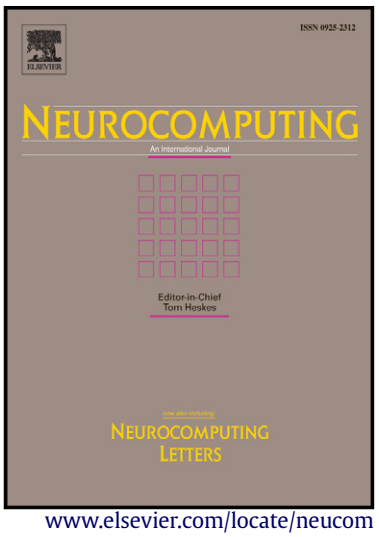
Author's Accepted Manuscript



www.elsevier.com/locate/neucom

Structured sparse representation for human action recognition

F. Moayedi, Z. Azimifar, R. Boostani

PII: S0925-2312(15)00214-3
DOI: http://dx.doi.org/10.1016/j.neucom.2014.10.089
Reference: NEUCOM15181

To appear in: *Neurocomputing*

Received date: 10 July 2014
Revised date: 26 October 2014
Accepted date: 27 October 2014

Cite this article as: F. Moayedi, Z. Azimifar, R. Boostani, Structured sparse representation for human action recognition, *Neurocomputing*, http://dx.doi.org/10.1016/j.neucom.2014.10.089

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.

ACCEPTED MANUSCRIPT

Structured Sparse Representation for Human Action Recognition

F. Moayedi[*,**], Z. Azimifar[*], R. Boostani[*]

[*] *School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran*
[**] *Visiting Researcher of Medical Management and Information School, Shiraz University of Medical Science, Shiraz, Iran*
{*moayyedi,azimifar,boostani*}*@ cse.shirazu.ac.ir*

Accepted manuscript

Abstract

Video understanding is an important goal of several computer vision problems. To achieve this goal, a video is decomposed into a set of key components and the interactions between the components are modeled. Human action recognition is a challenging example of video understanding in the field of computer vision. Modeling a vocabulary of local image features in a bag of visual words (BoW) is a common approach to extract the components of an action video. Since in a video recognition task, there is no direct mapping from the raw features to class label, higher level visual descriptors and indeed, more accurate dictionaries are required. Therefore, in order to extract intrinsic shape bases and to consider temporal structure of an action, in this paper we take the advantages of group sparse coding methods. In our proposed BoW method each video is represented as a histogram of the coefficients obtained from group sparse coding. The main contribution of this study is to explore the geometry of action components via structured sparse coefficients of visual words in a real-time manner. In comparison with the conventional BoW models, our proposed approach has other advantages including much less quantization error, higher level feature representation which leads to reduction in model parameters and memory complexity while considering temporal structure. We evaluate our method on standard human action datasets including KTH, Weismann, UCF-sports and UCF50 human action datasets. The experimental results

Download English Version:

https://daneshyari.com/en/article/6865788

Download Persian Version:

https://daneshyari.com/article/6865788

Daneshyari.com