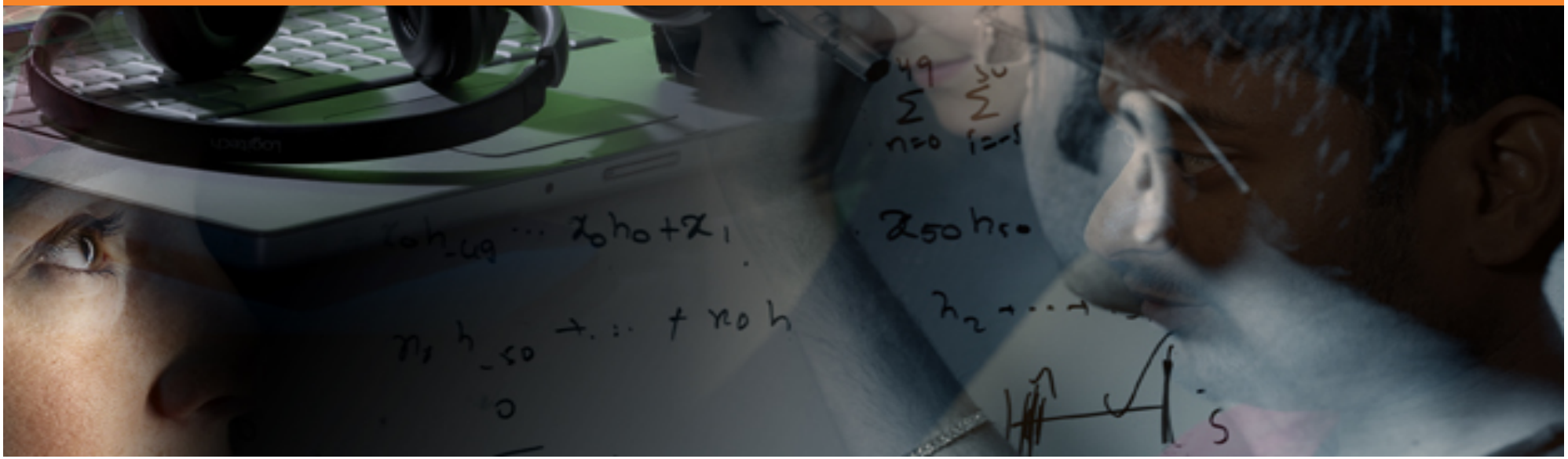
What's going on ?
Discovering temporal motifs from sensor logs

Jean-Marc Odobez

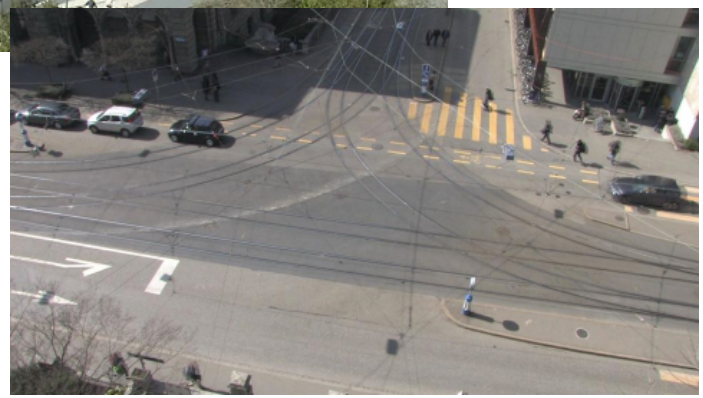
International Advisory Board Meeting, Idiap, Sep. 2, 2011

Research Activities



Motivation and applications

- More an more data generated (appliances, cell phones, web…)
  - Finding activity patterns, with main goals:
    - discover, explain, analyse (user, activities, relationships)
    - recommend, summarize
- Surveillance areas
  - Activity analysis is important



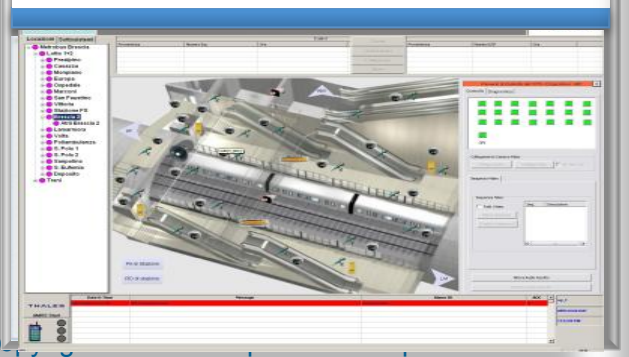
Motivation and applications

- Millions of cameras recording data passively everyday
  => used mainly for future references

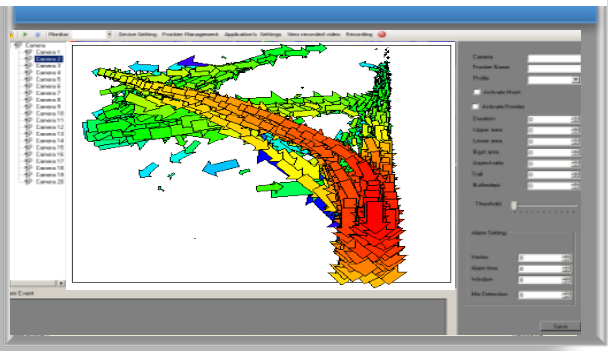
- Activity analysis useful for event detection and automatic behavior analysis
  - abnormal situation detection
  - alarm/stream selection in control rooms
  - activity models & statistics
    infrastructure planning and management

Autonomous stream selection

Planning applications

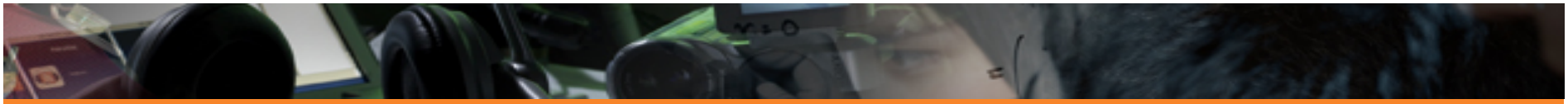


Collective behaviors building



Surveillance video monitoring

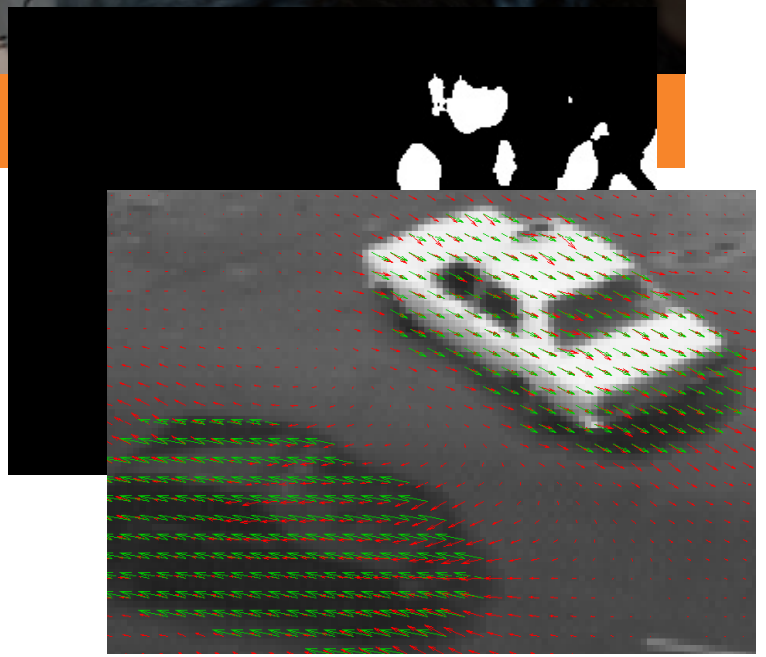
- Goal: automatic activity patterns discovery
  - extract temporal order of sub-events within the discovered activities
  - detect when activities occur
  - Defines normal activity => deviation = abnormality
- Issue
  - activities occur concurrently, with/without synchronization
  - apply to many cameras, low quality, weak activity structure



Outline

- Activity approach
  - General principles
  - Extensions
- Results and evaluation
- Conclusion and perspectives

Overall approach

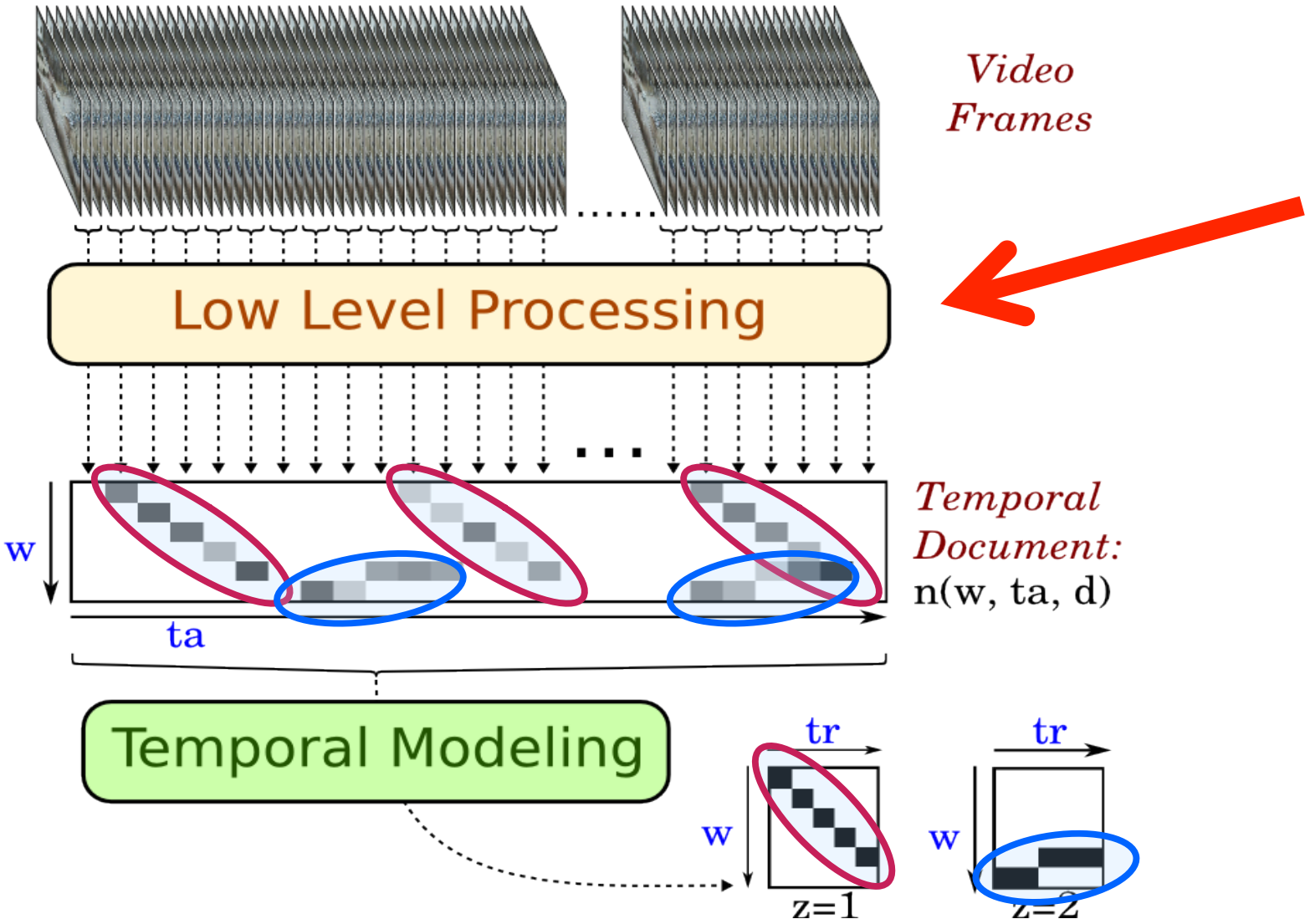


- Low-Level Features
  - object centric features trajectories – difficult in crowded scenarios
  - allows use of compressed domain features

- Unsupervised Methods
  - labeling data is time consuming & error prone

- Topic Models
  - efficient to mine dominant patterns through co-occurrence analysis

Approach overview



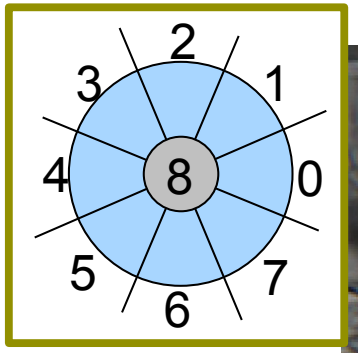
Temporal Motifs and When they occur

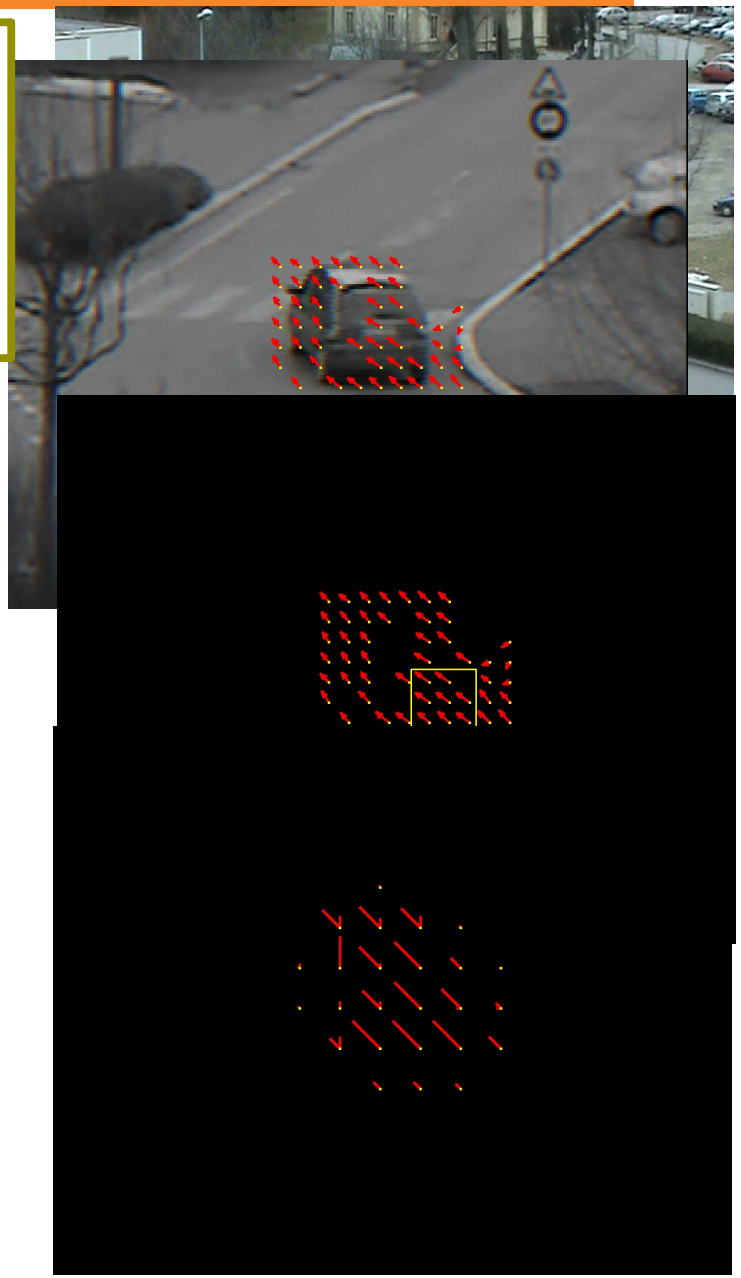
From Video to Low Level Documents

- Video

- Optical flow
- Quantization
  - 8 directions + very low motion
  - Spatial blocks

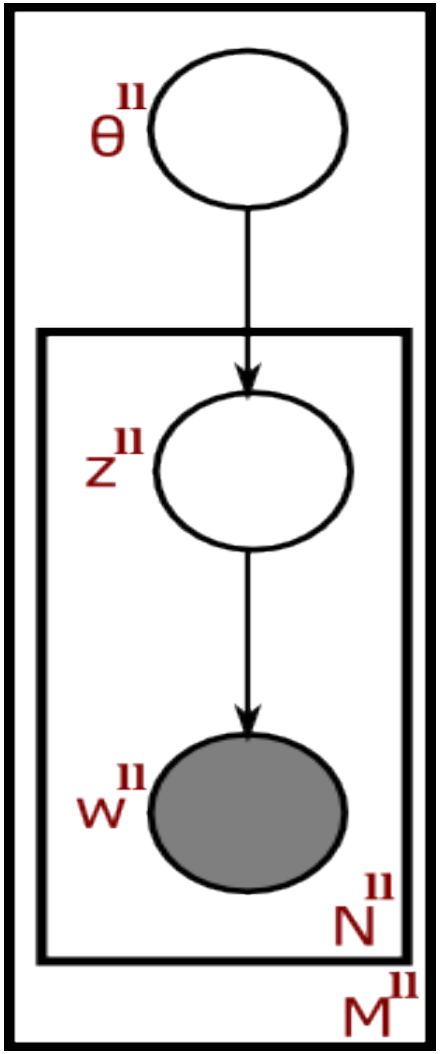
- Low-level documents
  - Low-level words: (position,motion)
  - Vocabulary size: 10k to 100k
  - Bag-of-words over 1 second window





Dimensionality reduction using PLSA (or HDP)

- Input document (vocabulary size: **10k-100k**)
- PLSA topic model
  - Decomposes document into mixture of topics
  - Soft clustering of data
  - Co-occurrence analysis
- Output
  - Topics (groups of low-level words)
  - Between **25-100** topics
  - Low-level topics => high-level words

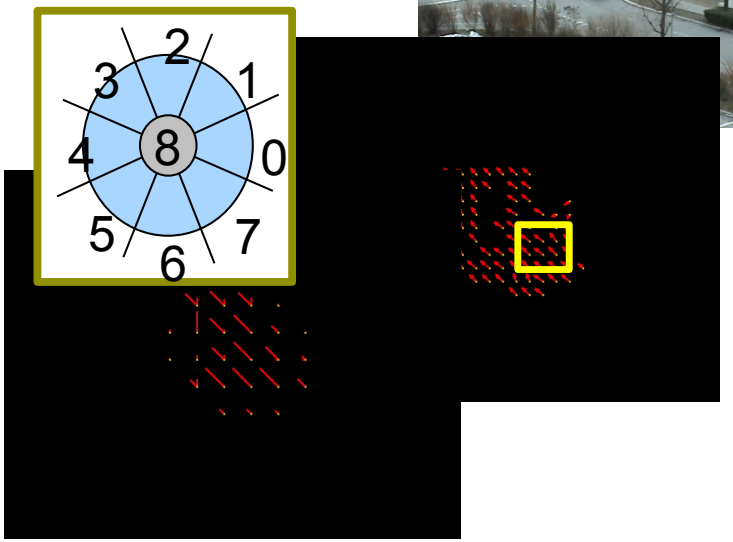


Low-level topic examples ≈ High-level words

- Note: topic distributions over (position,motion) words

From video to temporal document

- Low-level
  - Optical flow, quantization
  - PLSA : low-level topics



Temporal Document

- At each time instant (1 second)
  - Low-level topic presence = amount of high-level word

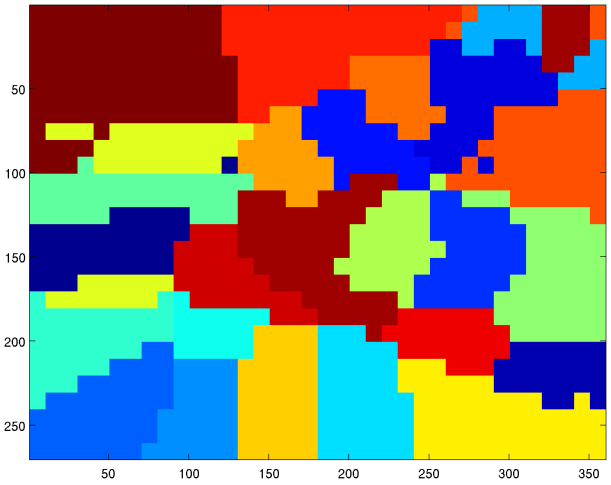
Main idea

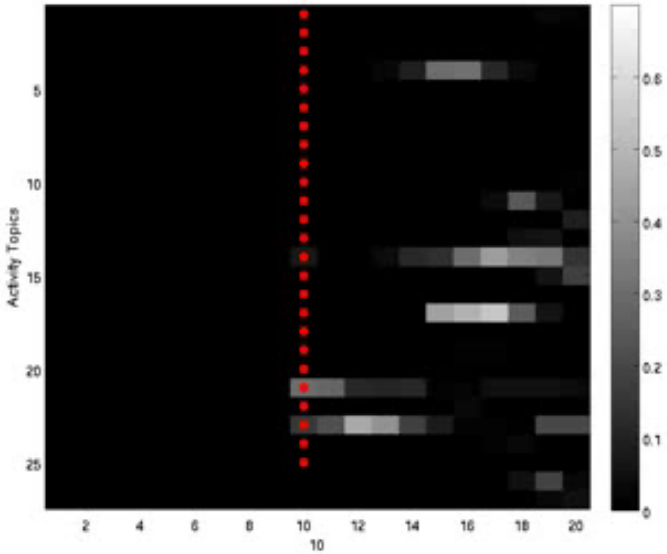
video

words

Low-level topics

Distribution over position and low-level features





low-level activities (high-level words)

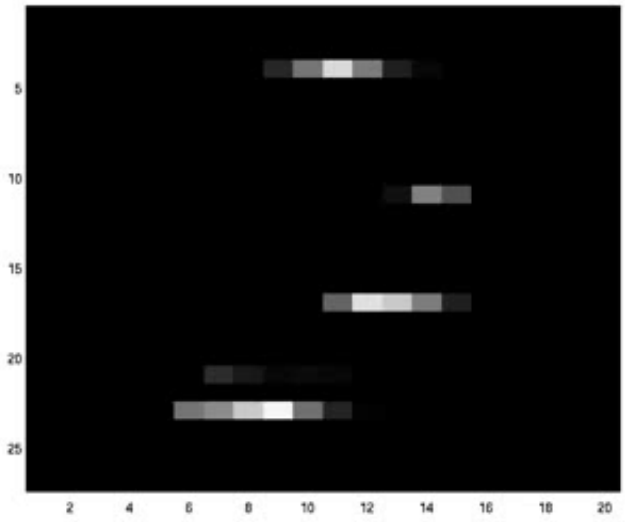
vs time

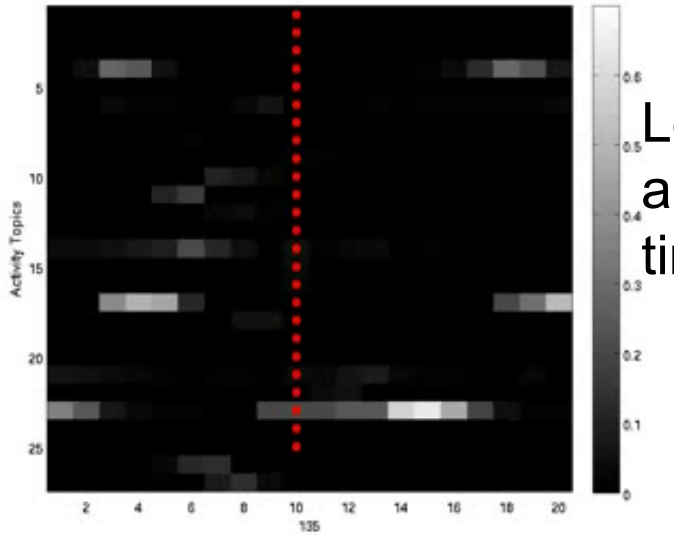
Main idea: activities = spatio-temporal patterns

video

words

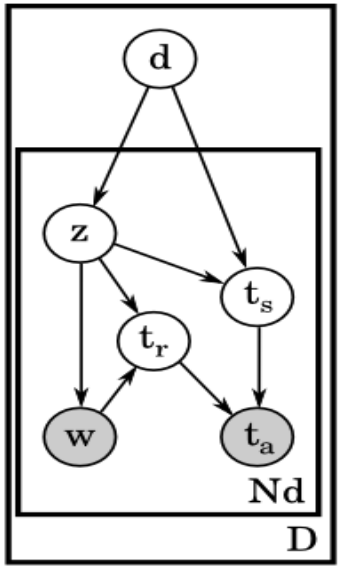
Temporal activity pattern



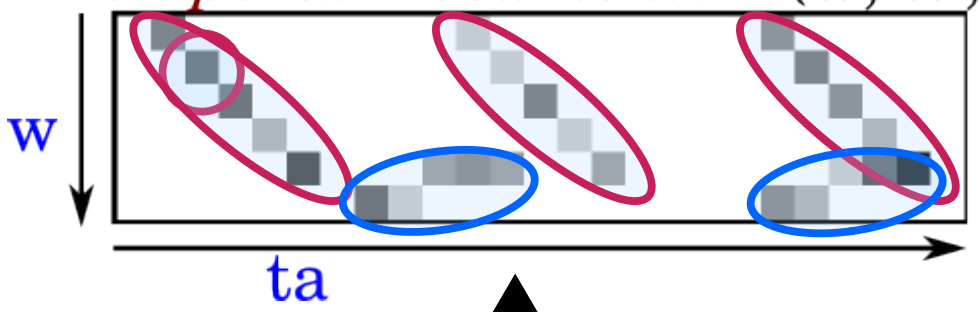


Local activities vs time

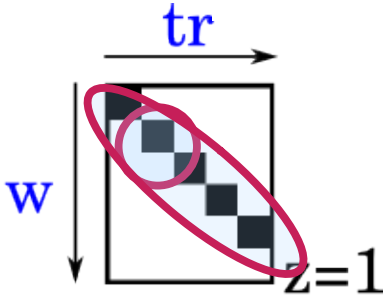
Probabilistic Latent Sequential Model (PLSM)

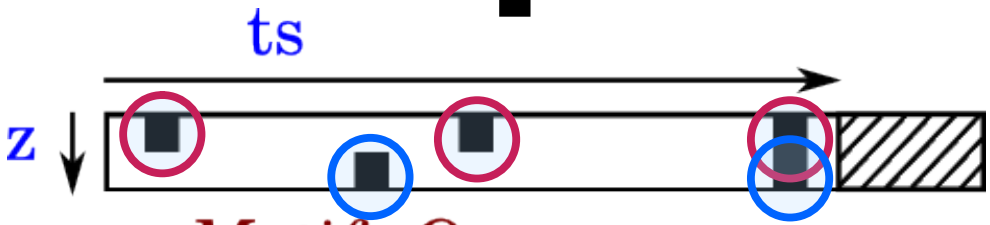


Temporal Document: n(w, ta, d)



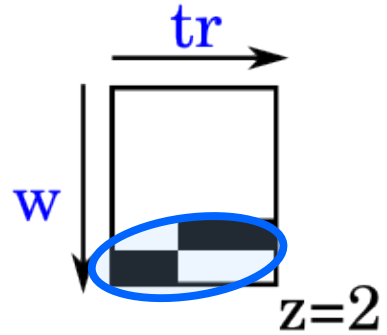
Generative Process





Motifs Occurences:

$p(z,ts \mid d)$

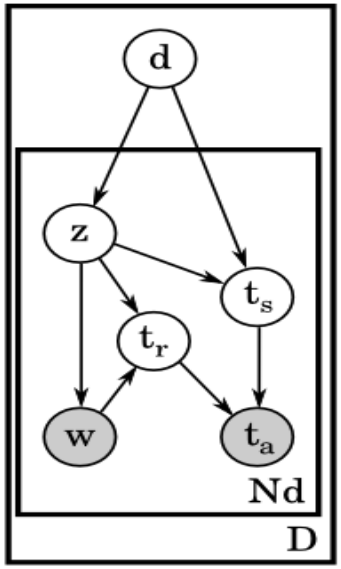


Temporal Motifs:

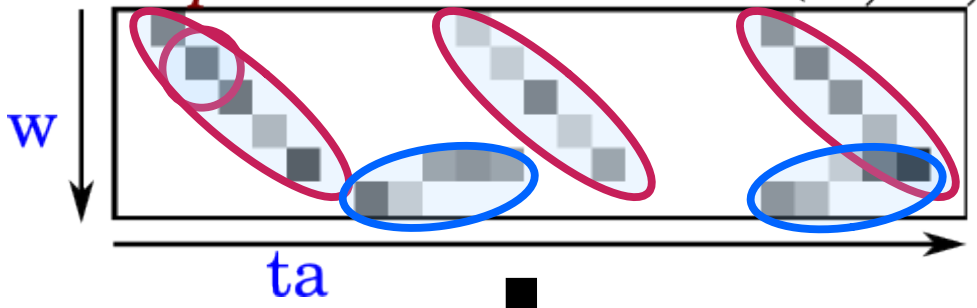
$p(w,tr \mid z)$

Probabilistic Latent Sequential Motifs: Discovering temporal activity patterns in video scenes, *Varadarajan, Emonet, Odobez*, **BMVC, 2010**

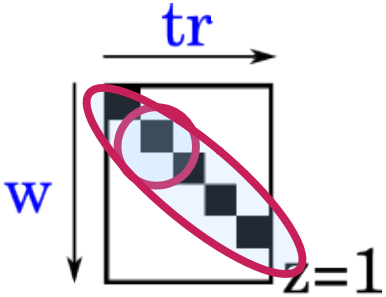
Probabilistic Latent Sequential Model (PLSM)

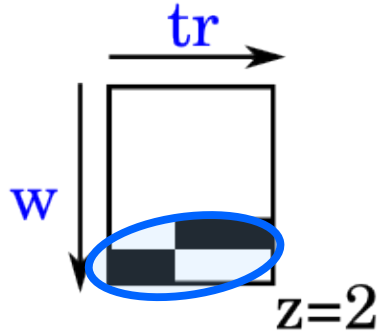


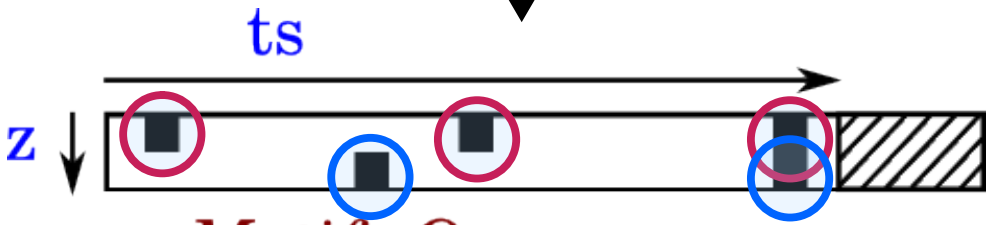
Temporal Document: n(w, ta, d)



Training:
Learning Motifs
Using an EM Algorithm





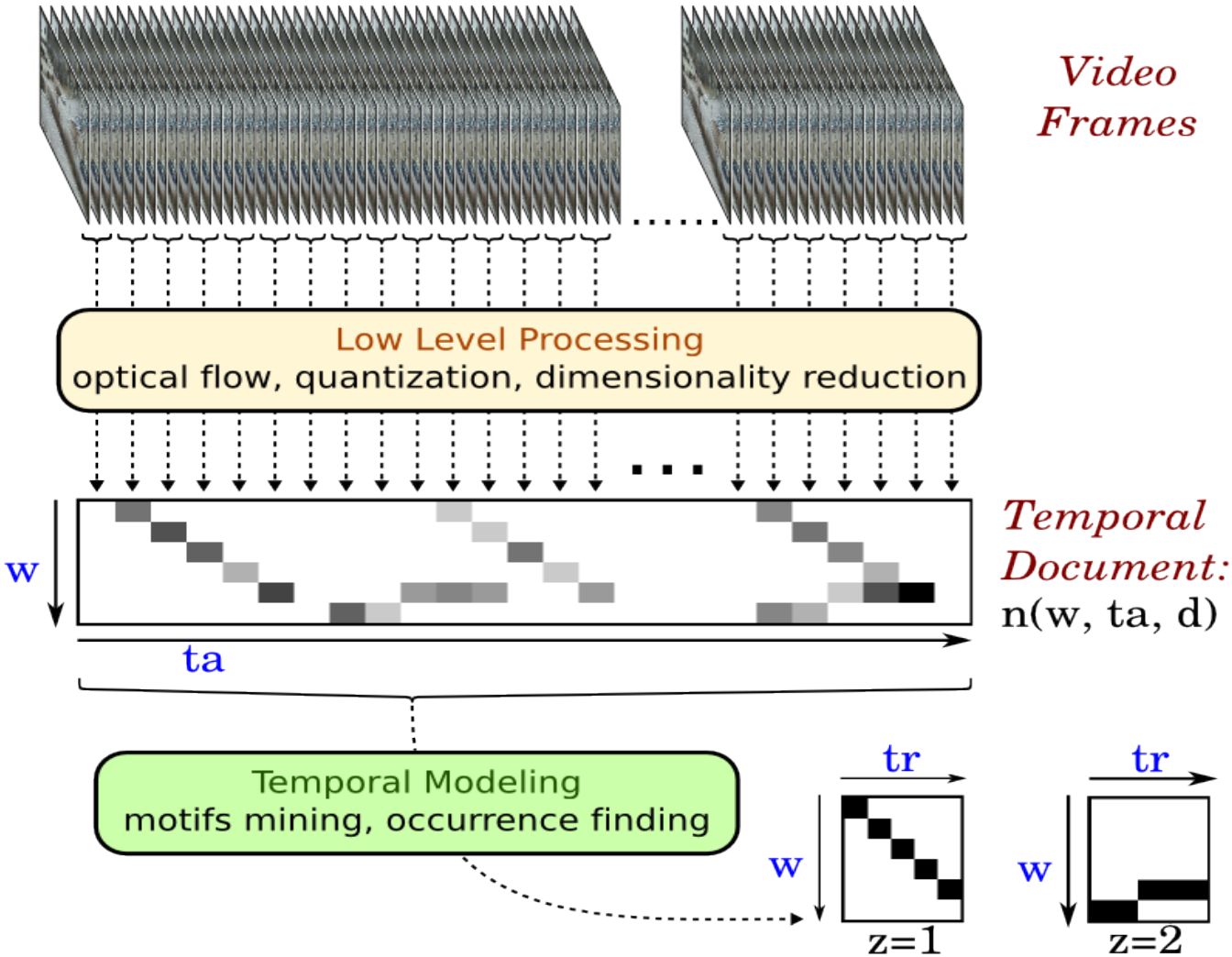


Motifs Occurences:
p(z,ts | d)

Temporal Motifs:
p(w,tr | z)

Probabilistic Latent Sequential Motifs: Discovering temporal activity patterns in video scenes, *Varadarajan, Emonet, Odobez*, **BMVC, 2010**

Model recapitulation



+ Extensions

Probabilistic Latent Sequential Motifs: Discovering temporal activity patterns in video scenes, *Varadarajan, Emonet, Odobez*, **BMVC, 2010**

Inference scheme: sparsity issue

- Learning issue: no constraints on distributions to be learnt
  - multinomial distributions => probability tables => many non-zero entries
  - in practice, one **expect multinomials to be peaky/sparse**
    e.g. topic starting times

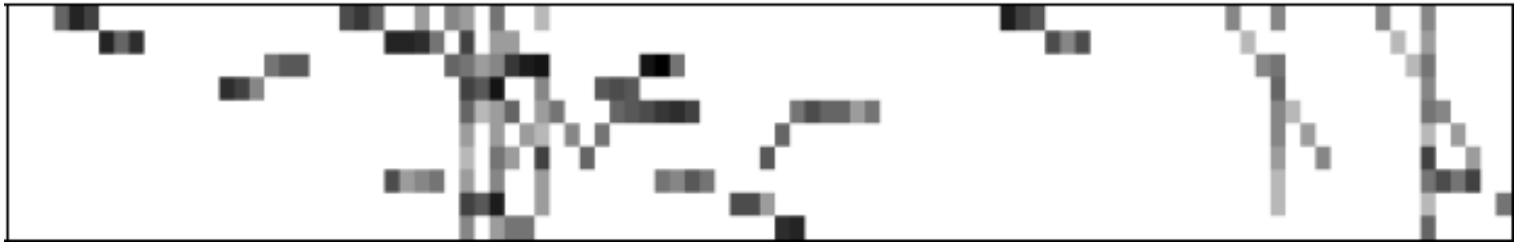
- Approach $\mathcal{L}_c(\mathcal{D}|\Theta) = \mathcal{L}(\mathcal{D}|\Theta)$
  - Sparsity constraint on $p(t_s|z,d)$
    - peaky or sparse distributions **=> small entropy**
    - indirectly achieved
      => maximize Kullback-Leibler divergence with uniform distribution U
  - Lead to simple modification of the EM algorithm

A Sparsity Constraint for Topic Models - Application to Temporal Activity Mining, *Varadarajan, Emonet, Odobez*, **NIPS workshop on Sparse Modeling**, 2010

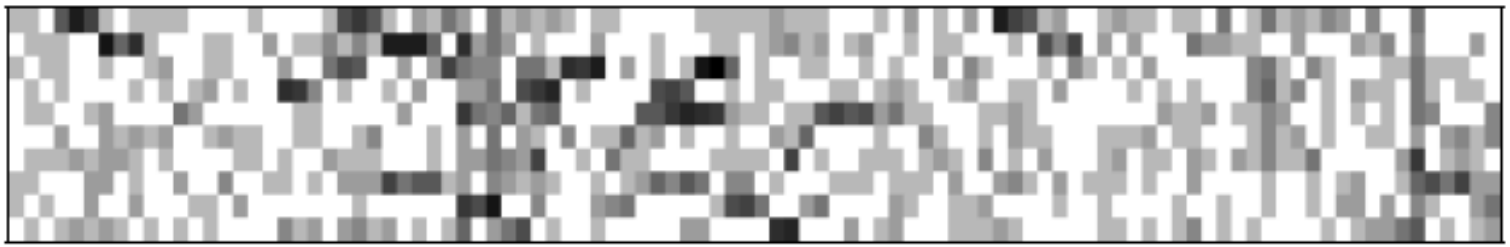
Experiments : synthetic data

- Document generation
  - 5 topics with 6 to 10 time steps

  - Random generation of 10 documents e.g.

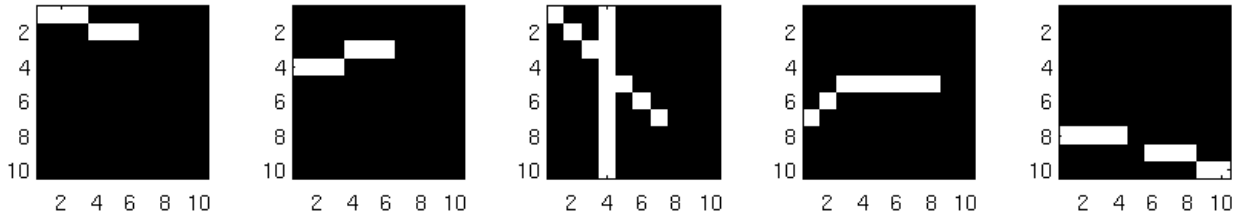


+ noise

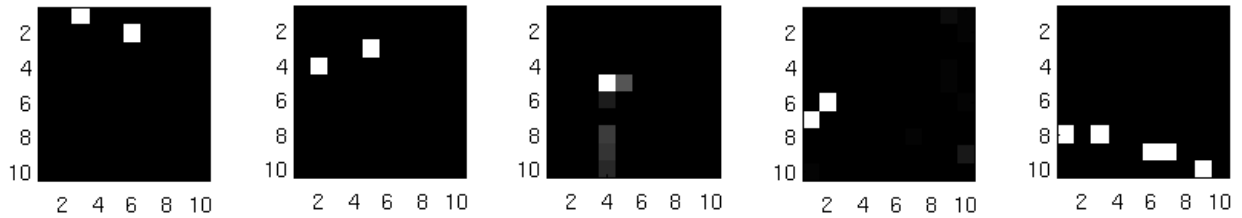


Synthetic experiments – sparsity impact

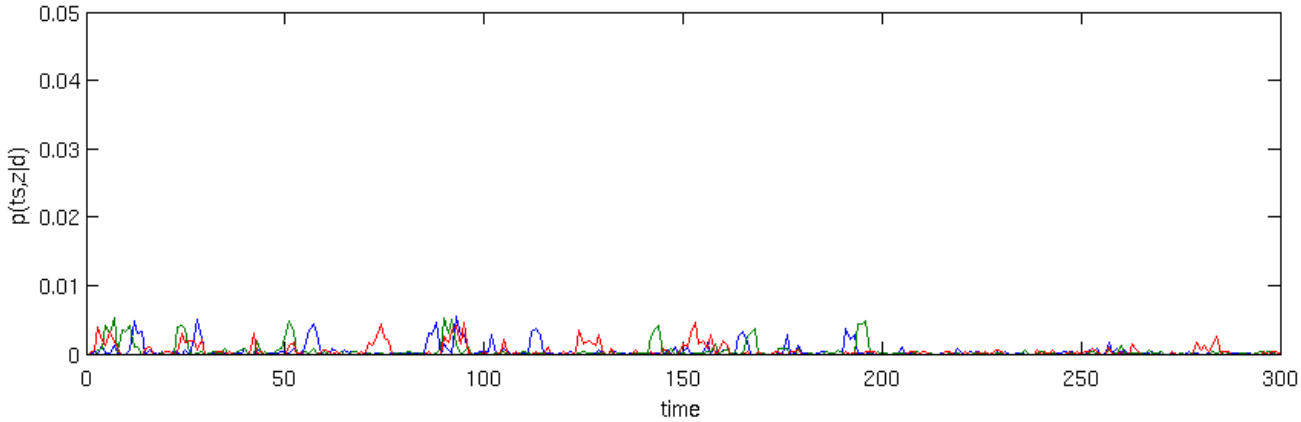
True topics



Recovered topics

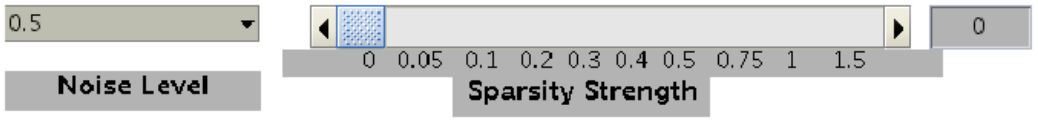


$p(t_s|z,d)$



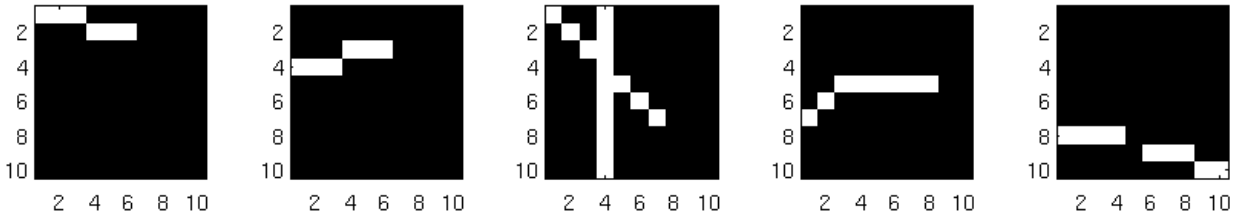
WITH Noise

No Sparsity constraint

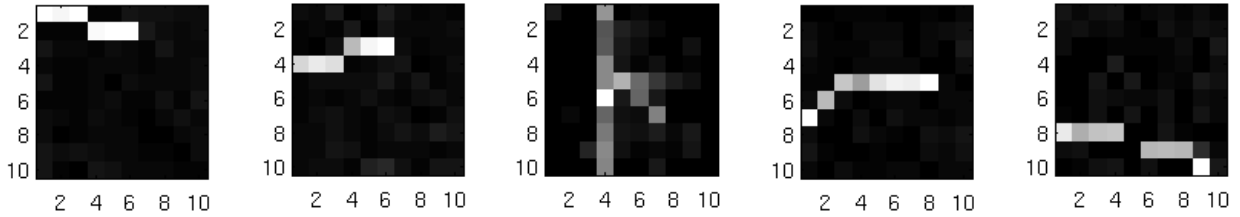


Synthetic experiments – sparsity impact

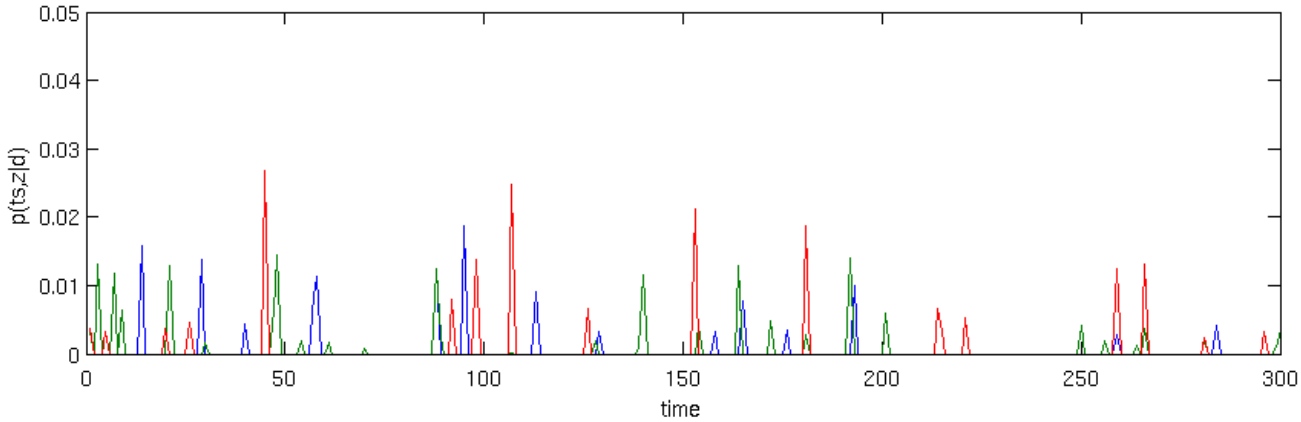
True topics



Recovered topics

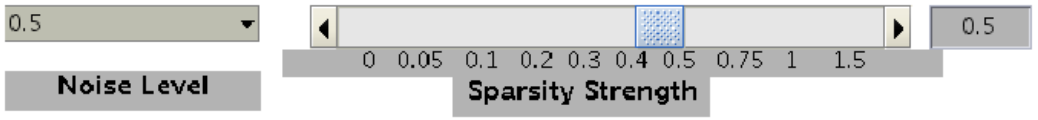


$p(t_s|z,d)$



WITH Noise

WITH Sparsity constraint



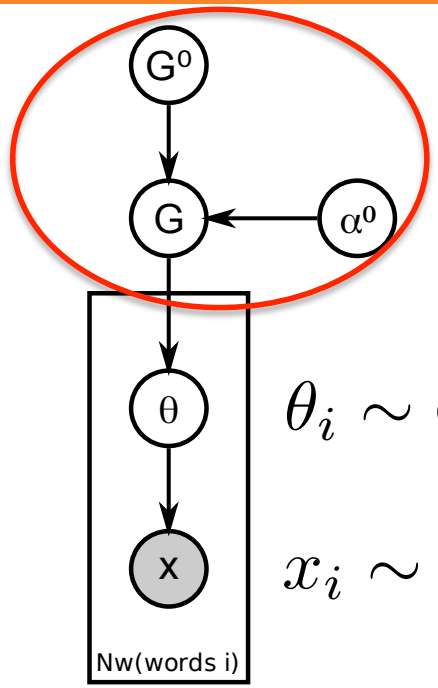
Model selection and Dirichlet Processes

- Limitations:
  - Parameter setting (number of topics, max of temporal extent)
  - No constraints on motif distributions to be learnt

- Solution: exploit Dirichlet Processe

Extracting and Locating Temporal Motifs in Video Scenes Using a Hierarchical Non Parametric Bayesian Model, *Emonet, Varadarajan, Odobez*, **CVPR 2011**

Dirichlet Processes



$$G = DP(\alpha_0, G^0)$$

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k} \quad \phi_k \sim G^0$$

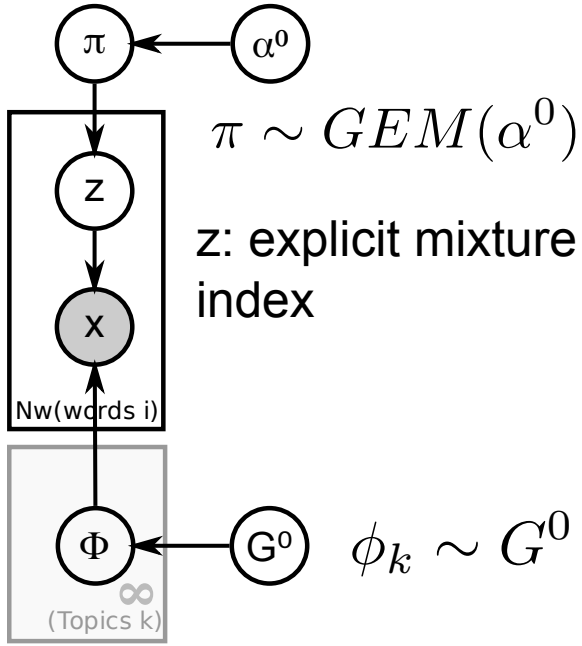
$$\pi \sim GEM(\alpha^0)$$

« stick-breaking process, chinese restaurant process »

$$\theta_i \sim G$$

$$x_i \sim f(\theta_i)$$

Compact form



$\pi \sim GEM(\alpha^0)$

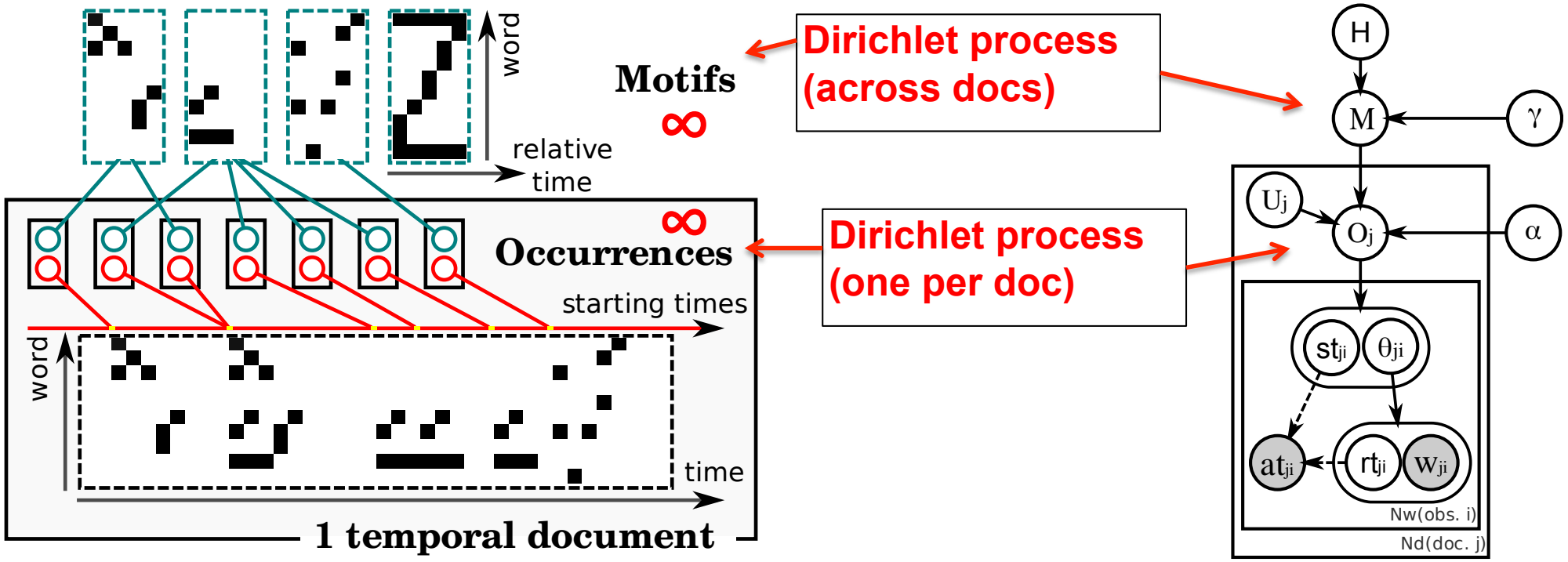
z: explicit mixture index

$\phi_k \sim G^0$

Expanded form

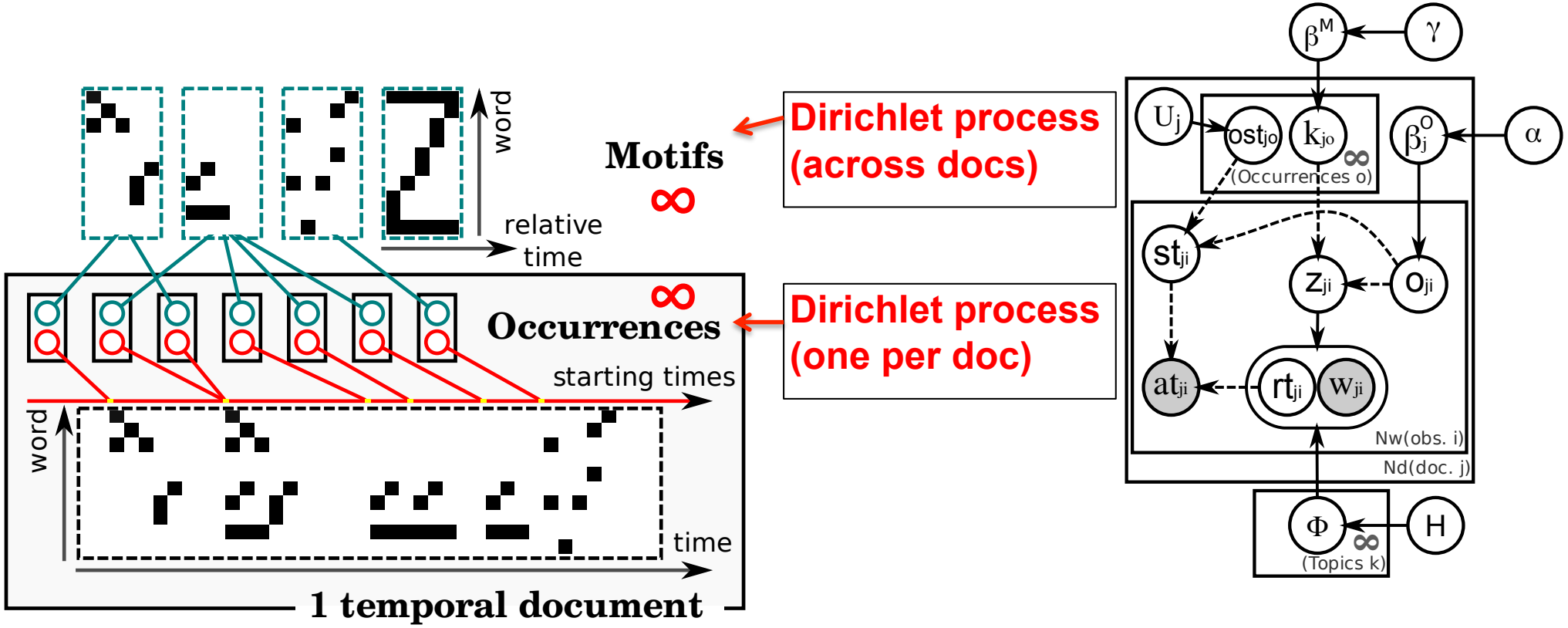
- Non-parametric approach to model arbitrary distributions over data points
- Defines a **prior distribution** over density distributions using **infinite mixture models** – e.g. Infinite Gaussian Mixture Models
- Two main 'parameters'
  - Base probability distribution G^o
  - Concentration parameter α^o (controls the mixture weights)

Dirichlet Processes for motif discovery



- Note: hierarchy introduced (motifs shared across occurrences)

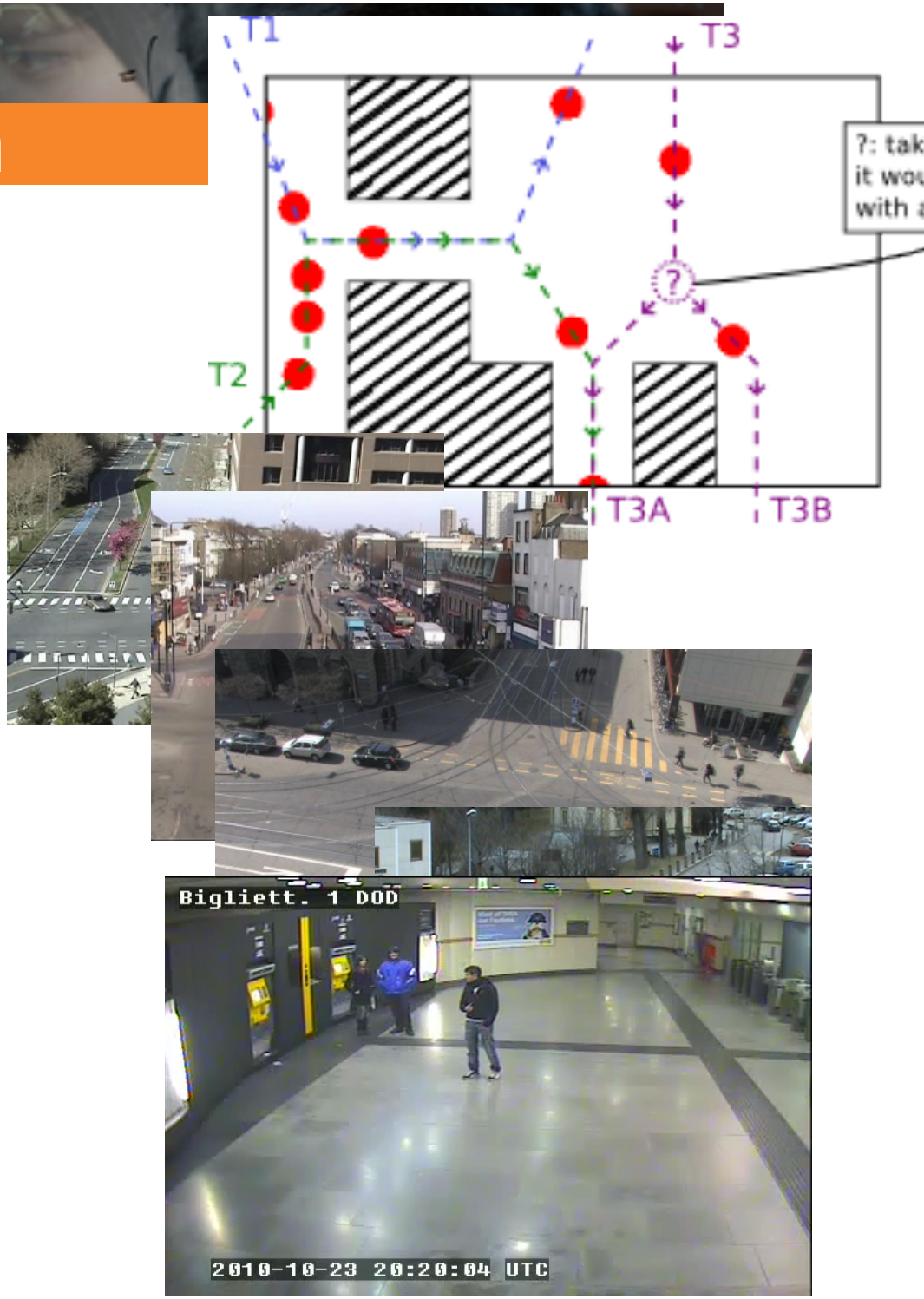
Dirichlet Processes for motif discovery



- Note: hierarchy introduced (motifs shared across occurrences)
- Solved with collapsed Gibbs sampling
  - Complexity proportional to amount of data
  - Occurrences: inherently sparse (compared to probability tables p(ts,z|d))

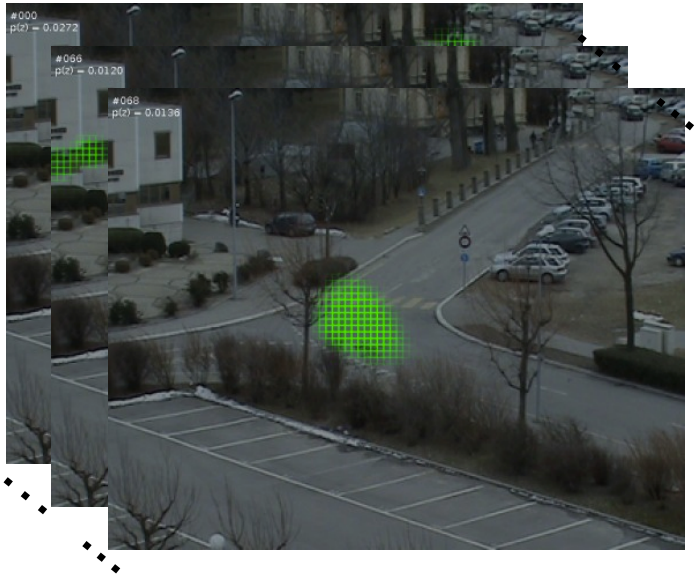
Results and evaluation

- Experiments with synthetic documents
  - Controlled setting, free ground truth
  - Test model behavior, draw curves
- Different datasets
  - Traffic: MIT, UQM, ETHZ, our data
  - VANAHEIM data
    - Loosely constrained behavior
    - Robust tracking impossible
    - Multi-camera handling
  - Audio data set
- Quantitative evaluation
  - Event precision recall
  - Prediction tasks



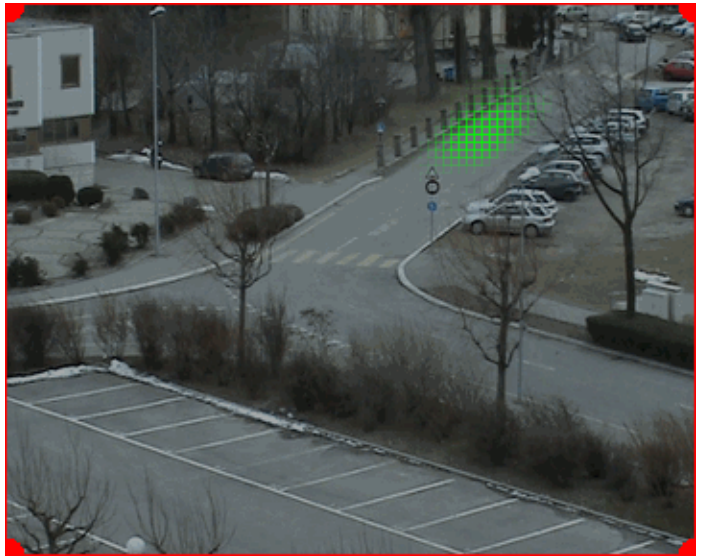
Motif representation and examples

Matrix

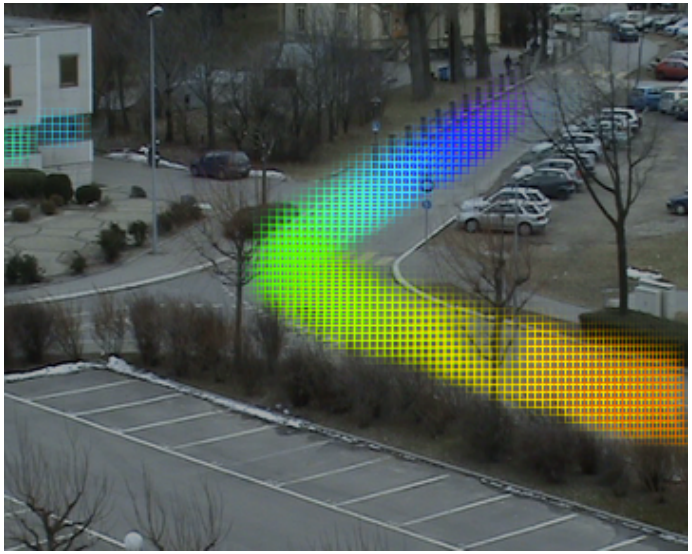


An image at each time step

(each column of the matrix)

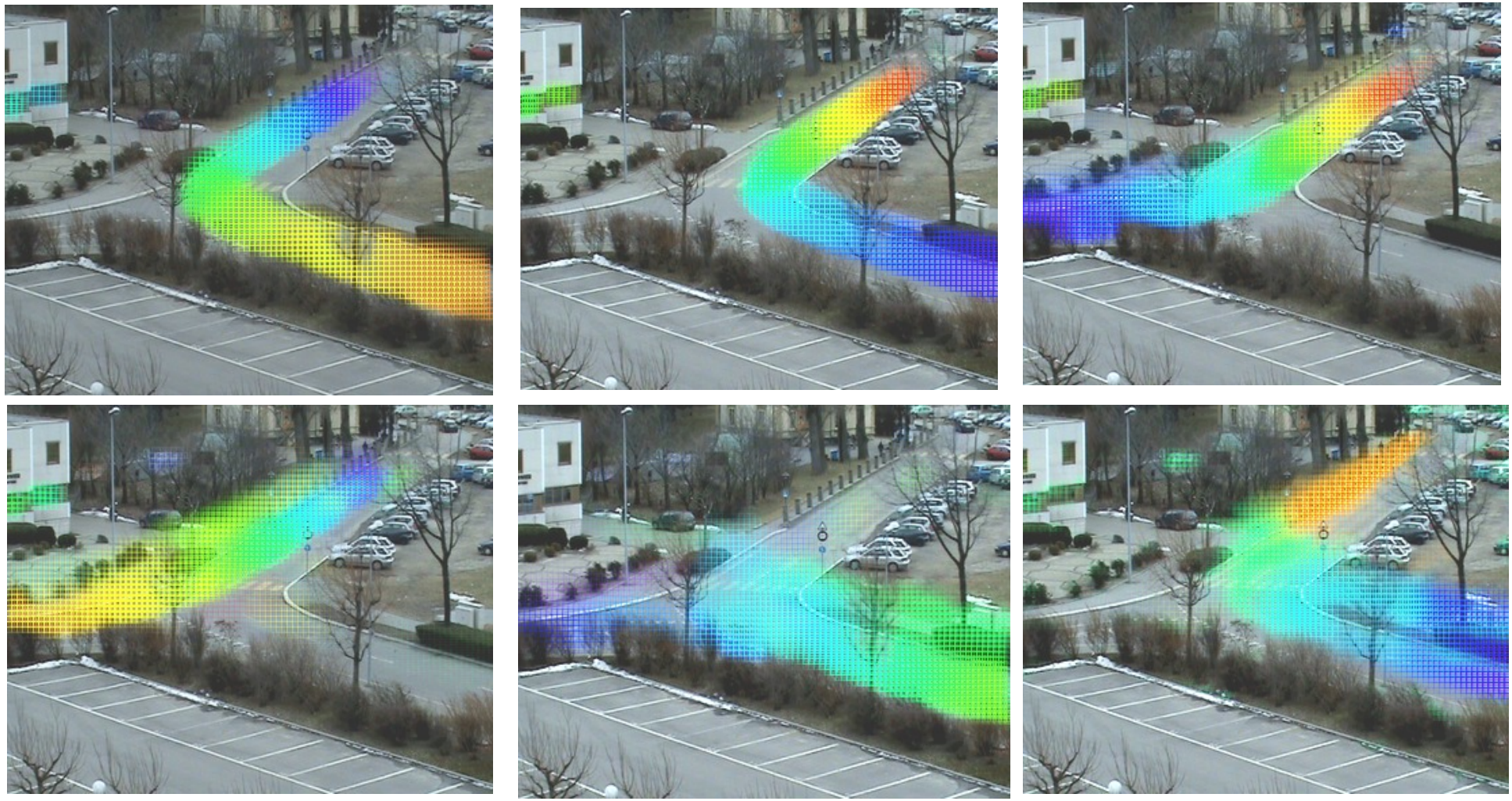


Animated GIF



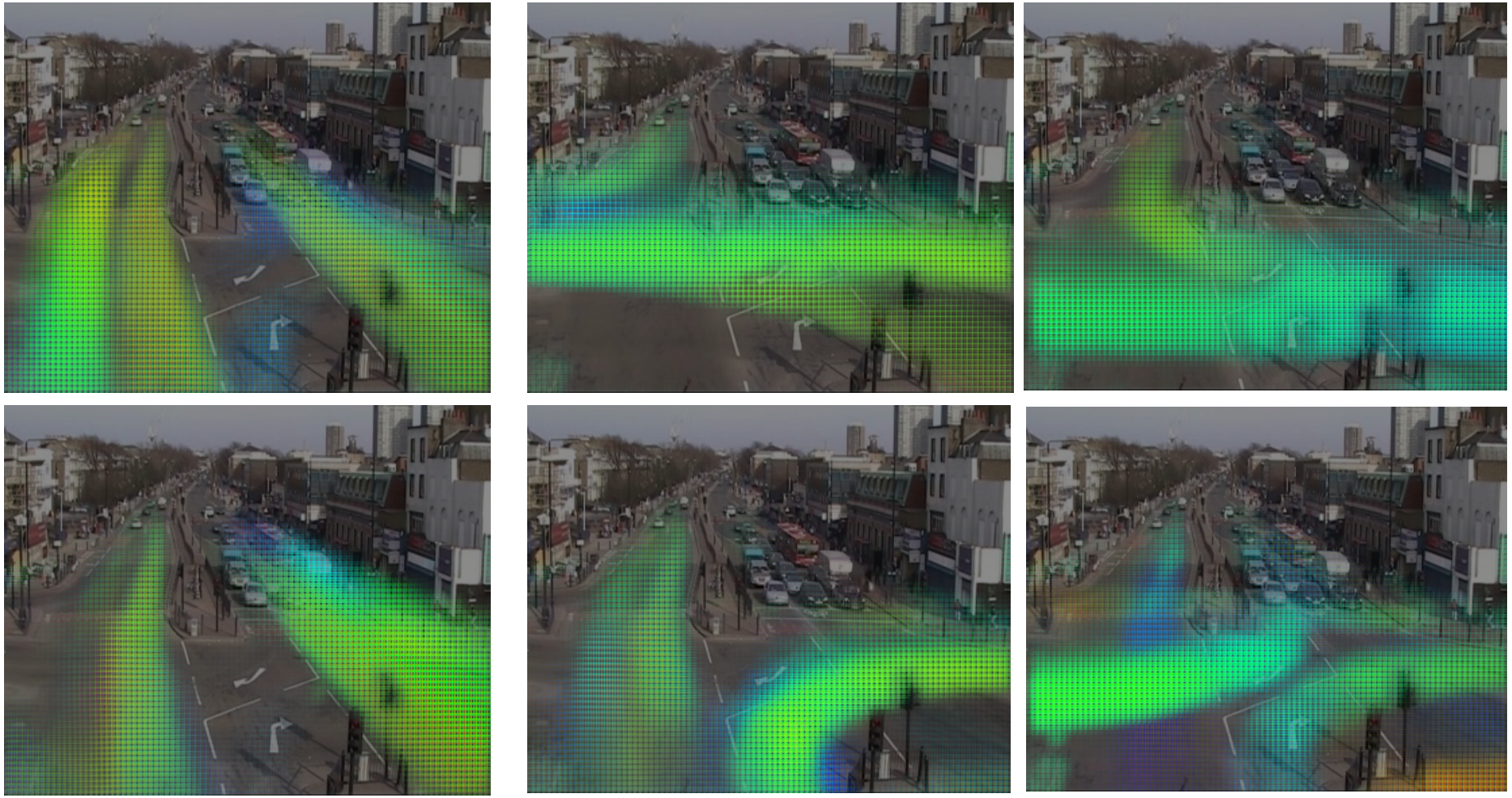
Color-coded dynamics

Result example 1 - Farfield



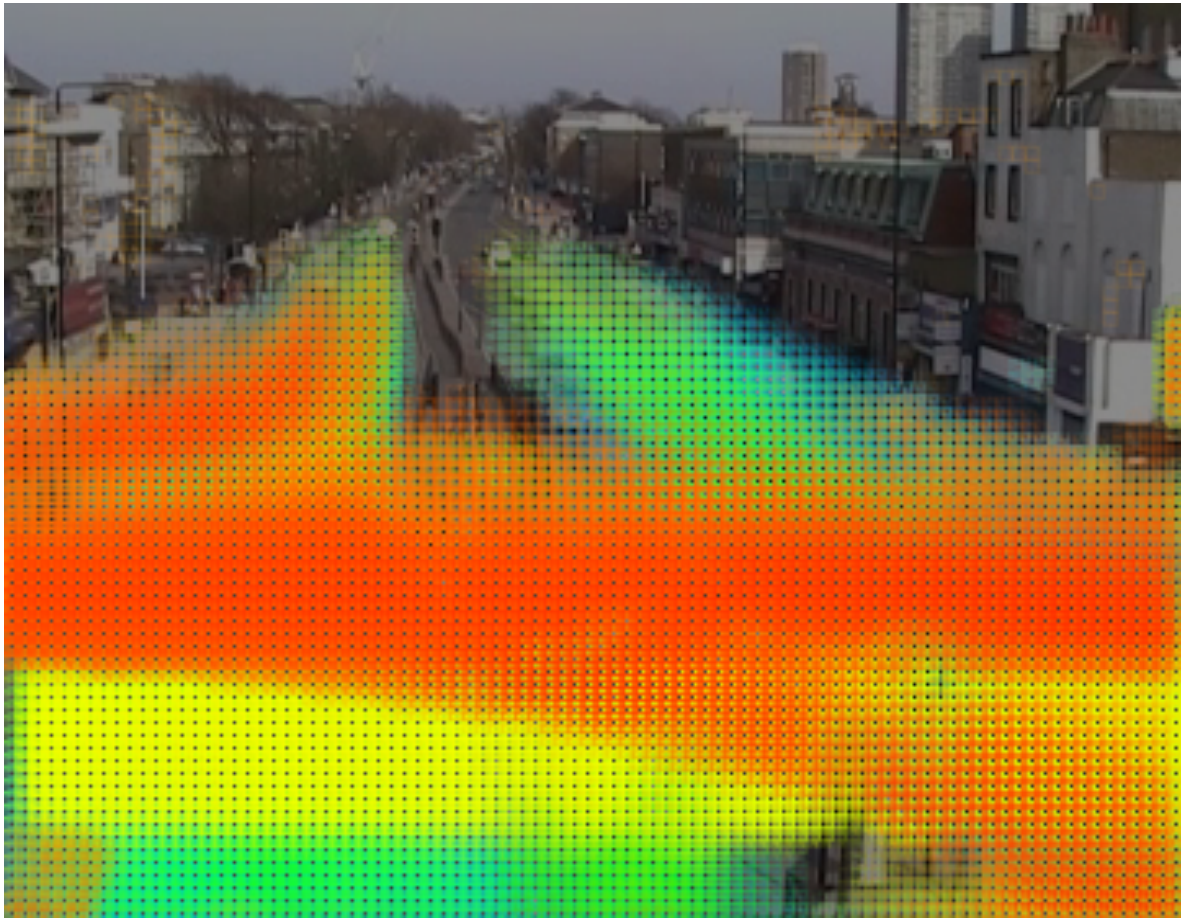
Top 6 topics, explaining more than 95% of the data
(looking for 12s max activities)

Example 2 – Traffic Junction (UQM)

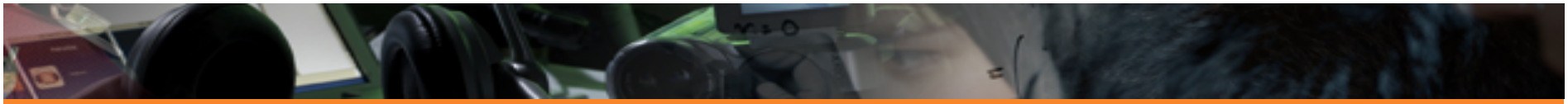


- Top 6 recovered topics, looking for 12s duration
  - lot of traffic => lot of co-occurring motion => topic cover more area
  - different traffic phases identifiable

Examples 2 - Traffic junction (UQM)



- Recovering 90s topics => automatically recovers only one motif



Example 3 - Metro

Montage 1
(almost no overlap)

Montage 2
(with overlap)

For each montage,
we compose a video from two cameras
and apply exactly the same approach.

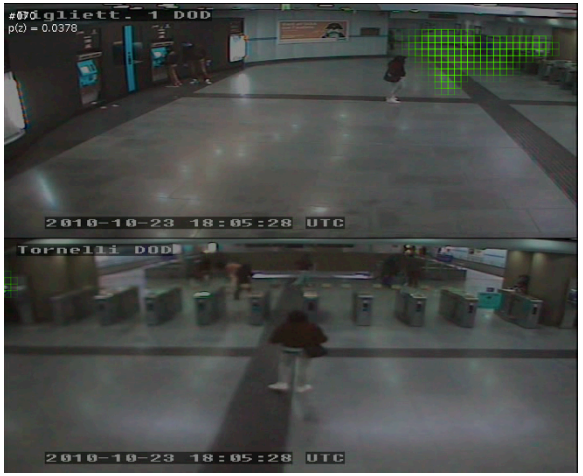
Example 3 - Metro

Montage 1

(almost no overlap)

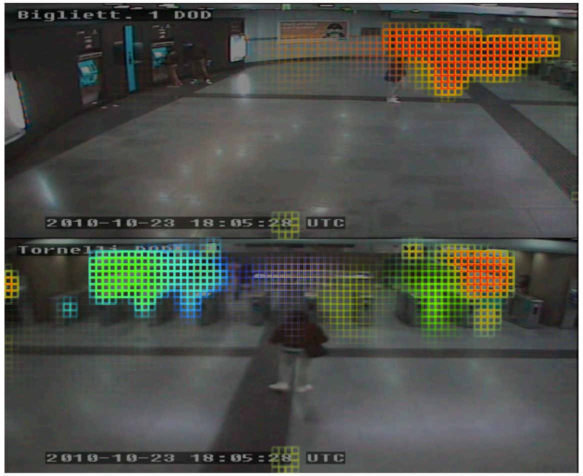
Montage 2

(with overlap)



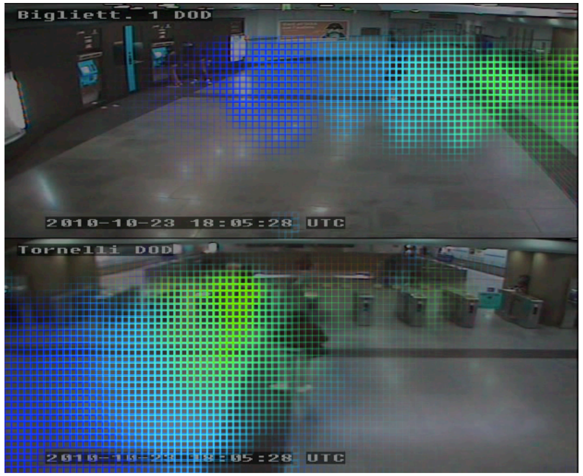
- Low-level topics: Approach does automatic soft-calibration

Recovered topics (montage 2)

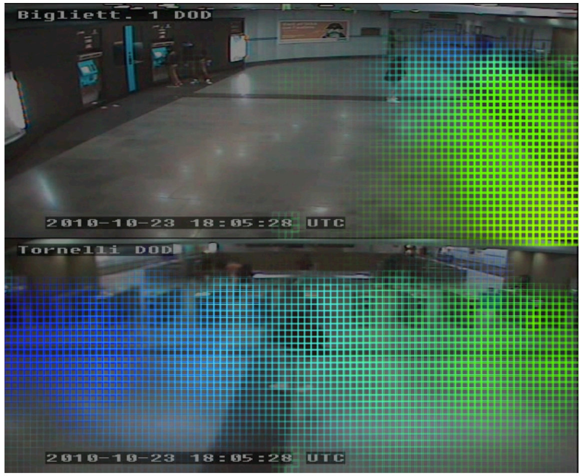


Leaving Station

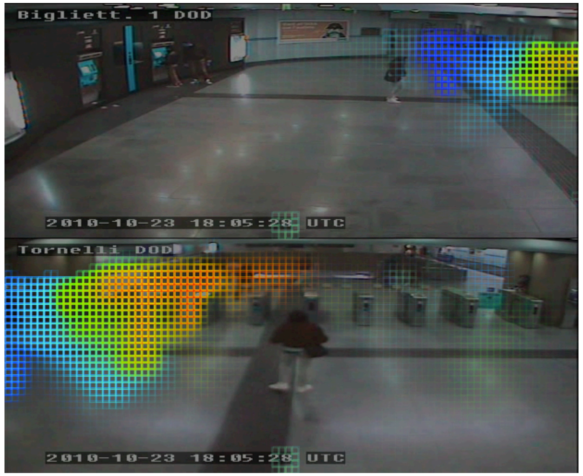
Entering Station (from the right)



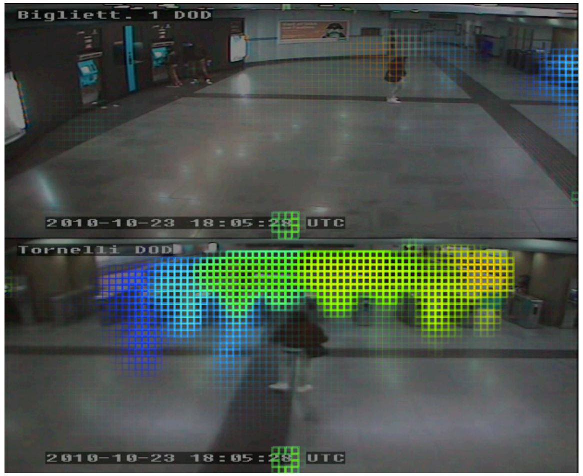
Vending Machine (leaving)



Just Crossing (left to right)



Entering Station (from the left)



Reaching Escalator and Elevator

Abnormality Rating

- Considered measures (at each time instant)
  - (log)likelihood
  - normalized (log)likelihood
  - reconstruction error
- **Reconstruction error**

$$abnorm(ta, d) = \sum_{w} \left| \frac{n(w, ta, d)}{n(d)} - p(w, ta|d) \right|$$

each time instant

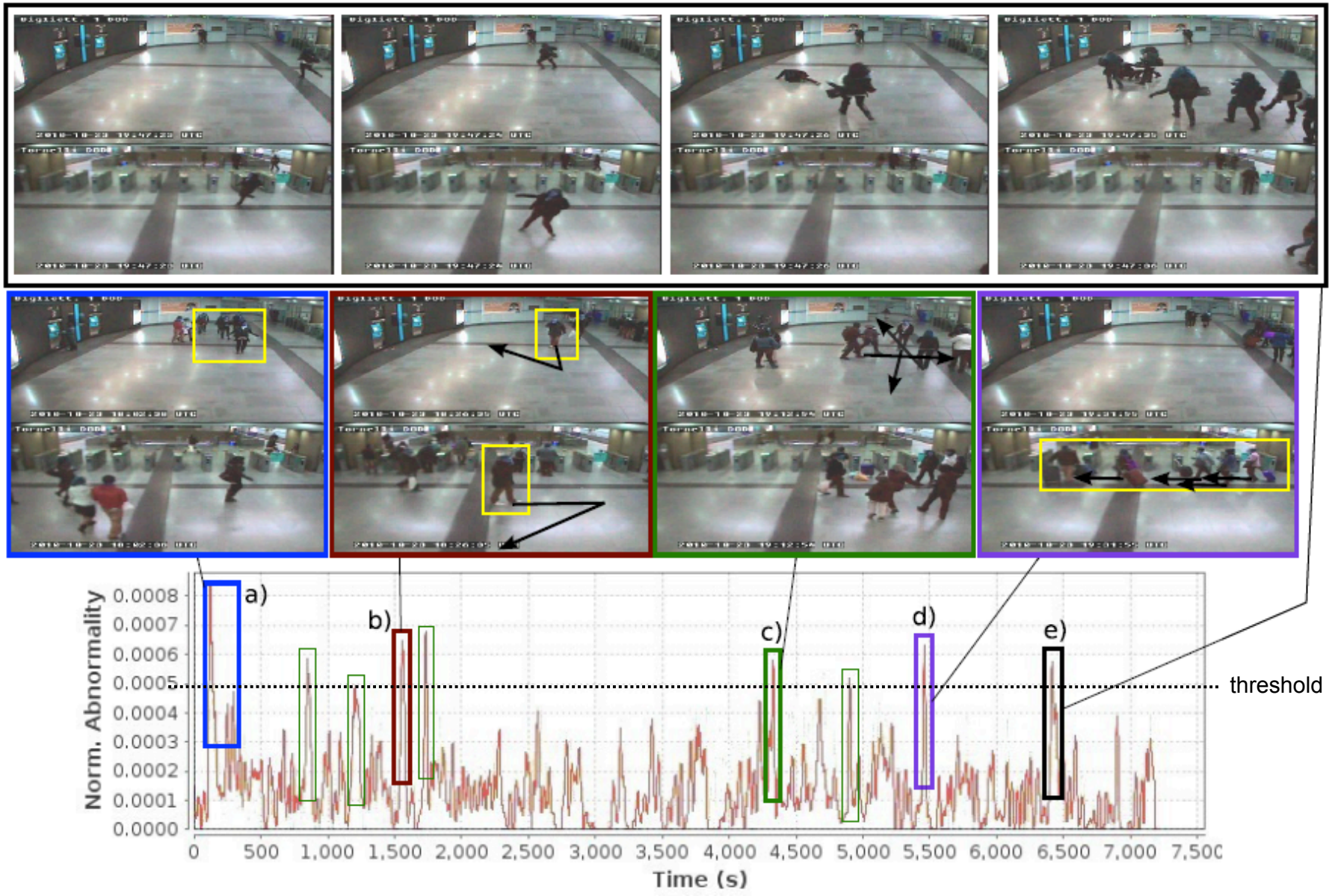
each document

sum over the vocabulary

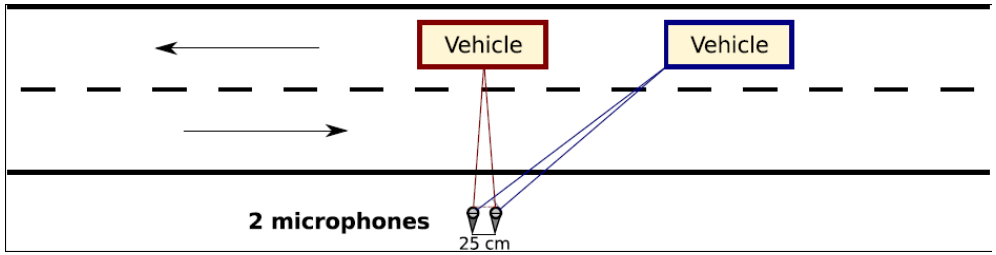
"Empirical Probability"

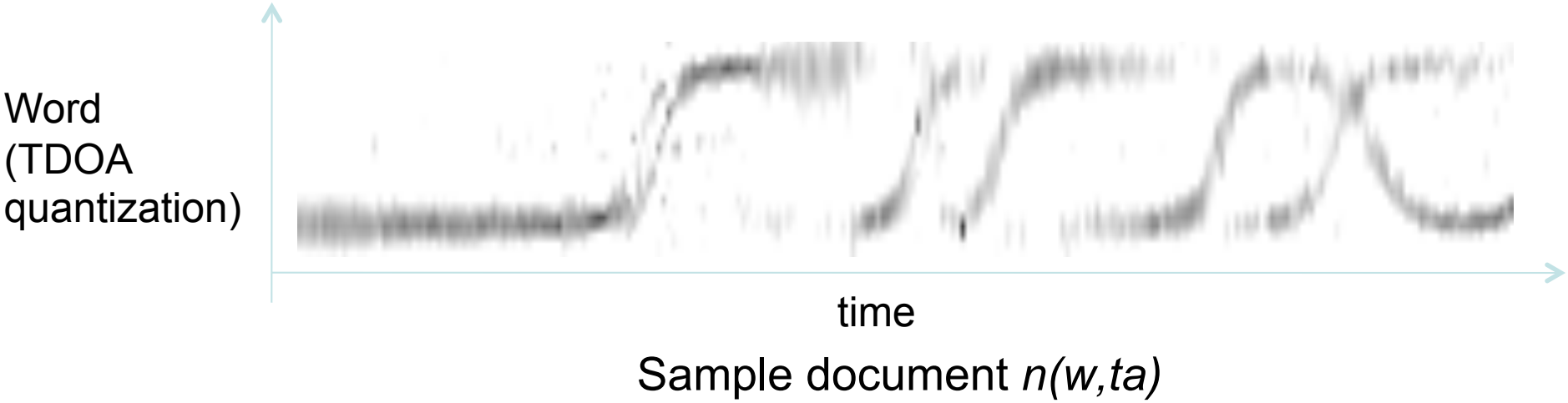
Reconstructed probability (from model fitting)

Example Abnormality Rating (Montage 2)



Mining audio data

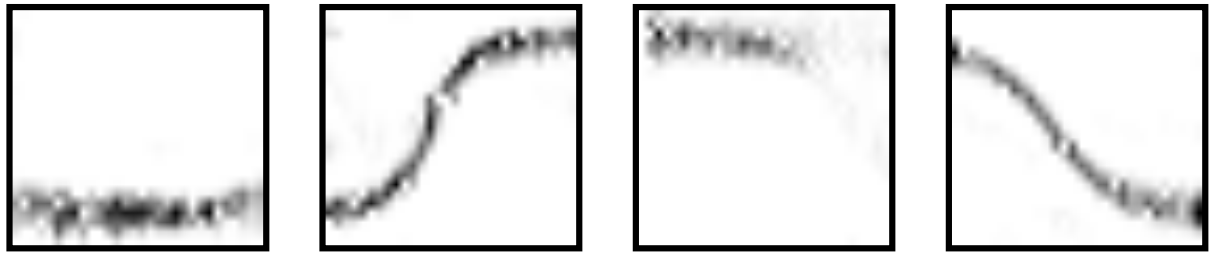




- TDOA features can be used as words

Resulting motifs

- Automatically recovers 4 motifs
  - 30 time steps

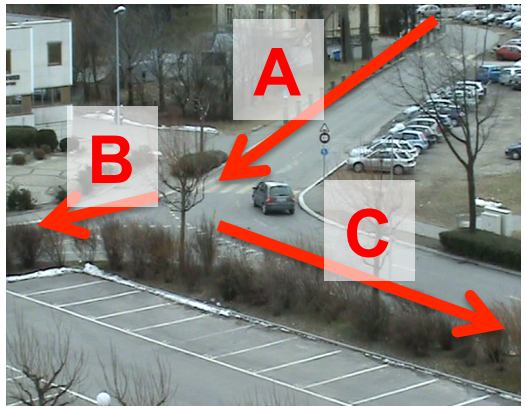


  - 60 time steps

    - Similar shapes

Conclusion, perspectives

- Temporal motif discovery model
  - Topic with explicit word temporal order, not only co-occurrence
  - Handles concurrent activities
  - Generic method: can be applid to any (word x time) count matrix
  - Fully automatic: from videos to (numer of) activities and abnormality
- Current work and Perspective
  - Apply model to Audio (marray)-Video => enhanced Scene model
  - Currently, no temporal modeling of activitiy starting times p(ts,z|d)
    - limitation, e.g. when dependencies exists
      A is followed by either B or C
      => pairwise analyses, or symbolic analysis
    - when scene goes through specific phases
      some activities are more likely to occur
      => HDP-HMM approach

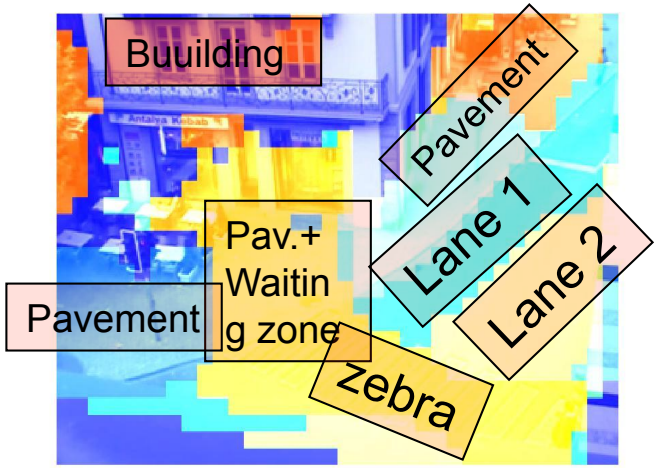


Perspectives

- Stream selection task (EU Vanaheim)
  - Unsupervised and supervised approaches (using little operator feedback)

- Human activity recognition
  - DGA french funding
  - Use of Spatio-Temporal-Interest points (STIP) as descriptor for words
  - Motifs define super-word features

- Urban scene semantic labeling (Thales funding)
  - Automatic scene segmentation + semantic labels
  - Use dynamics features (learned activities) + visual features and object detection responses
  - Apply known semantic rules into structured scenes for abnormal event detection



Thank you for your attention.

Questions ?