# EXPLOITING PROBABILISTIC RELATIONSHIPS BETWEEN ACTION CONCEPTS FOR COMPLEX EVENT CLASSIFICATION

Somayeh Keshavarz, Imran Saleemi,

and George Atia

*College of Engineering and Computer Science, University of Central Florida*

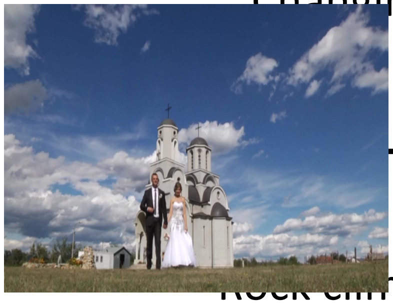
*ICIP 2017, Beijing, China*



UCF

Changing tire

Feeding animal



Rock climbing

Wedding C

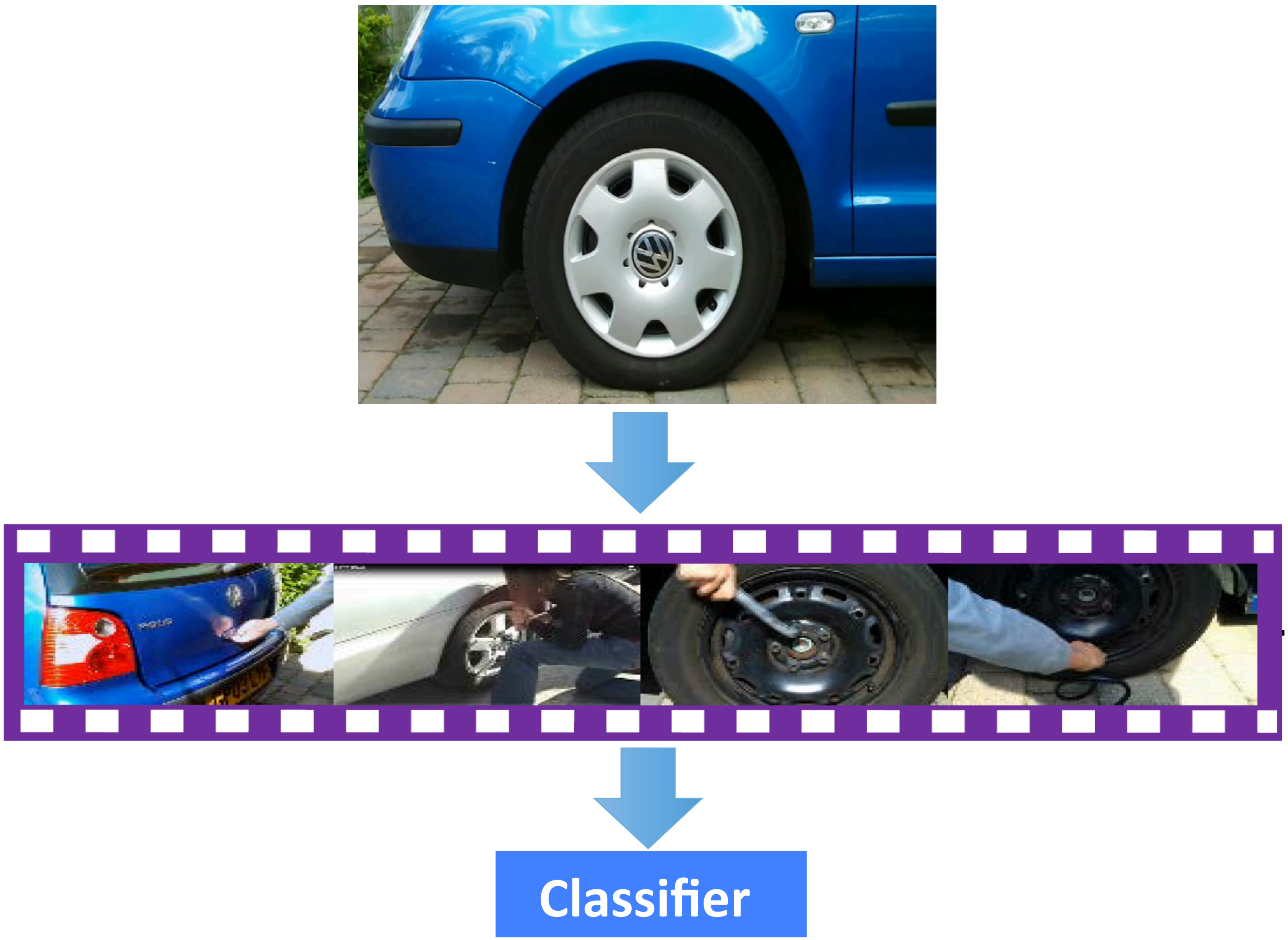
# Challenges of video classification

- Camera motion
- Illumination condition changing
- Background clutter
- Diversity in scene setting


# Bag of words pipeline

- Feature extraction
- Encoding
- Pooling
- Classification

# Decomposing a Video to Concepts



# Outline

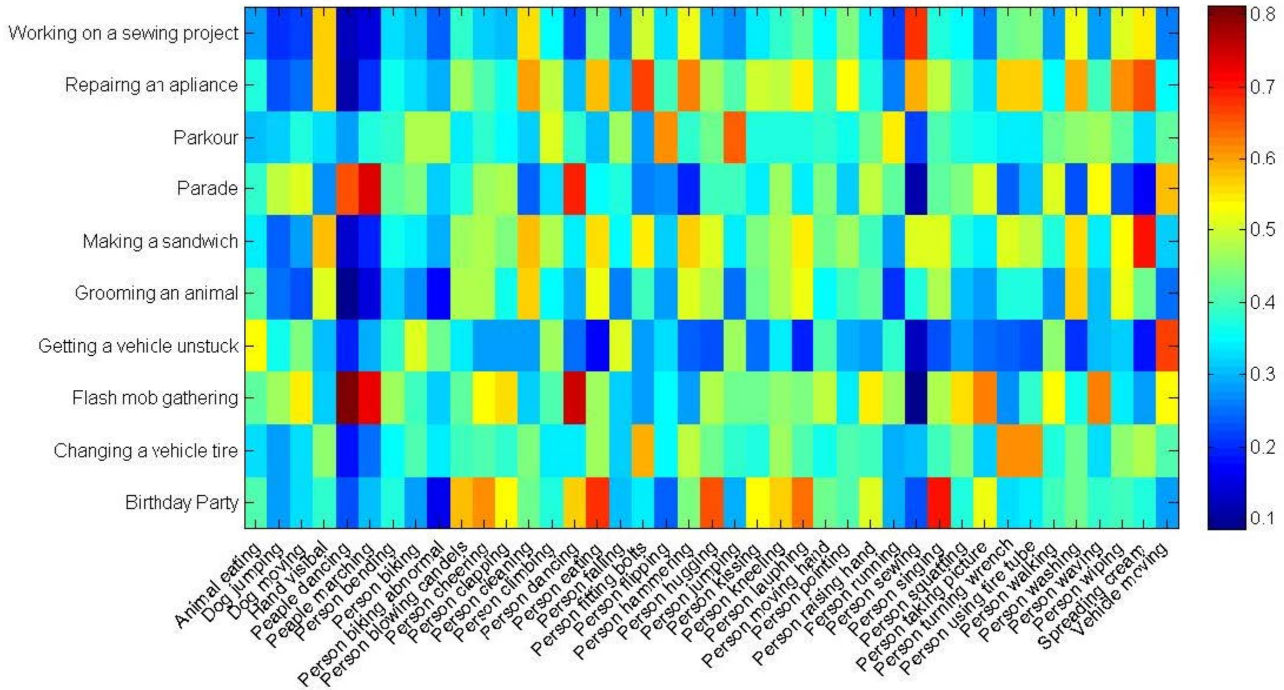
- Proposed classification method
- Dataset
- Result

UCF

# Bayes Classifier

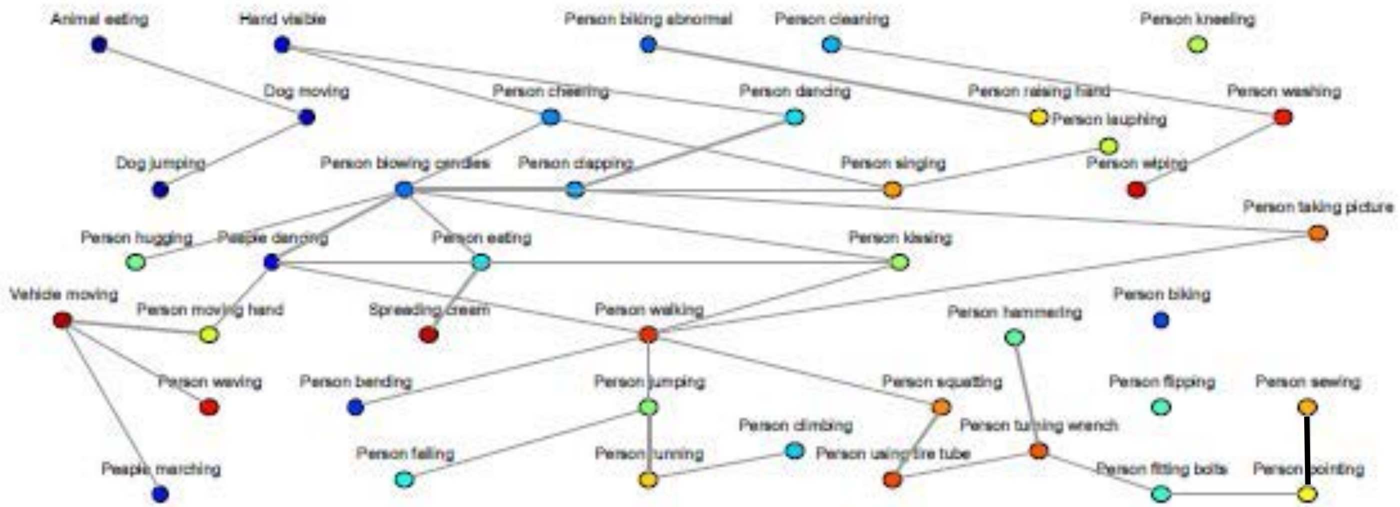

$$\hat{e}_{ML} = \arg\max_{e \in \mathcal{E}} p\left(v \mid e\right)$$

# Naive Bayes classifier

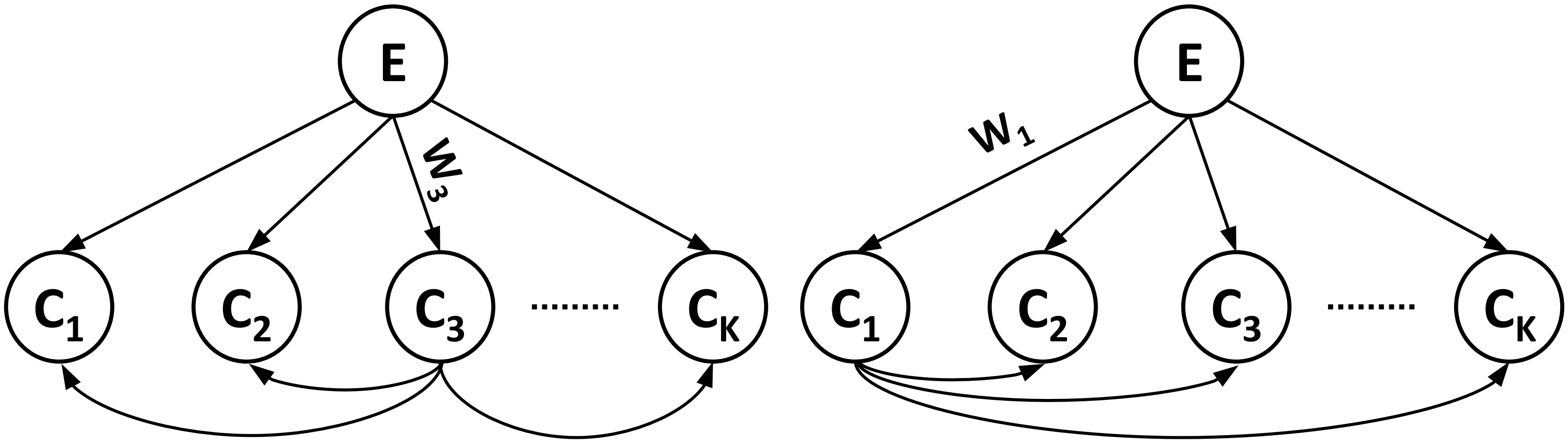



$$\hat{e}_{ML} = \arg\max_{e \in \mathcal{E}} \prod_{i=1}^{k} p\left(c_i \mid e\right)$$

# Concepts co-occurrence



# Weighted-Average One Dependence Estimator




$$\hat{e}_{ML} = \arg\max_{e \in \mathcal{E}} \sum_{i=1}^{r} W_i \cdot p(c_i, e) \prod_{\substack{j=1 \\ j \neq i}}^{k} p\left(c_j \mid c_i, e\right)$$

# Estimation of Probabilities

$$\hat{e}_{ML} = \arg\max_{e \in \mathcal{E}} \sum_{i=1}^{r} W_i \cdot p(c_i, e) \prod_{\substack{j=1 \\ j \neq i}}^{k} p\left(c_j \mid c_i, e\right)$$

- Probability of $c_i$ in e

$$p\left(c_i, e\right) = \frac{F\left(c_i, e\right) + \frac{1}{2n}}{n+1}$$

- Probability of $c_j$ and $c_i$ in e

$$p\left(c_j \mid c_i, e\right) = \frac{F\left(c_j, c_i, e\right) + \frac{1}{2}}{F\left(c_j, c_i\right) + 1}$$

# Weights of Root Concepts

$$\hat{e}_{ML} = \arg\max_{e \in \mathcal{E}} \sum_{i=1}^{r} W_i \cdot p(c_i, e) \prod_{\substack{j=1 \\ j \neq i}}^{k} p\left(c_j \mid c_i, e\right)$$

- Weights are defined based on the method of information gains

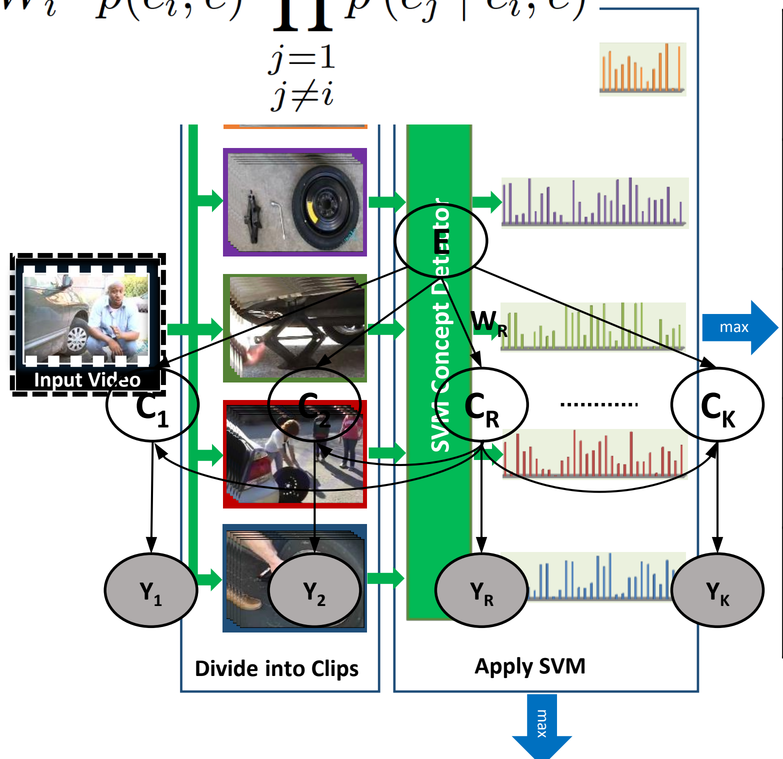
$$\mathrm{W}_i = IG(\mathcal{S}, C_i) = H(\mathcal{S}) - \frac{|\mathcal{S}_i|}{|\mathcal{S}|} H(\mathcal{S}_i) - \frac{|\bar{\mathcal{S}}_i|}{|\mathcal{S}|} H(\bar{\mathcal{S}}_i)$$

- Entropy is defined by summing over all events:

$$H(\mathcal{S}_i) = -\sum_{e=1}^{m} \frac{|\mathcal{S}_i(e)|}{|\mathcal{S}_i|} \log \frac{|\mathcal{S}_i(e)|}{|\mathcal{S}_i|}$$

# Inferring Concepts

$$\hat{e}_{ML} = \arg\max_{e \in \mathcal{E}} \sum_{i=1}^{r} W_i \cdot p(c_i, e) \prod_{\substack{j=1 \\ j \neq i}}^{k} p\left(c_j \mid c_i, e\right)$$



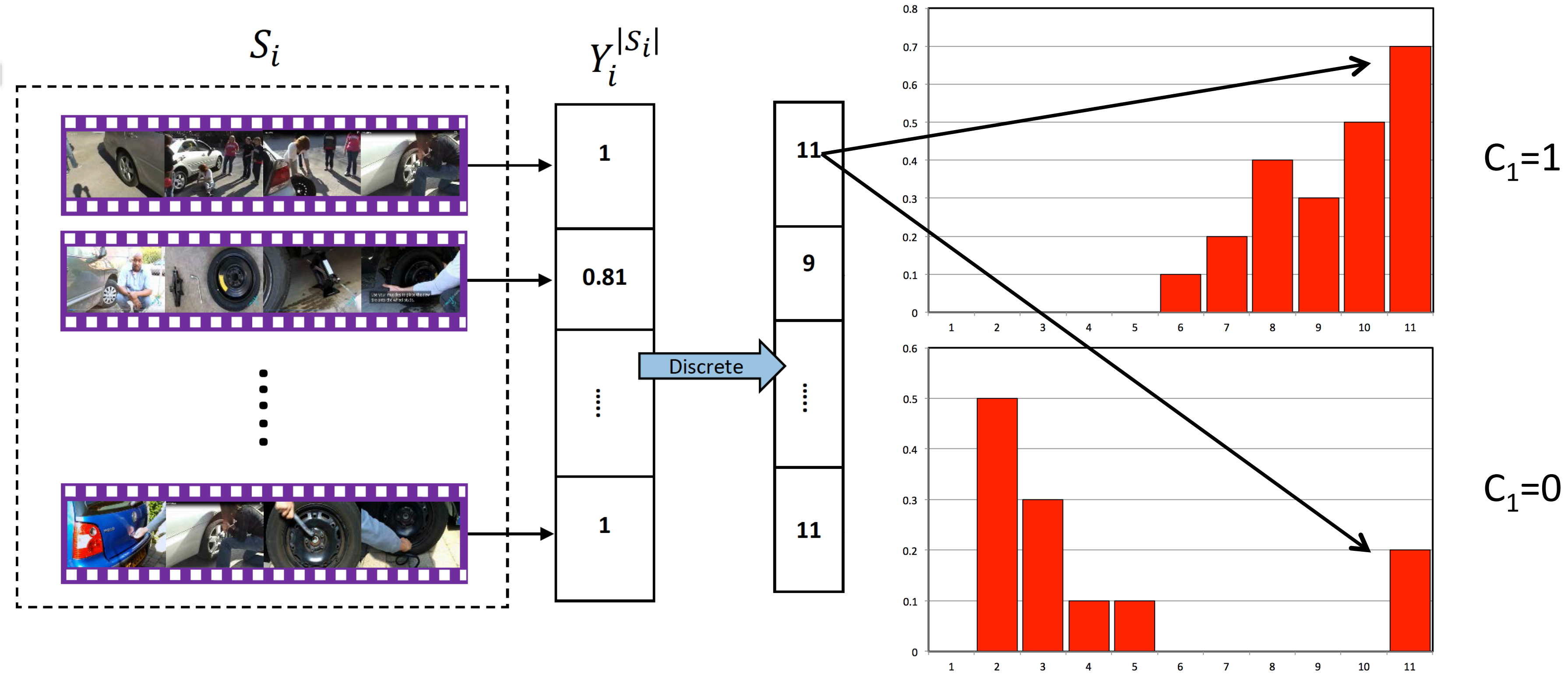

(b)

| | |
|---|---|
| $a_1$ | $C_{20}$ |
| $a_2$ | $C_5$ |
| $a_3$ | $C_{62}$ |
| $a_4$ | $C_{31}$ |
| $a_5$ | $C_{47}$ |

(c)

| $Y_1$ | $Y_2$ | $Y_3$ | ... | $Y_K$ |
|---|---|---|---|---|
| 0.14 | 0.37 | 0.69 | ... | 0.7 |

# Calibrated concept detection




Concept detector result

Annotation information

# Outline

- Proposed classification method
- Dataset
- result

# Datasets

- DEVT (Subset of TRECVID11)
  - 15 complex events
  - Total of 8100 videos

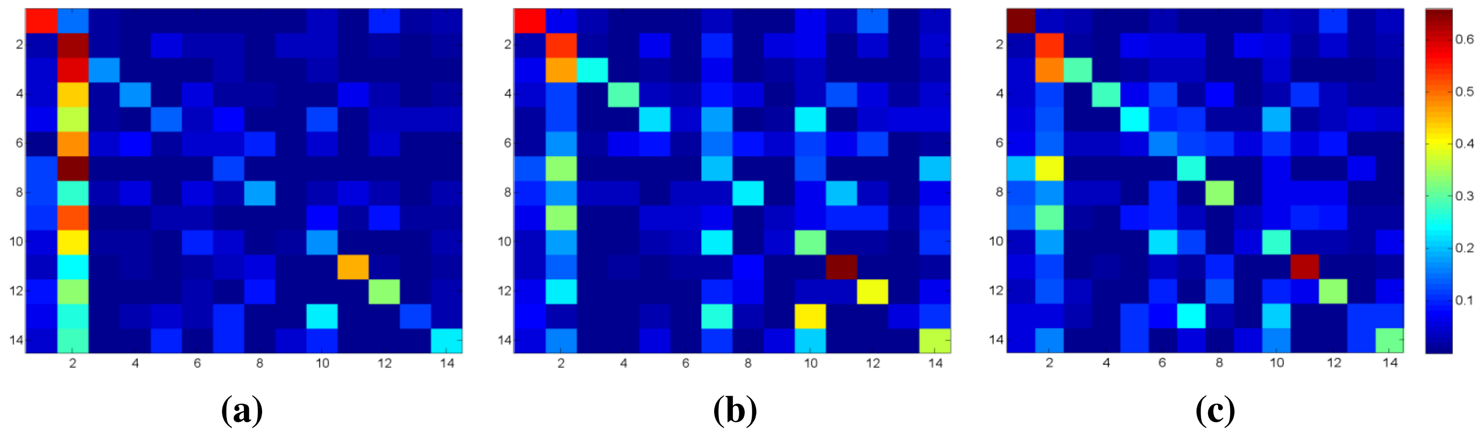
- EC (Subset of TRECVID11 and TRECVID12)
  - 25 complex events
  - Total of 2062 videos

- Concept detector
- Extract Motion Boundary Histogram (MBH) features
- train 93 binary SVM

# Outline

- Proposed classification method
- Dataset
- Result

# Experimental results on DEVT

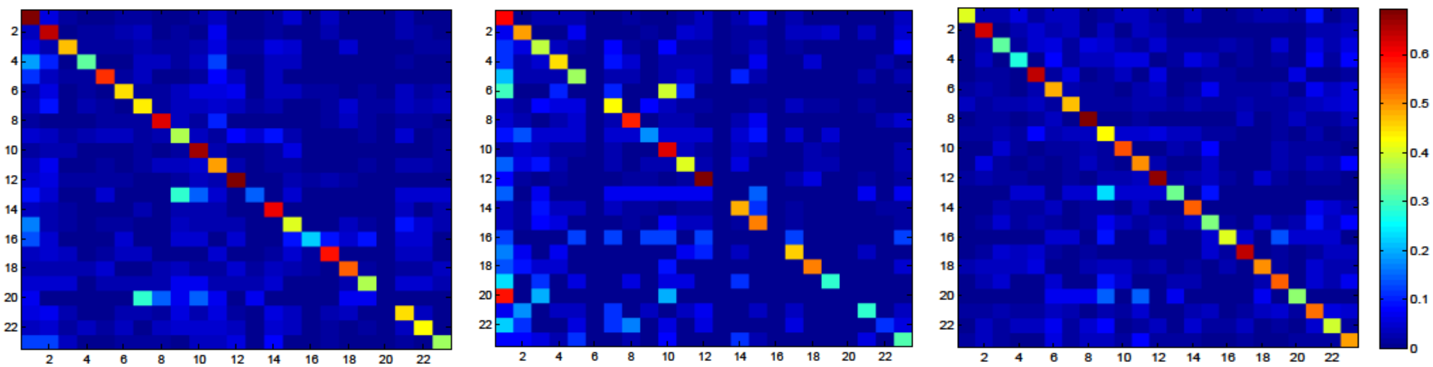


(a) SVM event classifiers AC = 26:59

(b) H-S [19] AC = 29:36

(c) Proposed method AC =32.31%

# Experimental results on EC



(a) (b) (c)

(a) SVM method AC = 46:23

(b) H-S [19] AC = 36:06

(c) Proposed method AC = 48.5%

# Conclusion

- We have proposed a novel probabilistic inference framework for complex video event recognition using supervised action concepts.
- To the best of our knowledge, this is the first principled approach to attempt to model the conditional relationships between events and concepts by constraining dependencies to pairwise joint distributions while avoiding the need to manually re encode new graph structures as the number of concepts increases.
- This method outperforms state-of-the-art techniques on multiple challenging data sets of complex event videos

# Thank You

UCF