Author's Accepted Manuscript

Learning discriminative trajectorylet detector sets for accurate skeleton-based action recognition

Ruizhi Qiao, Lingqiao Liu, Chunhua Shen, Anton van den Hengel

Cite this article as: Ruizhi Qiao, Lingqiao Liu, Chunhua Shen and Anton van den Hengel, Learning discriminative trajectorylet detector sets for accurate skeleton-based action recognition, *Pattern Recognition*, http://dx.doi.org/10.1016/j.patcog.2017.01.015

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Learning discriminative trajectorylet detector sets for accurate skeleton-based action recognition

Ruizhi Qiao, Lingqiao Liu, Chunhua Shen, and Anton van den Hengel

School of Computer Science, University of Adelaide. Adelaide, SA 5005, Australia

Abstract

Devising a representation suitable for characterising human actions on the basis of a sequence of pose estimates generated by an RGBD sensor remains a research challenge. We here provide two insights into this challenge. First, we show that discriminate sequence of poses typically occur over a short time window, and thus we propose a simple-but-effective local descriptor called a trajectorylet to capture the static and kinematic information within this interval. Second, we show that state of the art recognition results can be achieved by encoding each trajectorylet using a discriminative trajectorylet detector set which is selected from a large number of candidate detectors trained through exemplar-SVMs. The action-level representation is obtained by pooling trajectorylet encodings. Evaluating on standard datasets acquired from the Kinect sensor, it is demonstrated that our method obtains superior results over existing approaches under various experimental setups.

Key words: Action recognition, Kinect, Motion capture, Feature learning, exemplar-SVM.

1. Introduction

The advent of low-cost RGB-D sensors, and their ability to rapidly capture sequences of human pose estimates, has promoted a large amount of research interest in skeleton-based human action recognition. [1]. Intuitively, a temporal sequence of 3D skeleton joint locations captures sufficient information to distinguish between actions, but recording such skeleton sequences was previously very expensive with the traditional motion capture technology [2]. Recently, the advent of RGB-D cameras such