Actom Sequence Models for Efficient Action Detection

LEAR – INRIA Grenoble

Adrien Gaidon

Zaid Harchaoui

Cordelia Schmid

Presentation by Benoit Massé

Introduction

- Video : Big Data
- Automatisation ?
  - Semantic analysis
  - Retrieval

Problem :

Find **if** and **when** a specific action happen

State of the art

- Training
  - Define the action
  - Choose the features
  - Train
- Retrieval
  - Classification
  - Detection

State of the art

- Training
  - Define the action => Spatio-temporal extent
  - Choose the features => HoG, HoF, SP interest Point
  - Train => Bag-of-Feature
- Retrieval
  - Classification => SVM, Bayesian Network
  - Detection => ?

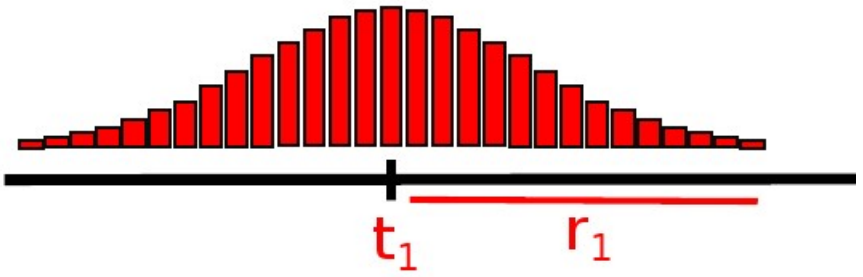
Actoms

- **Actom :** short atomic action

Actoms

An actom has

- A location t
- A radius r



Actom descriptors : Set of visual words

- Bag of Features applied on HoG, HoF, Harris Interest points...
- Ponderated sum from t - r to t + r

Interest of Actoms

- An **action** is composed of several **actoms**
  - New goal : find an ordered sequence of actoms
  - No temporal dependance inside an action
    - Gap between actoms
    - Overlap

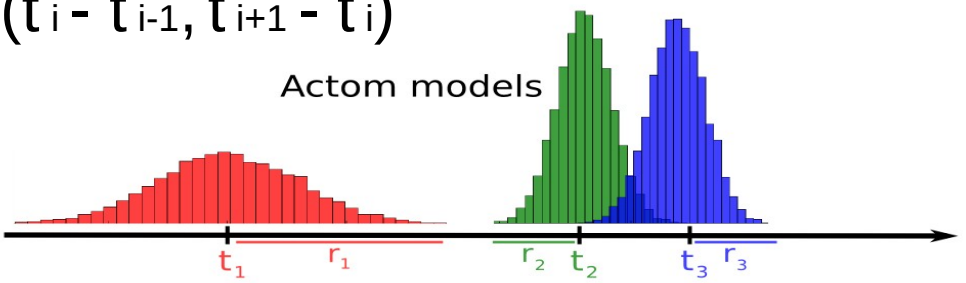
- An action can be composed of very different parts

  => Classic methods compute the average

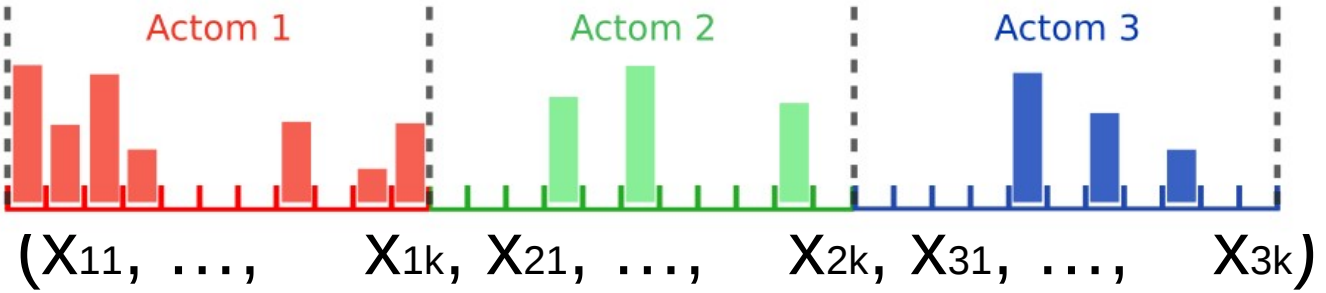
Actom Sequence Model (ASM)

One Action = One Actom Sequence

- The radius r_i of actom i depends on its distance to the closer other actoms : $\min(t_i - t_{i-1}, t_{i+1} - t_i)$



- ASM : concatenation of actoms words



Classification

- Given a new ASM (x_{11}, ... x_{nk}), does it corresponds to the trained action ? (for instance : « drinking »)
  - Classic machine learning problem
  - Chosen solution : SVM
  - Including negative examples improves the classifier

Detection

- Given a video, find all the occurences of the trained action. (for instance : « drinking »)

For every 5 frames

Set the current frame as the middle actom

Generate candidates for other actoms

Apply classification on the result

End

Delete non-maximal overlapping actions

Detection

Tricky step : Generating the other actoms

We must estimate the distance between actoms

- Training : Build the multivariate distribution $\{t_{i+1} - t_i\}$

  Remove the outliers

- Estimation : Try all the possible combinations

  (starting from the middle limit the error propagation)

Experiments

4 kind of actions

- Drinking
- Smoking
- Open a door
- Sit down

Criteria

- OV20 (20 % Overlap)
- OVAA (All Actoms Overlap)

State of the art Comparison

- Bag of Features
- Bag of Features with a grid
- Other published methods

Results

Method	"Drinking"	"Smoking"
matching criterion: OV20		
DLSBP [3]	40	NA
LP [12]	49	NA
KMSZ [9]	54.1	24.5
BOF	36 (±1)	19 (±1)
BOF T3	44 (±2)	23 (±3)
ASM	**57** (±3)	**31** (±2)
matching criterion: OVAA		
BOF	11 (±2)	1 (±0)
BOF T3	18 (±3)	4 (±1)
ASM	**50** (±5)	**22** (±2)

Method	"Open Door"	"Sit Down"
matching criterion: OV20		
DLSBP [3]	13.9	14.4
BOF	12.2	14.2
BOF T3	11.5	17.7
ASM	**16.4**	**19.8**
matching criterion: OVAA		
BOF	9.9	5.8
BOF T3	5.1	13.1
ASM	**14.9**	**16.7**

Conclusion

ASM gives better result than state-of-the-art, using the same data sets.

=> Actoms are particularly adapted for representing the temporal structure of actions into videos

QUESTIONS ?