Recognizing human actions in still images: a study of bag-of-features and part-based representations

Vincent Delaitre, Ivan Laptev and Josef Sivic

July 2, 2010

1. Introducing a new dataset

2. Bag-of-features classifier
    - Image representation
    - Investigated Kernels
    - Dealing with bounding boxes

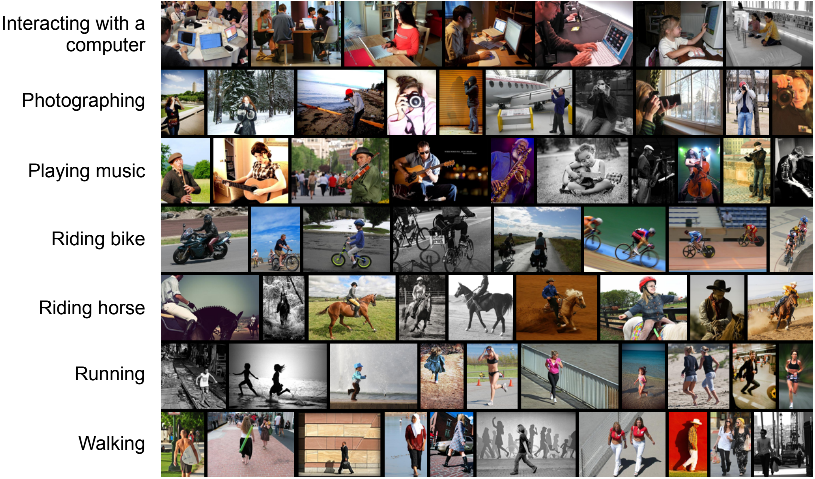
3. Discriminatively trained part-based model

4. Experimental results

Introducing a new dataset

We collected a new challenging dataset for real-life human actions. It is composed of 968 images collected from Flickr representing natural variations in terms of camera view-point, human pose, clothing, occlusions and scene background.
Pictures are distributed among 7 different classes:

- Interacting with a computer
- Taking a photograph
- Playing music
- Riding bike
- Riding horse
- Running
- Walking


Interacting with a computer
Photographing
Playing music
Riding bike
Riding horse
Running
Walking

Classification task

Each person is annoted with a bounding box (smallest rectangle containing its visible pixels) and the action being executed.

In the following, we are interessted in the 7-class classification problem. The training set consists in 70 images of each type of action, so that at least 48 images per class remain for test.

We mesure the performances using:

i *the classification accuracy*: average of the diagonal of the confusion table

ii *the mean average precision (mAP)*: mean area under the precision-recall curve of each 1-vs-all classifiers.