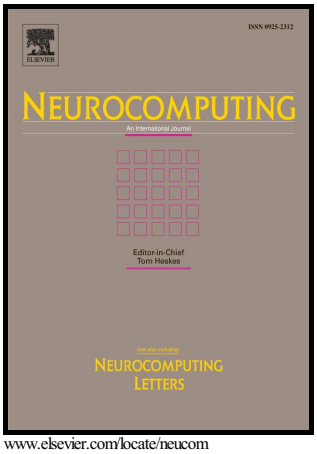

www.elsevier.com/locate/neucom

A Novel Local Feature Descriptor Based on Energy Information for Human Activity Recognition

Yongxiong Wang, Yubo Shi, Guoliang Wei

PII: S0925-2312(16)31267-X
DOI: http://dx.doi.org/10.1016/j.neucom.2016.07.058
Reference: NEUCOM17664

To appear in: *Neurocomputing*

Received date: 21 December 2015
Revised date: 23 June 2016
Accepted date: 14 July 2016

Cite this article as: Yongxiong Wang, Yubo Shi and Guoliang Wei, A Novel Local Feature Descriptor Based on Energy Information for Human Activity Recognition, *Neurocomputing,* http://dx.doi.org/10.1016/j.neucom.2016.07.058

A Novel Local Feature Descriptor Based on Energy Information for Human Activity Recognition

Yongxiong Wang*, Yubo Shi, and Guoliang Wei

Shanghai Key Lab of Modern Optical System and Engineering Research Center of Optical Instrument and System, Ministry of Education, University of Shanghai for Science and Technology, 200093, China

Abstract In this paper we propose a novel local feature descriptor based on energy information for human activity recognition. Instead of detecting spatio-temporal interest points, we combine the kinetic energy, gesture potential energy of 3D skeleton joints and others as a feature matrix. The semantic features are obtained by the Bag of Word (BOW) based on k-means clustering. These features conform to not only kinematics and biology of human action, but also the natural visual saliency for action recognition. During the activity recognition, we first present a temporal segmentation method based on kinetic features of human skeleton to cut the long videos into the sub-action segments. Then the sub-action units are iteratively incorporated in the meaningful groups by considering similarity of feature information. Finally, SVM based on kernel function is used to carry out human activity recognition. The experimental results show that our approach outperforms several state-of-the-art algorithms based on our proposed low dimensional features of energy information.

***Keywords*:** Activity recognition; 3D skeleton; Local feature descriptor; Bag of word; SVM

1 Introduction

Human activity recognition plays an extremely important role in many applications, such as in intelligent furniture, service robot, and intelligent surveillance. In recent decades, human activity recognition based on vision has become one of the most popular research topics and made great progress. Especially, there are many works of human behavior recognition based on 2D video [1, 2, 3]. However, due to the limitations of 2D video, light changes, occlusion, background changes and other factors often greatly degenerate recognition accuracy in the actual application. Because of the affordable depth image devices, image segmentation and object recognition become more simple and robust, therefore many researchers have pay more attention to it in recent years. We can conveniently and easily acquire color image and depth information (RGB-D) from a 3D sensor device, such as Microsoft Kinect and ASUS Xtion Pro Live. And human skeleton can be precisely extracted from this RGB-D data [4].

Feature description or feature selection of human behavior is a key problem in