition [2], to motion analysis [3]. The core idea of DTW is to search for the best alignment of two sequential patterns by optimizing a warping function, which specifies the sequential correspondence between them. Since the number of possible combinations of warped patterns is exponentially large, to avoid exhaustive search dynamic programming has been used, which can effectively optimize the alignment score and produce the alignment path of the most similar warped patterns.

Although DTW is a very useful and widely applicable tool for sequence analysis, it has several limitations when applied to tasks of classifying multiple sequences, such as gesture recognition with many kinds of hand shapes and personal identification by gait recognition. Here are the issues that we will address in this paper.

1. Since dynamic programming is basically a deterministic approach, the obtained solution is likely to be sub-optimal for the sequential data that contains large intra-variation in the temporal structure.
2. The alignment is typically done by trying to match an input sequence to each reference sequence in a given set. This can lead to a high computational cost when the number of the reference sequences to be considered is large.
3. DTW has no internal mechanism to remove or ignore irrelevant variation that may affect the classification result. For example, variation of lighting conditions in video data or speakers in speech data can significantly lower the performance of a classification method using DTW. That is, DTW-based classification methods are sensitive to these undesirable effects.

To tackle these issues, we generalize the notion of DTW to construct a new method for sequential data analysis, which is termed *randomized time warping* (RTW). The core idea of RTW is essentially to simultaneously search for the most similar warped patterns from a number of candidates which are prepared beforehand through randomization. Fig. 1 illustrates the difference between DTW and our RTW approach.

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

* Corresponding author.
E-mail addresses: chendra@cvlab.cs.tsukuba.ac.jp (C. Suryanto), jinghao.xue@ucl.ac.uk (J. Xue), kfukui@cs.tsukuba.ac.jp (K. Fukui).

http://dx.doi.org/10.1016/j.imavis.2016.07.003
0262-8856/

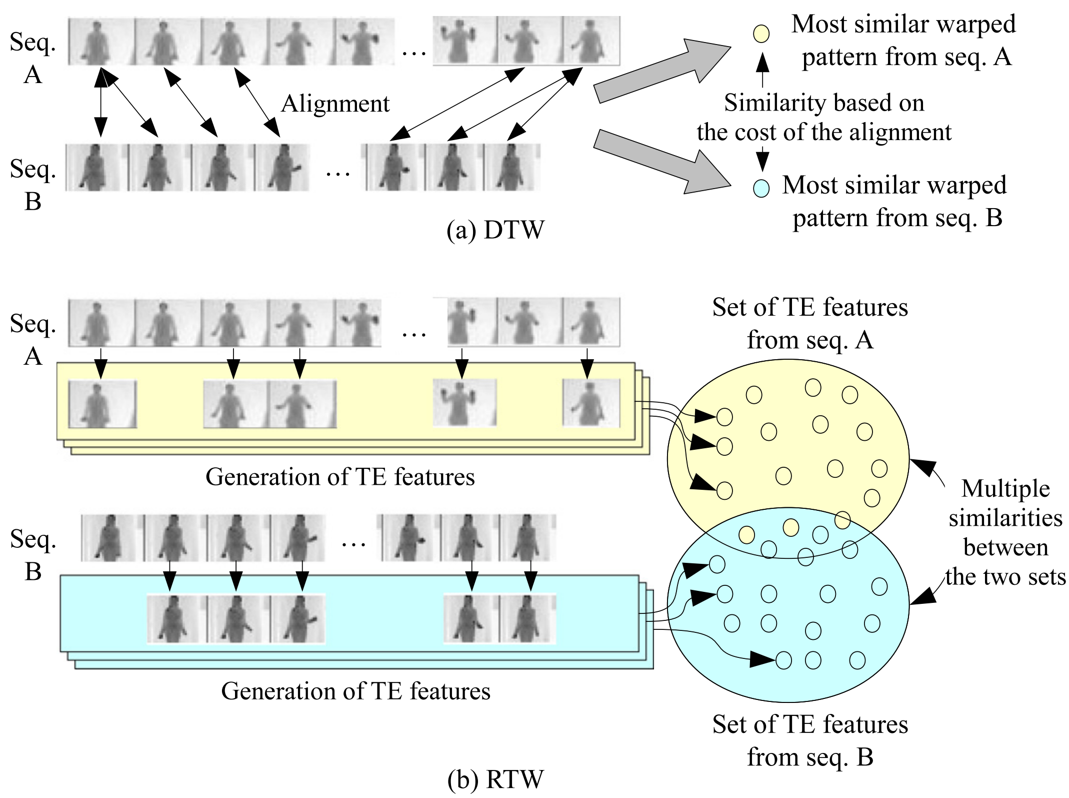

Fig. 1. Comparison between DTW and RTW. (a) DTW searches for the most optimal alignment in a large space through dynamic programming. The outputs of DTW are the most similar warped patterns and the cost of the alignment. (b) RTW generates many candidate warped patterns, called time elastic (TE) features, and then compares the sets of the candidates. The outputs of RTW are multiples of the highest similarities between the two sets.

Instead of searching for the most similar warped patterns using dynamic programming, RTW progressively generates a set of time warped patterns, called *time elastic* (TE) features, through repeated random sub-sampling while preserving the original temporal order. We utilize this bagging-like strategy to ensure that the set of the TE features contains sufficient discriminative frames with high probability. The use of TE features converts the comparison of two sequences to the comparison of two sets of TE features. Fig. 2 shows the comparison process between two sets of the TE features. The comparison is conducted using a subspace-based method, in which each set of TE features is represented as a low-dimensional subspace, called a sequence *hypothesis* (Hypo) subspace. Finally, the similarity between the two sequences is defined by the average of multiple canonical angles θ_i between the two Hypo subspaces. We regard the canonical vectors that form the canonical angles as pseudo-warped patterns (further discussion is provided in Section 3.2). This approach can provide a promising solution to each of the DTW issues previously mentioned:

1. Since random sampling is able to generate a large number of time warped patterns (TE features), our RTW approach is non-deterministic and can deal with a huge number of possible combinations of warped patterns with various time-scales, and thus is able to tackle the issue with large intra-variation in the temporal structure.
2. Since our approach uses the compact subspace-representation, exhaustive matching between all possible TE features is avoided. Each Hypo subspace can contain the TE features from multiple sequences and the canonical angles between two subspaces can be calculated with simple linear algebra. Hence RTW can alleviate the issue of high computational costs.
3. Our approach is based on a subspace method, which can remove or reduce the undesirable effects of irrelevant features. This enables RTW to mitigate the third issue and thus improve the performance of classification

To demonstrate the effectiveness of our approach, we focus on gesture recognition in this paper. We conducted experiments on gesture recognition using three public datasets, namely, the Cambridge

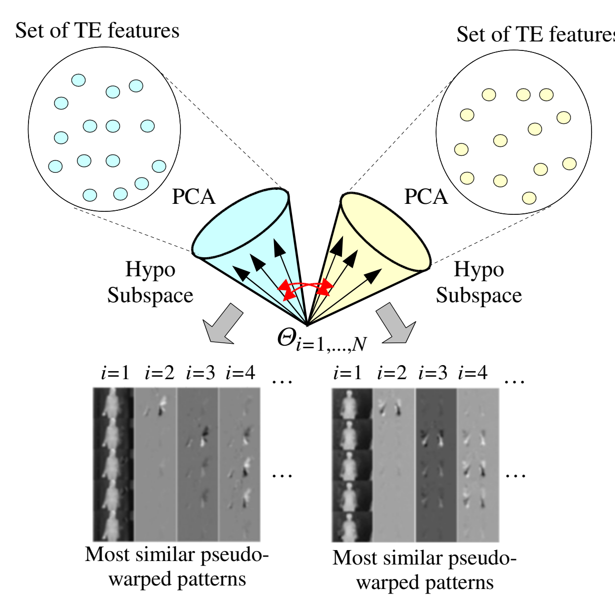

Fig. 2. The comparison process for two sets of TE features in RTW.