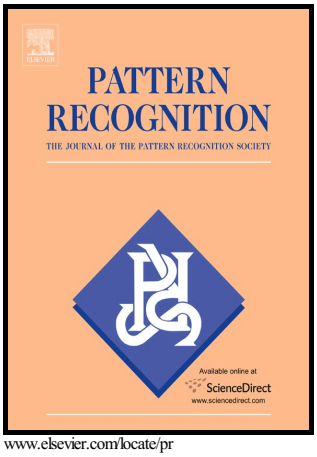
Human Action Recognition with Graph-Based Multiple-Instance Learning

Yang Yi, Maoqing Lin



www.elsevier.com/locate/pr

PII: S0031-3203(15)00442-2
DOI: http://dx.doi.org/10.1016/j.patcog.2015.11.022
Reference: PR5585

To appear in: *Pattern Recognition*

Received date: 26 April 2015
Revised date: 31 August 2015
Accepted date: 28 November 2015

Cite this article as: Yang Yi and Maoqing Lin, Human Action Recognition with Graph-Based Multiple-Instance Learning, *Pattern Recognition*, http://dx.doi.org/10.1016/j.patcog.2015.11.022

Human Action Recognition with Graph-Based Multiple-Instance Learning

Yang Yi [a, b, *], Maoqing Lin [a]

[a] School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China

[b] Xinhua College of Sun Yat-sen University, Guangzhou 510520, China

[*]Corresponding author at: School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China. Tel.: +86 13902295111. E-mail address: issyy@mail.sysu.edu.cn.

***Abstract*:**

A new approach to human action recognition from realistic videos is presented in this paper. First, an affine motion model is utilized to compensate background motion for the purpose of extracting dense foreground trajectories. Then, a trajectory spectral embedding is introduced to split up foreground action into multiple spatio-temporal action parts for constructing a mid-level representation. To deal with over-segmentation, a novel density discontinuity detector is proposed for the sake of generating semantically salient action parts. Finally, to handle the ambiguity in the training set, action classification is formulated within the multiple-instance learning framework, which a spatio-temporal graph model is incorporated into. Extensive experiments show that the proposed approach achieves competitive results to state of the art on UCF Sports, Kisses/Slaps, YouTube, and Hollywood datasets.

***Keywords*:**

Action recognition; Dense trajectory; Motion compensation; Spectral embedding; Multiple-instance learning

1. Introduction

Human action recognition from videos is considered as an important topic in the field of computer vision due to the large number of potential applications in the areas of automatic visual surveillance, advanced human-computer interaction and content-based video retrieval, etc. Recent interest in human action recognition research has shifted from the simple action recognition in a well-controlled scene, to the more realistic action identification under an unconstrained environment which is also referred to as "in the wild" (e.g., videos recorded by an amateur using a hand-held camera, feature films [1, 2], sports broadcast [3, 4] and home videos on YouTube [5]). The action recognition from a complex environment is challenging owing to camera motion,