Efficient Activity Detection with Max-Subgraph Search

Chao-Yeh Chen and Kristen Grauman
The University of Texas at Austin

Problem
How to detect human activity in continuous video?

Status quo approaches:
- Expensive sliding window search
- Restricted shapes (gobo detection)
- Lack context (tracking humans)

Main Idea
We pose activity detection as a maximum-weight connected subgraph problem over a learned space-time graph.

Approach
Define weighted nodes
- Frame-level nodes
- Space-time nodes

Search for the max-weight subgraph
- Transform max-weight subgraph problem into a prize-collecting Steiner tree problem.
- Solve efficiently with branch and cut method from [Ljubic et al. 2006].

Results
UCF-concatenated
Dense + HoG3D
STIP + HoG/HoF or high-level
STIP + HoG/HoF

Temporal detection
UCF concatenated
Hollywood uncropped
MSR Action

Space-time detection
Possible subvolume types
- T-Sliding
- ST-Cube
- ST-Cube-Subvolume

Classifier training for feature weights
Classifier must be defined such that features computed within local space-time region S can be combined additively. Suitable classifiers include boosting, Naïve Bayes, or linear SVM:

Conclusions
- Reduced computation time for detection vs. sliding window search.
- Flexible node structure offers more robust detection in noisy backgrounds.
- High-level descriptor shows promise for complex activities by incorporating semantic relationships between humans and objects in video.