

# Human Activity Recognition

Niranjan

Under the guidance from

Dr. Subhashis Banerjee and Dr. Parag Singla

Department of Computer Science, IIT Delhi

M.Tech Project Presentation – September 20, 2013

# Problem Statement

- What is Video Activity Recognition?



Play

# Approach

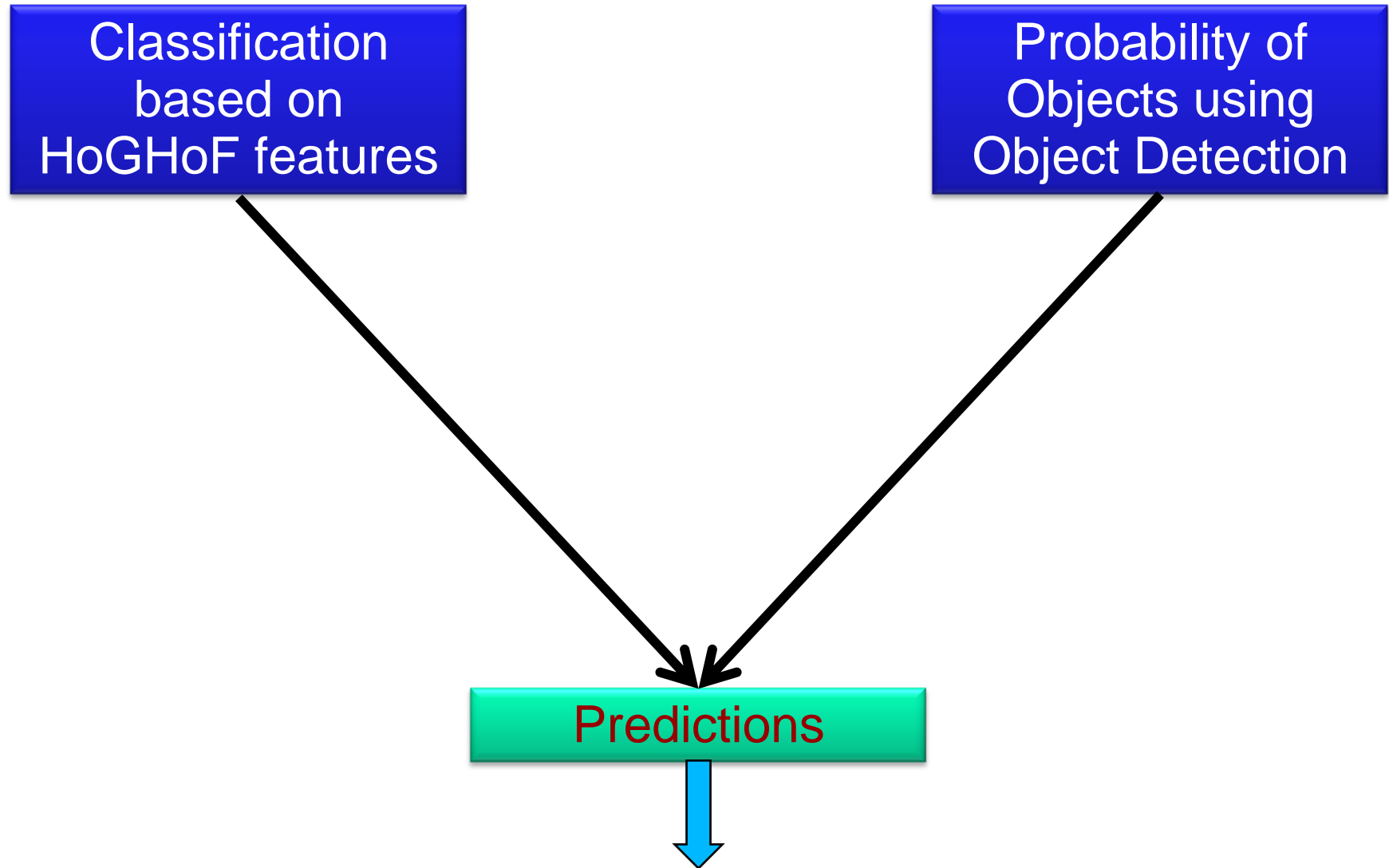
## *Actions in Context*

by Ivan Laptev et. al. in Proc. CVPR'09

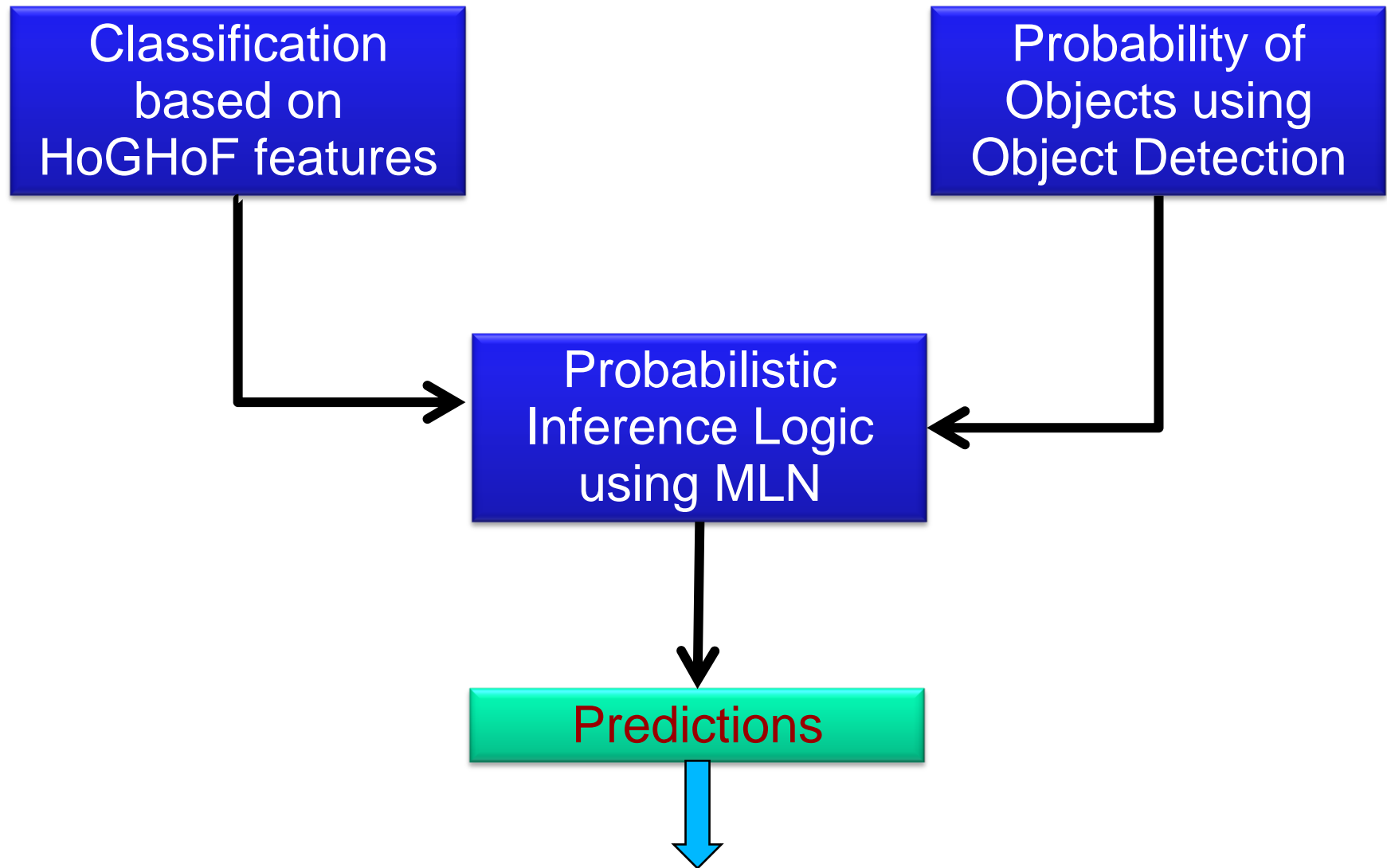
## *Improving Video Activity Recognition using Object Recognition and Text Mining*

by Tanvi Motwani and Raymond J. Mooney,  
ECAI-2012

# Approach

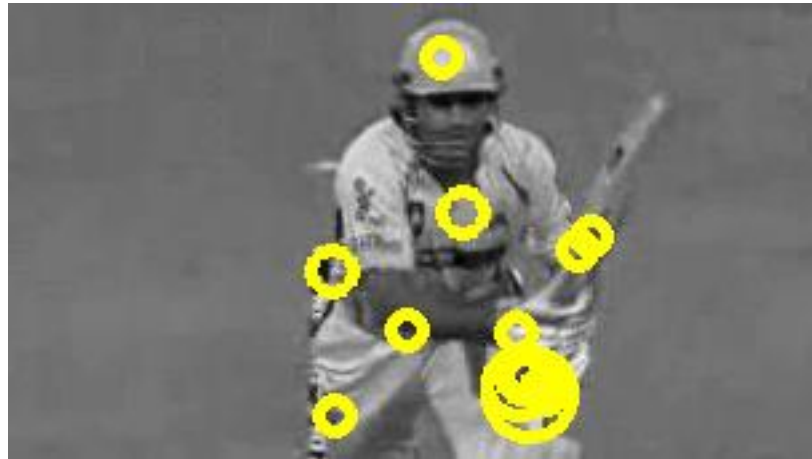


# Approach++



# Classification using HoGHoF

- Extracting STIP features



STIP HoG and HoF feature vector :

0.627496	0.0892087	0.0293946	0.253901	0.668772
0.160494	0.000758835	0.169975	0.414533	0.508073 ...

.....

.....

# Classification using HoGHoF

- ❑ Describe a video clip as bag-of-features.
  - Cluster all HoGHoF feature descriptors using k-means.
  - Represent clip as a histogram over these clusters
- ❑ Train a classifier
  - ❑ Supervised - Dataset is pre labeled

# Object Detection

- Using Discriminatively Trained Deformable Part Models
  - Pre-trained object detector for 19 objects



# Object Detection



# Object Detection



# Relation between Activity and Objects

- English Gigaword Corpus – 15 GB of raw text
- **Occurrence counts:**
  - of an activity  $A_i$ : occurrence of the verbs
  - of an object  $O_j$ : occurrence of object noun  $O_j$  or its synonym.
- **Co-occurrence of an Activity and an Object:**
  - *POS Tagging*
    - Using Stanford tagger.
    - Occurrence of the object ( tagged as noun ) within a window of  $w$  or fewer words of an occurrence of the activity ( tagged as verb ).

# Relation between Activity and Objects

Probability of each activity given each object

$$P(A_i|O_j) = (\text{Count}(A_i, O_j) + 1) / (\text{Count}(O_j) + |A|)$$

# Integrated Activity Recogniser

- $P(A_i | F_v)$  – Calculated in 1st part.

# Integrated Activity Recogniser

- $P(A_i | F_v)$  – Calculated in 1st part.
- $P(A_i | F_o)$  -

$$P(A_i | F_o) = \sum_{j=1}^{|O|} \underbrace{P(A_i | O_j)}_{\text{Gigaword Corpus}} * \underbrace{P(O_j | F_o)}_{\text{Object Detector}}$$

# Integrated Activity Recogniser

- $P(A_i | F_v)$  – Calculated in 1st part.
- $P(A_i | F_o)$  -

$$P(A_i | F_o) = \sum_{j=1}^{|O|} P(A_i | O_j) * P(O_j | F_o)$$

- Consider only  $P(A_i | F_v)$  when no object is detected and  $P(A_i | F_o, F_v)$  when objects are recognized

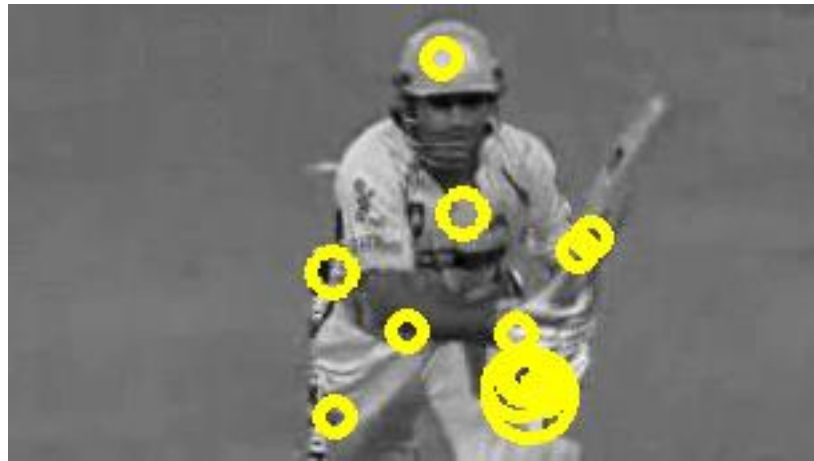
# Morkov Logic Network

- Inference using Knowledge – First Order Logic
- Probabilities – Handle probabilities with FOL
- Problems with Pure Logic and Pure Probabilities
- MLN – Assign real valued weights to rules
  - Learn the weights



# Work Done

- STIP features extraction done.



# Work Done

- STIP features extraction done.
- Clustering – k-means clustering integrated in pipeline.

# Work Done

- STIP features extraction done.
- Clustering – k-means clustering integrated in pipeline.
- Bag-of-Features representation of video clips.

# Work Done

- STIP features extraction done.
- Clustering – k-means clustering integrated in pipeline.
- Bag-of-Features representation of video clips.
- Learning a model

# Work Done

- **STIP features extraction** done.
- **Clustering** – k-means clustering integrated in pipeline.
- **Bag-of-Features** representation of video clips.
- **Learning** a model

# Work To be Done

- Object Detection, Learning Gigaword Corpus.
- Markov Logic Network implementation.

# References

- *Actions in Context* by M. Marszalek, I. Laptev and C. Schmid; in Proc. [CVPR-2009](#)
- *Improving Video Activity Recognition using Object Recognition and Text Mining* by Tanvi Motwani and Raymond J. Mooney, [ECAI-2012](#)
- *Markov Logic* by Pedro Domingos, Parag Singla, et.al., [Probabilistic Inductive Logic Programming](#) (pp. 92-117), [2008](#). New York: Springer.