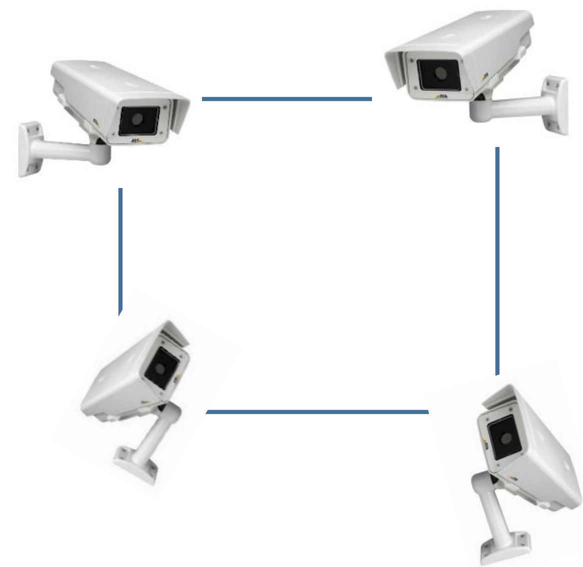
LSSCN Consortium

Activity recognition and anomaly detection in video using temporal and discriminative analysis

Ioannis Kaloskampis, Yulia Hicks
Cardiff University

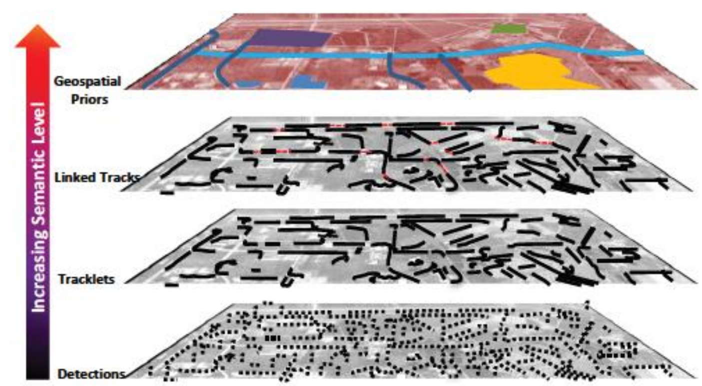
Framework for Activity Recognition and Anomaly Detection – Problem Statement



- Area surveillance problem

- Multiple sensors
  - Overlapping/non overlapping field of views
  - Variable sensor configuration

- Multiple data modalities
  - Video
  - Radar data
  - HCI data
  - Background maps



DARPA WPAFB 2009

Framework for Activity Recognition and Anomaly Detection - Requirements

- Data-driven framework
- Computationally efficient
- Can handle heterogeneous spatio-temporal data
- Can incorporate domain knowledge.

Contributions

- Framework for Activity Recognition and Anomaly Detection
- High-level inference model for activity recognition
  - Hierarchy based on 'importance' of actions
  - Prolonged activities

Anomaly Detection in Video - Framework for Activity Recognition

STAGE 1
LOW LEVEL FEATURES

→

STAGE 2
MID-LEVEL REPRESENTATION

→

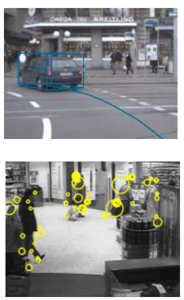
STAGE 3
ACTION RECOGNITION

→

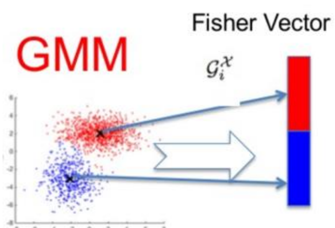
STAGE 4
SITUATIONAL ASSESSMENT

→

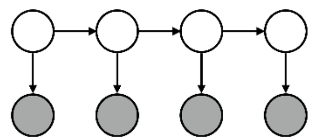
STAGE 5
HIGH-LEVEL INFERENCE MODEL



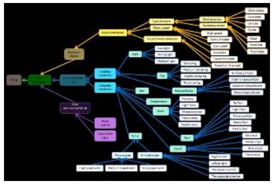
Track data, local features



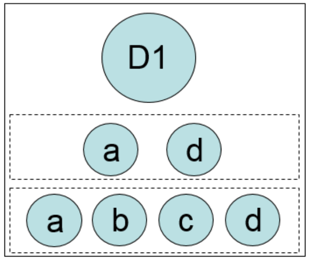
video segmentation, bag-of-words, Fisher vectors (soft clustering)



Continuous HMMs
Detects action primitives



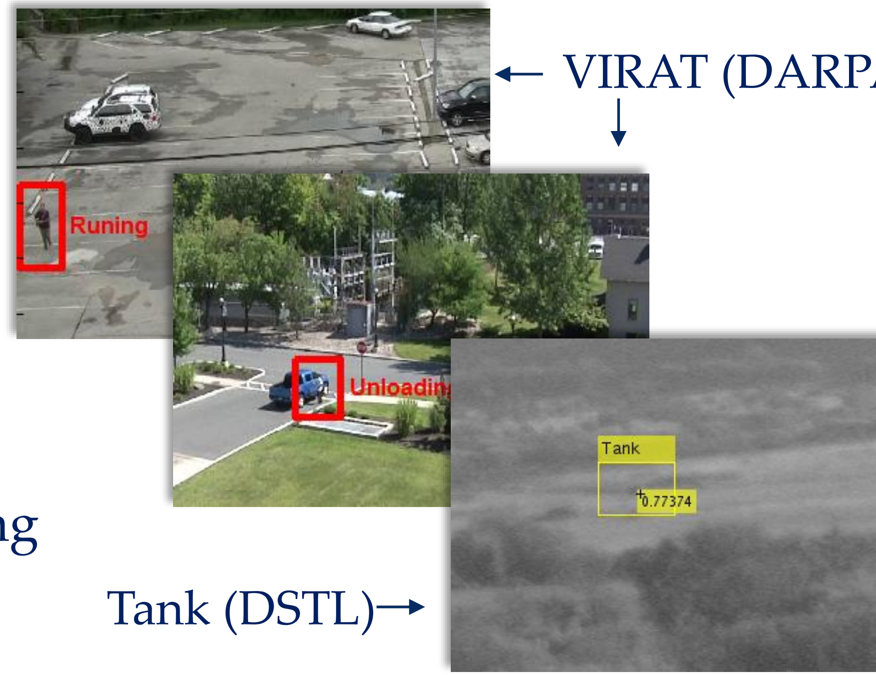
Ontologies, temporal models



Hierarchical Importance Model
Detects activities in heterogeneous data streams

Low-level features

- We employ two approaches:
  1. Objects of interest in a scene detected and tracked; events modelled as trajectories
     - VJ / HOG detectors
     - Kalman / Particle filter tracking
     - Rely on accuracy of tracker/ detector
  2. Local spatio-temporal features
     - Repeatable across different videos
     - HOGHOF, DTF
     - Unsupervised detection
     - Generic approach
     - Large vectors (D=162→426)



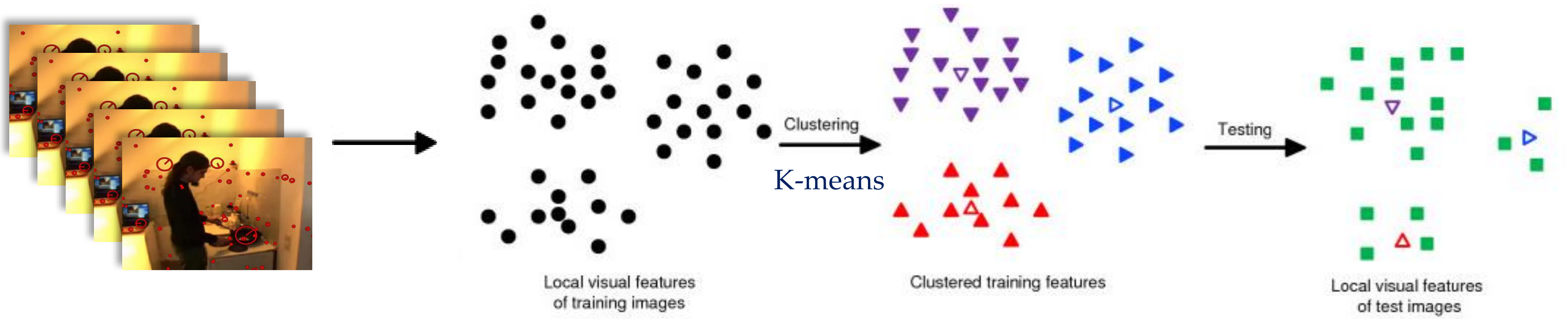
← VIRAT (DARPA)

Tank (DSTL)→

BF (Brown Univ.)→

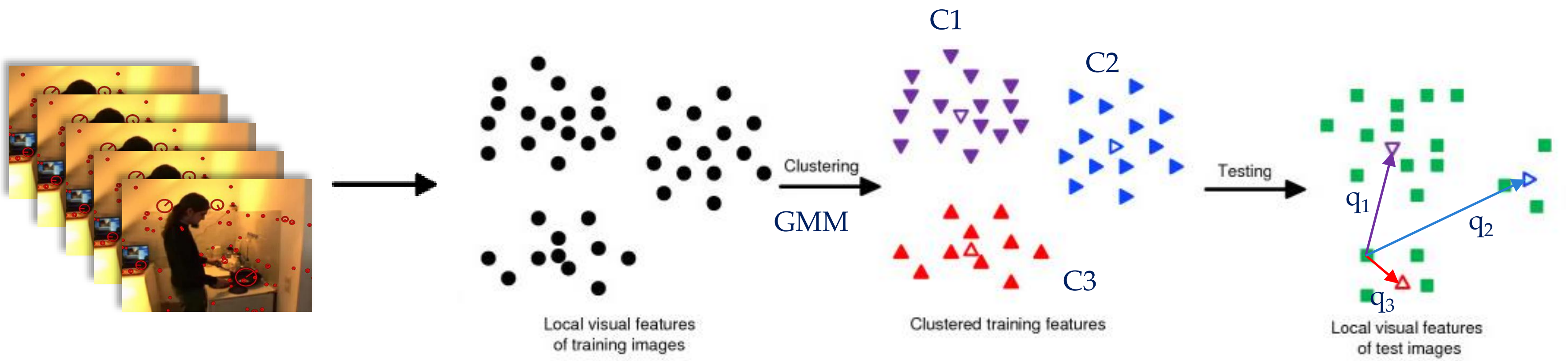
Mid-level representation

- Classic approach: Bag-of-Words
  - Build a dictionary of ‘words’ by clustering training features
  - Hard assign features to clusters - ‘words’



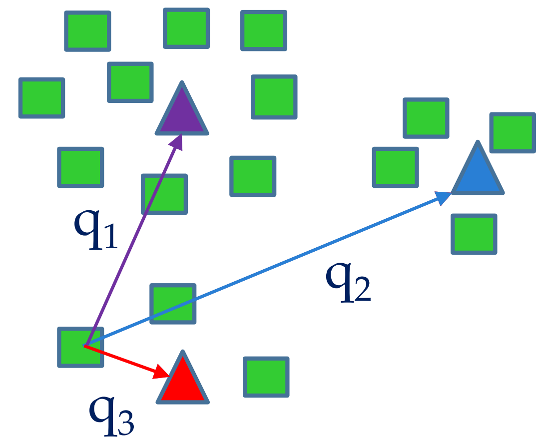
Mid-level representation

- Modern approach: Fisher vectors (Peronnin and Dance 2007)
  - Soft assignment of features to clusters
  - More detailed and accurate approach



Mid-level representation

- Modern approach: Fisher vectors
  - Soft assignment of features to clusters
  - $I = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ set of D dim. feature vector
  - GMM: $\Theta = (\mu_k, \Sigma_k, \pi_k : k = 1,\ldots,K)$
  - Associate each $\mathbf{x}_i$ to a mode k with strength:



$$
\left.
\begin{aligned}
q_{ik} &= \frac{\exp\left[-\frac{1}{2}(\mathbf{x}_i - \mu_k)^T \Sigma_k^{-1} (\mathbf{x}_i - \mu_k)\right]}{\sum_{t=1}^{K} \exp\left[-\frac{1}{2}(\mathbf{x}_i - \mu_t)^T \Sigma_k^{-1} (\mathbf{x}_i - \mu_t)\right]} \\
u_{jk} &= \frac{1}{N\sqrt{\pi_k}} \sum_{i=1}^{N} q_{ik} \frac{x_{ji} - \mu_{jk}}{\sigma_{jk}}, \qquad j = 1,2,\ldots D \\
v_{jk} &= \frac{1}{N\sqrt{2\pi_k}} \sum_{i=1}^{N} q_{ik} \left[\left(\frac{x_{ji} - \mu_{jk}}{\sigma_{jk}}\right)^2 - 1\right]
\end{aligned}
\right\}
\qquad
\Phi(I) = \begin{bmatrix} \vdots \\ \mathbf{u}_k \\ \vdots \\ \mathbf{v}_k \\ \vdots \end{bmatrix}
$$

Mid-level representation

- Modern approach: Fisher vectors
  - Problem dimensionality increases
    - FV.dim = 2 x (Gaussians) x (Dims)
    - FV.dim = 6000 ~ 30000
  - PCA with efficient SVD algorithms
    - LM-SVD (Liu et al., 2013)
    - F-SVD (Halko et al., 2010)
    - SVDS (MATLAB)
  - dim → 64, 128, 256

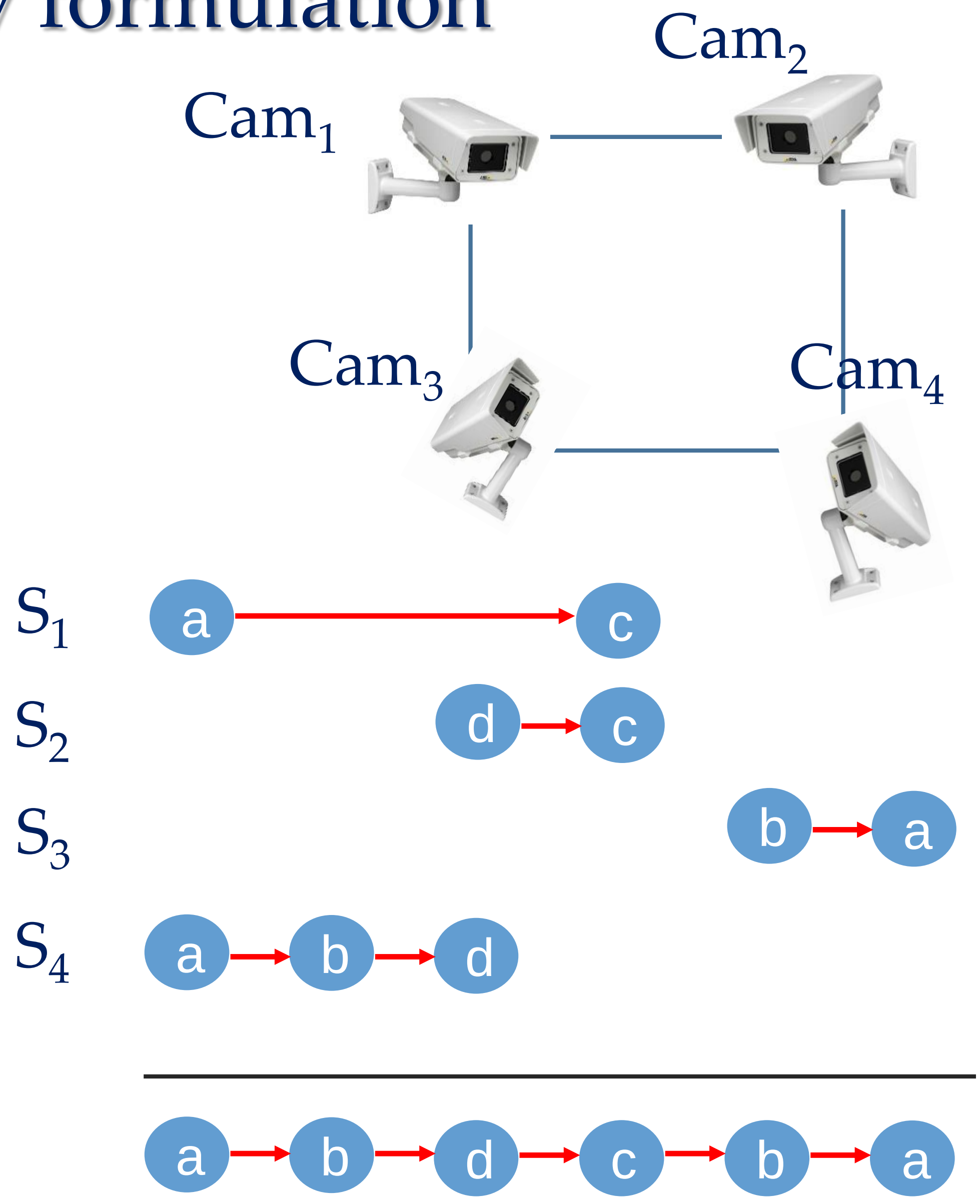
Dim 6816 → 64	
Algorithm	**Computation time (sec)**
LM-SVD	1.95
F-SVD	0.89
SVDS	83.04

Recognition and anomaly detection at action level

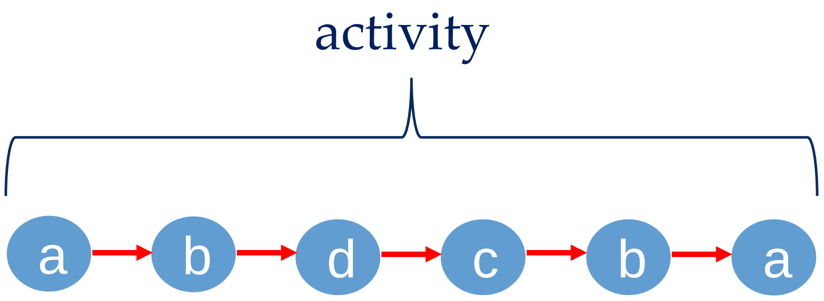
- Action detection
  - HTK framework
  - Based on continuous HMMs
  - Output: Probability for each of the *n* possible actions for each testing sample
- Anomaly detection
  - Unrecognised actions
  - Actions with low confidence score (thresholds)

Data fusion and activity formulation

- Each sensor outputs a time sequence of actions
- Actions are put together in chronological order to formulate the activity sequence
- Actions can be represented by their start and end points to handle overlaps



Activity recognition

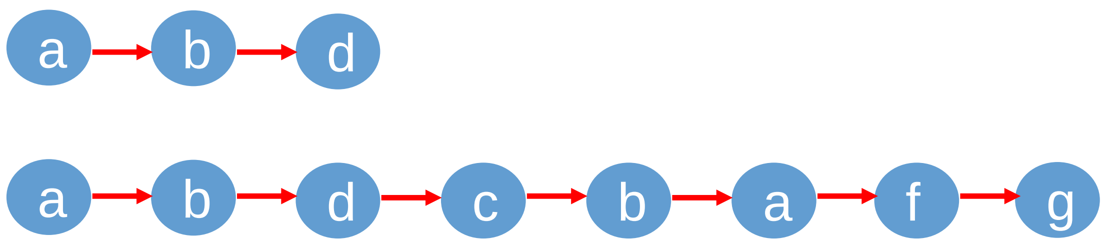


- High level inference
  - HIM algorithm (proposed)
  - Output: One of the m possible activities for each testing sample

- Anomaly detection at high level
  - Proximity to activity classes (thresholds)
  - Define classes corresponding to abnormal behaviour

Action detection and inference

- Handling prolonged activities



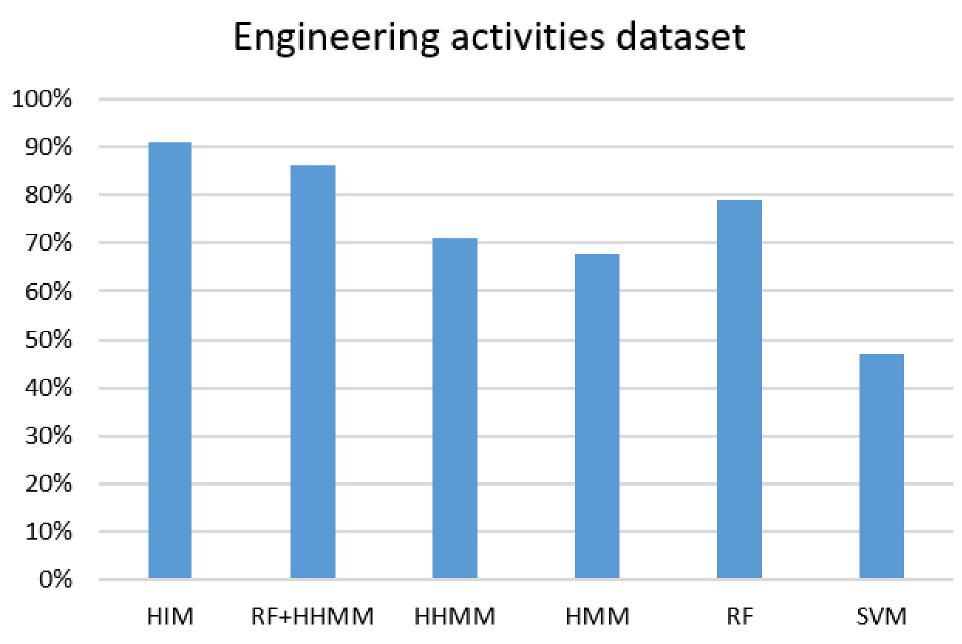
Simple

Prolonged

- Challenge 1: temporal dependencies cannot be efficiently encoded by a simple state model (e.g. HMM)
- Challenge 2: discriminative properties of simple state model inadequate for long sequences
- Solutions:
  - ➢ Hybrid classifiers ~RF+HHMM (Kaloskampis et al. 2011)
  - ➢ HIM algorithm (proposed)

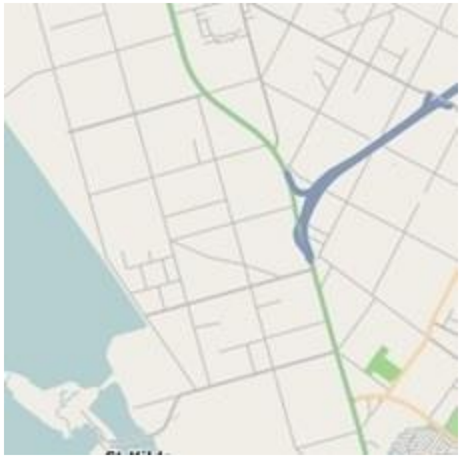
Framework Evaluation: ENGIN dataset

- Activity recognition in video
  - Cardiff ENGIN dataset (IMA 2014).
- Prolonged action sequences
- Temporal dependencies and discriminative properties
- Anomaly detection: erroneously executed activities



Framework Evaluation – Next Steps

- WASABI dataset
- Wide area surveillance challenge
- Sample received from DSTL (Richard Green)
- Available modalities:
  - Background maps
  - Full motion video
  - Wide area motion imagery
  - Radar data
  - Track data
- Challenges
  - Anomaly detection
  - Activity understanding



Framework Evaluation - Next Steps

- FMV data received:
  - 68 HD videos
  - Res: 1920x1080
  - Frame rate: 25 FPS
- 3 video types
  - Low, High, IR
- Initial approach
  - Exploit track data
  - Set up grid on the surveilled area
  - Learn model of normal behaviours from a number of trajectories
  - HIM behaviour analysis algorithm
  - Detect outliers