The importance of high level representation in Action Recognition

Advisors:
Francesc SERRATOSA
francesc.serratosa@urv.cat
http://deim.urv.cat/~francesc.serratosa/

Mario VENTO
mvento@unisa.it
http://docenti.unisa.it/005501/en/home

Application
Action recognition can be profitably used in several application fields, ranging from surveillance systems and ambient assisted living, with the possibility of identifying some events of interest occurring into a video (an aggression, a robbery, a tumble of an elder person, etc), to video retrieval, by allowing a human operator to easily retrieve specific types of actions performed by persons in a video footage.

The recognition of an action is typically performed at two different layers: at a low level, a set of features (local or global) is analysed frame by frame; at a high level, the temporal information is explored for taking the final decision regarding the action to be recognized.

The aim of this thesis is to realize a framework for experimentally evaluating if and how much the different high level representations impact in the accuracy of action recognition methodologies.

Figura 1 - Example of actions
Methodology
Some standard low level representation will be defined (for instance, LBP, Hu moment and SIFT+BOW) and will be used for testing the system. Then, the different high level representations used in the literature will be compared so as to understand the importance and the effect of the different representations on the accuracy of the system. In particular, the following representations need to be considered: bag of words [3], VLAD, temporal bag of words, n-grams, strings [1,2,4].

Implementation and Experimentation
The methods will be implemented in C/C++, using Opencv Libraries. The experimentation will be performed by using standard datasets, such as MHAD, MIVIA Action dataset, both providing the background image.

References