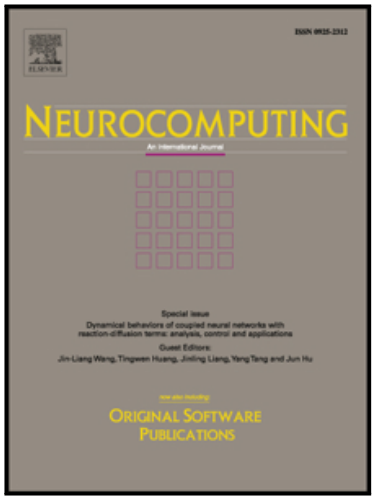
Communicated by Jungong Han

Accepted Manuscript



Region-based Mixture Models for Human Action Recognition in Low-resolution Videos

Ying Zhao, Huijun Di, Jian Zhang, Yao Lu, Feng Lv, Yufang Li

PII: S0925-2312(17)30541-6
DOI: 10.1016/j.neucom.2017.03.033
Reference: NEUCOM 18254

To appear in: *Neurocomputing*

Received date: 19 April 2016
Revised date: 22 February 2017
Accepted date: 19 March 2017

Please cite this article as: Ying Zhao, Huijun Di, Jian Zhang, Yao Lu, Feng Lv, Yufang Li, Region-based Mixture Models for Human Action Recognition in Low-resolution Videos, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.03.033

Region-based Mixture Models for Human Action Recognition in Low-resolution Videos

Ying Zhao[a,b,c], Huijun Di[b], Jian Zhang[c], Yao Lu[b,*], Feng Lv[b], Yufang Li[a]

[a] *Teachers College, Beijing Union University, Beijing, China*
[b] *Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing, China*
[c] *Advanced Analytics Institute, University of Technology, Sydney, Australia*

Abstract

State-of-the-art performance in human action recognition is achieved by the use of dense trajectories which are extracted by optical flow algorithms. However, optical flow algorithms are far from perfect in low-resolution (LR) videos. In addition, the spatial and temporal layout of features is a powerful cue for action discrimination. While, most existing methods encode the layout by previously segmenting body parts which is not feasible in LR videos. Addressing the problems, we adopt the Layered Elastic Motion Tracking (LEMT) method to extract a set of long-term motion trajectories and a long-term common shape from each video sequence, where the extracted trajectories are much denser than those of sparse interest points (SIPs); then we present a hybrid feature representation to integrate both of the shape and motion features; and finally we propose a Region-based Mixture Model (RMM) to be utilized for action classification. The RMM encodes the spatial layout of features without any needs of body parts segmentation. Experimental results show that the approach is effective and, more importantly, the approach is more general for LR recognition tasks.

Keywords: Low-resolution, action recognition, elastic motion tracking, mixture model, Expectation Maximization (EM) algorithm

*Corresponding author. Present address: 5 South Zhongguancun Street, Haidian District, Beijing 100081, China.
Email address: vis_yl@bit.edu.cn (Yao Lu)