Signal Processing 124 (2016) 36–44

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Multi-task human action recognition via exploring super-category

Yanhua Yang, Ruishan Liu, Cheng Deng, Xinbo Gao *

School of Electronic Engineering, Xidian University, Xi'an 710071, China

ARTICLE INFO

Article history:
Received 16 July 2015
Received in revised form
13 October 2015
Accepted 30 October 2015
Available online 22 November 2015

Keywords:
Action recognition
Super-category
Multi-task learning
Mutual information
Fisher Vector

ABSTRACT

There is indeed a relationship among various action categories, with which many correlated action categories can be clustered into a same group, named *super-category*. Knowledge sharing within super-category is an effective strategy to achieve good generalization performance. In this paper, we propose a novel human action recognition method based on multi-task learning framework with super-category. We employ Fisher vector as the action representation by concatenating the gradients of log likelihood with respect to mean vector and covariance parameters of Gaussion Mixture Model. Considering the occupancy probability of each Gaussian component is different, we naturally discover the relationship among different action categories by evaluating the importance of each Gaussian component in classifying each category. For these categories, the more related to the same Gaussian component, the more possible belonging to the same super-category, and vice versa. By applying the explored super-category information as a prior, feature sharing within super-category and feature competition between super-categories are simultaneously encouraged in multi-task learning framework. Experimental results on large and realistic datasets HMDB51 and UCF50 show that the proposed method achieves higher accuracy with less dimensions of features over several state-of-the-art approaches.

1. Introduction

Recognizing human action [1,2] has been an important research area in computer vision while enabling very broad applications, e.g., smart surveillance, video analysis, human–computer interaction and so forth. Earlier research mainly focused on datasets comprised of simple actions recorded in a controlled environment such as Weizmann dataset [3]. However, there is still a great challenge to recognize actions from large and more realistic datasets (e.g., UCF50 dataset [4], HMDB51 dataset [5]) due to various difficulties, such as background clutter, viewpoint changes, occlusion and motion style variation.

Most of the recent work show that, basing on more elaborately designed spatial-temporal local features [6–8], the performance could be improved significantly by Bag-of-Visual-Words (BoVW) model [9] or its variants [10,11]. Among its variants, Fisher Vector [11] which could describe higher order statistic of features in conjunction with Gaussian Mixture Model has achieved promising performance in several action recognition datasets [12–14]. Fisher Vector is the concatenation of gradients of log likelihood with respect to mean vector, and covariance matrix of each Gaussian component. As a popular clustering algorithm, each component of GMM is considered as a cluster center. From this perspective, each Gaussian component can be understood to have the ability to reflect specific feature characteristics, such as one of a kind of motion patterns for motion features. As the occupancy probability that each sample data is generated by different Gaussian component is dissimilar, we assume that the importance of each

* Corresponding author.
E-mail address: xbgao@mail.xidian.edu.cn (X. Gao).

http://dx.doi.org/10.1016/j.sigpro.2015.10.035

Gaussian component to each action category is dissimilar and the correlated action categories prefer to have a closer relationship with the same Gaussian components. For example, as shown in Fig. 1, action shoot gun and shoot ball are highly correlated and prone to share similar information captured by the same Gaussian component. In this paper, the above mentioned relationship between Gaussian components and action categories is explored as an important cue to reflect the relationship between various action categories.

To exploit the knowledge shared among categories, multi-task learning framework [15,16] is always employed, which could share useful information by learning a common representation space for all tasks. The key to improve the generalization performance with MTL is that those tasks are indeed related. However most existing multi-task learning methods [15–17] are assumed that all of the tasks are correlated, which may be too strong in realistic setting. When the assumption does not hold, learning jointly with dissimilar or outlier tasks will lead to negative transfer occurs. As a result, the performance is deteriorated. Thus the crucial issue in applying MTL is how to construct reasonable task correlation to guide the knowledge sharing. Fortunately, we observe that action categories with high correlation could be clustered into a same group based on Fisher Vector and categories in the same group are prone to share high-dimension feature from the same Gaussian and vice versa. A part example of category groups in the HMDB51 dataset is shown in Fig. 1.

Based on the above observations, this paper proposes a new approach to measure the similarity of categories by computing mutual information. Concretely, the mutual information is calculated firstly and then the important score of a Gaussian component to each category is estimated based on mutual information. The Gaussian component with a low important score is discarded and vice

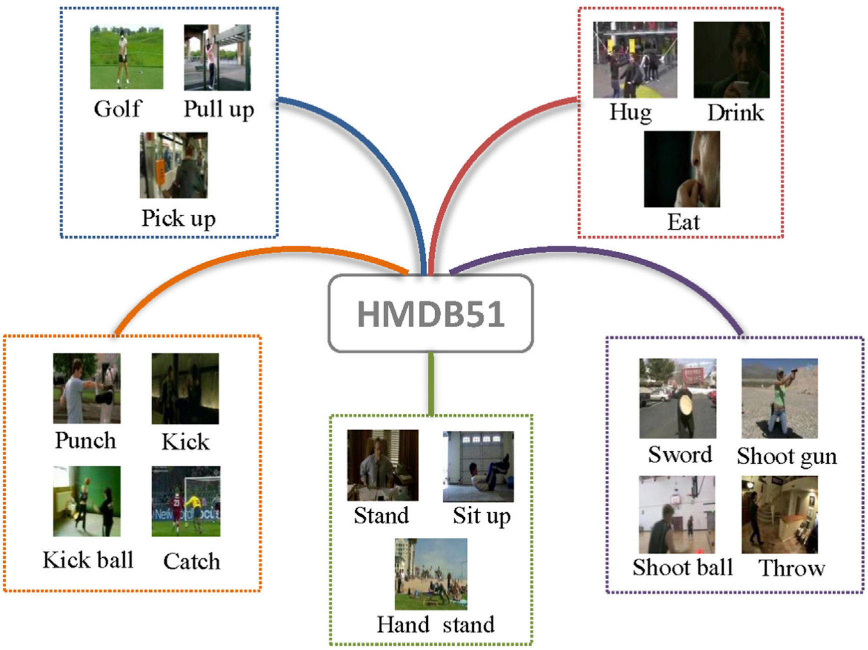

Fig. 1. Part example of category groups in the HMDB51 dataset obtained by the proposed work.

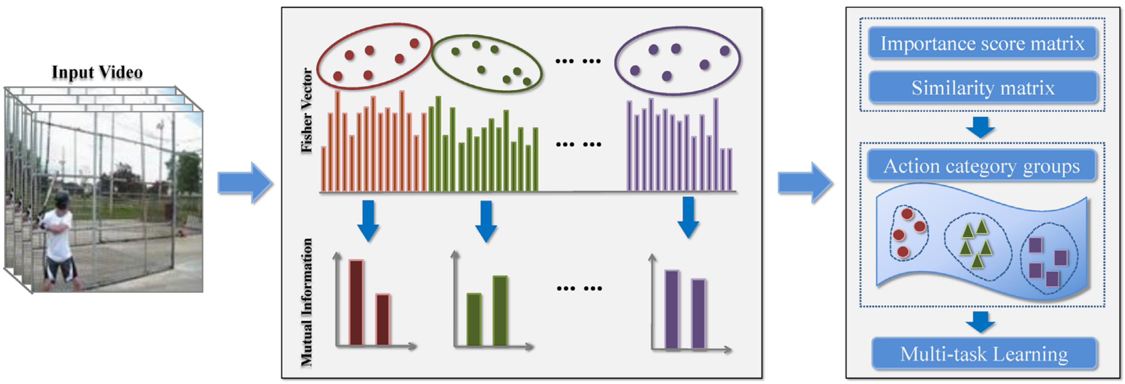

Fig. 2. Learning procedure of the proposed method.