

Human Action Recognition in Still Images using Bag of Latent Poselets

Moin Nabi, Mohammad Rahmati¹

¹{m.nabi, m.rahmati}@aut.ac.ir IPPR Lab, Amirkabir University of Technology, Tehran, Iran.

Abstract

In this paper we proposed a new method for the problem of structural human action recognition in single images. In this work, we first extract all Poselets in the images for using as the descriptor of human's activity. Then, we model the latent topics of human poses by using extracted vectors and P-LSA. Finally recognize human's action in a query image by using the trained SVM on the extracted bag of latent Poselets. We tested our method on PASCAL VOC2010 action classification dataset and the results show the significant improvements in some action classes such as Walking and Running.

Keywords: Latent pose, Poselet, Topic modelling, P-LSA, SVM.

1 Introduction

Most of the works in the field of human action recognition are focused on video, and recognizing action from still images, on the other hand, has not been widely studied. But we believe this problem is important, not only to be applied to videos, but also for being used directly for several new applications like image search and retrieval on Internet [1]. The rest of the paper is structured as follows. After a brief explanation of the method with a comprehensive workflow of it (Sec.2), some qualitative results are shown and the method is compared with one of the recent papers on same problem.

2 Proposed Method

In this work, we first train Poselets [2] on the set of all images without paying attention to their action labels. Then prune some of these Poselets that are none-discriminative and find the action specific Poselets using the ad-hoc method by Maji et al. [3]. These extracted Poselets can lead us to build a Poselet Activation Vector (PAV) for every image. PAV is an efficient descriptor for images based on the active Poselets inside it.

We then concatenate these PAVs to make a Poselet by Image matrix. This matrix is similar to the matrix of word by document in the text space; so we use text-based approaches for extracting the basic concepts of human pose (Latent Poselets). For this purpose Topic Modelling approaches are efficient and P-LSA method is what we used for this task.

Finally, we extract the introduced descriptor of *bag of latent Poselet* for all training images and train a linear SVM using these feature vectors. This trained model is what we used for recognizing the action label of human inside a new query image in classification phase.

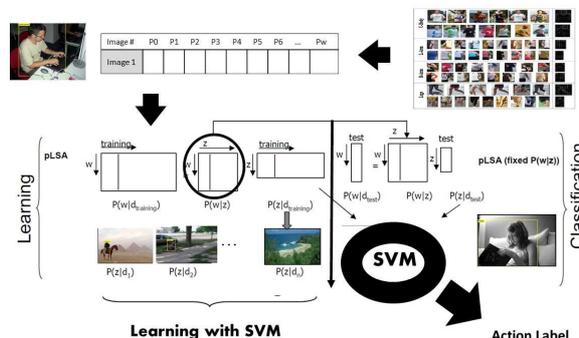


Figure 1 : Workflow of the proposed method

#	Category	PAV[3]	Our Method
1	Phoning	63.3	67.3
2	Play instruments	44.2	38.9
3	Reading	37.4	43.3
4	Riding bike	62	64.4
5	Riding horse	91.1	86.8
6	Running	82.4	88.5
7	Taking photo	21.1	20.2
8	Using computer	54.2	44.3
9	Walking	82	84.3
	Ave.	59.7	58.6

Table 1: Average Precision



Figure 2 : Some qualitative results

3 Results

This work was tested on recently introduced PASCAL VOC 2010 dataset and our method achieves an average AP of 58.6 on the validation set which is comparable to the winning techniques in PASCAL VOC 2010.

Acknowledgements

This research was done as a part of my master thesis in IPPR laboratory at CEIT department of Amirkabir University of Technology.

References

- [1] W. Yang, Y. Wang, and G. Mori. Recognizing human actions from still images. CVPR 2010.
- [2] L. Bourdev and J. Malik. Poselets: Body part detectors training using 3d human pose annotations. ICCV 2009.
- [3] S. Maji, L. Bourdev and J. Malik, Action Recognition from a Distributed Representation of Pose and Appearance, CVPR 2011.