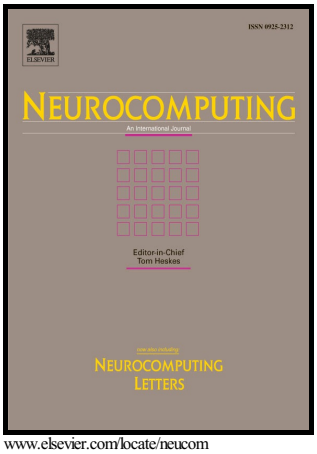
Author's Accepted Manuscript



www.elsevier.com/locate/neucom

Spatial and Temporal Scoring for Egocentric Video Summarization

Zhao Guo, Lianli Gao, Xiantong Zhen, Fuhao Zou, Fumin Shen, Kai Zheng

PII: S0925-2312(16)30480-5
DOI: http://dx.doi.org/10.1016/j.neucom.2016.03.083
Reference: NEUCOM17105

To appear in: *Neurocomputing*

Received date: 1 November 2015
Revised date: 27 February 2016
Accepted date: 18 March 2016

Cite this article as: Zhao Guo, Lianli Gao, Xiantong Zhen, Fuhao Zou, Fumi Shen and Kai Zheng, Spatial and Temporal Scoring for Egocentric Videc S u m m a r i z a t i o n , *Neurocomputing* http://dx.doi.org/10.1016/j.neucom.2016.03.083

This is a PDF file of an unedited manuscript that has been accepted fo publication. As a service to our customers we are providing this early version o the manuscript. The manuscript will undergo copyediting, typesetting, an 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 whic could affect the content, and all legal disclaimers that apply to the journal pertain

ACCEPTED MANUSCRIPT

Spatial and Temporal Scoring for Egocentric Video Summarization

Zhao Guo[a], Lianli Gao[a,*], Xiantong Zhen[b], Fuhao Zou[c], Fumin Shen[a], Kai Zheng[d]

[a] *University of Electronic Science and Technology of China, China*
[b] *University of Western Ontario, Canada*
[c] *Huazhong University of Science and Technology, China*
[d] *Soochow University, China*

Abstract

We present a summarization approach for egocentric video. Given hours of video, the proposed method produces a compact storyboard summary of the camera wearer's day. In contrast to traditional keyframe selection techniques, the resulting summary focuses on the most important video shots which reflect high stable salience, discrimination and representativeness. To accomplish this, we utilize egocentric salience cues, motion cues and a selection model to capture stable salience weight, discriminative weight and representative weight of a video shot respectively. We further combine these weights in a unified framework to predict the importance score of a shot, based on which, important shots are selected for the storyboard. Critically, the approach is neither camera-wearer-specific nor object-specific; that means the learned importance metric need not be trained for a given user or context, and it can predict the importance of shots that have never been seen previously. Experimental results on three video datasets across various genres demonstrate that our proposed approach clearly outperforms several state-of-the-art methods.

Keywords: egocentric videos, video summarization, sparse coding

1. Introduction

With the widespread availability of wearable cameras, massive amounts of videos are recorded every day to capture special moments or keep track of daily activities. It has been never easier to create and access a large amount of video content. There has been an ever increasing demand for efficient and effective tools to help users consume such a large amount of disorganized video content. Video summarization is one of the prime techniques that respond to this need, which produces a condensed and

*Corresponding author
Email addresses: lianli.gao@uestc.edu.cn (Lianli Gao), zhenxt@gmail.com (Xiantong Zhen), fuhao_zou@hust.edu.cn (Fuhao Zou), fshen@uestc.edu.cn (Fumin Shen), zhengkai@suda.edu.cn (Kai Zheng)

Accepted manuscript

Download English Version:

https://daneshyari.com/en/article/4948627

Download Persian Version:

https://daneshyari.com/article/4948627

Daneshyari.com