Activity Analysis Using Topic Models

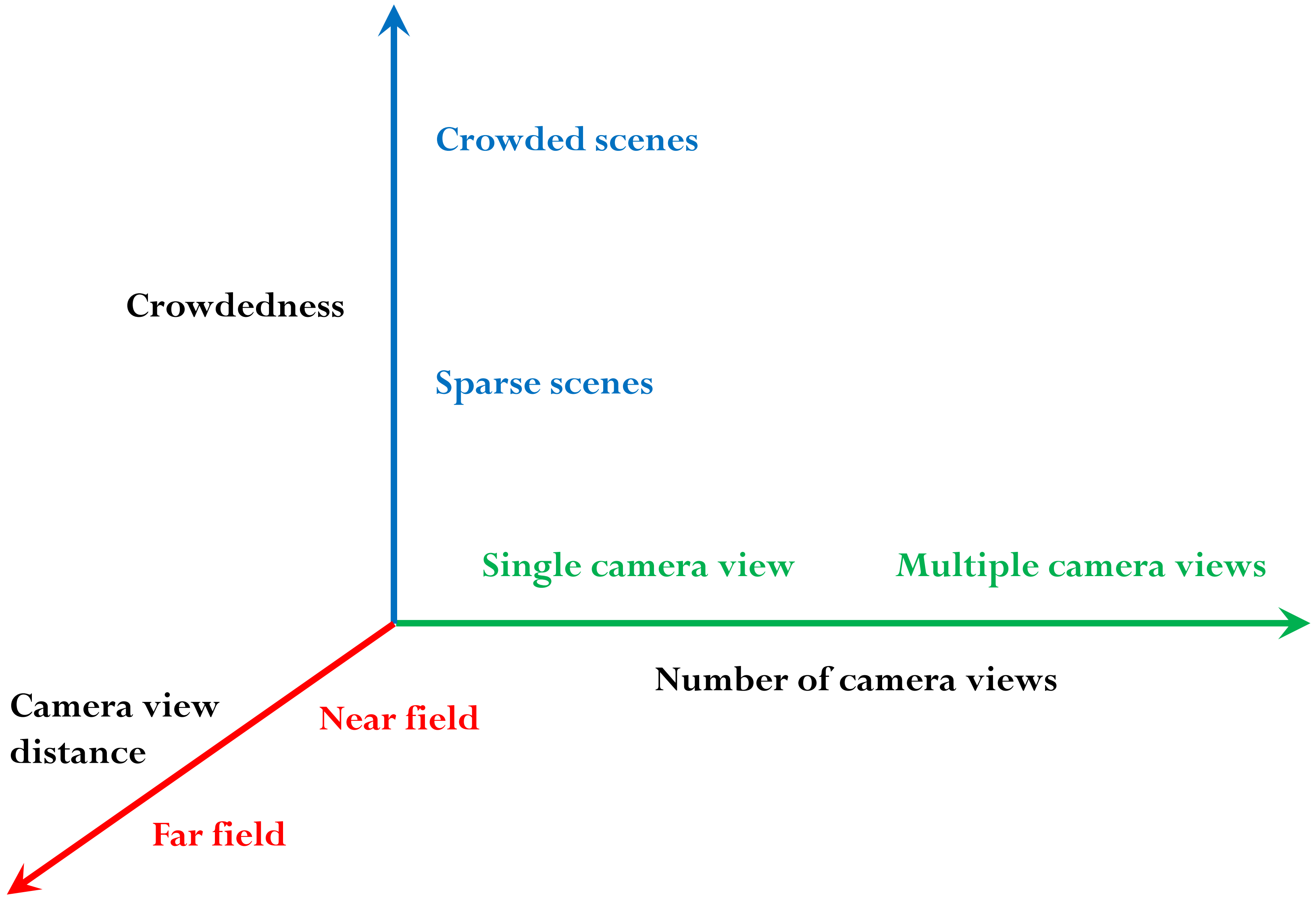
Xiaogang Wang
Department of Electronic Engineering
The Chinese University of Hong Kong

November 07, 2011

Outline

- **Introduction**
- Topic Models
- Activity analysis in sparse scenes
  - Based on trajectories
  - Single camera view
  - Multiple camera views
- Activity analysis in crowded scenes
  - Based on tracklets
  - Based on local motions
- Activity analysis in near fields

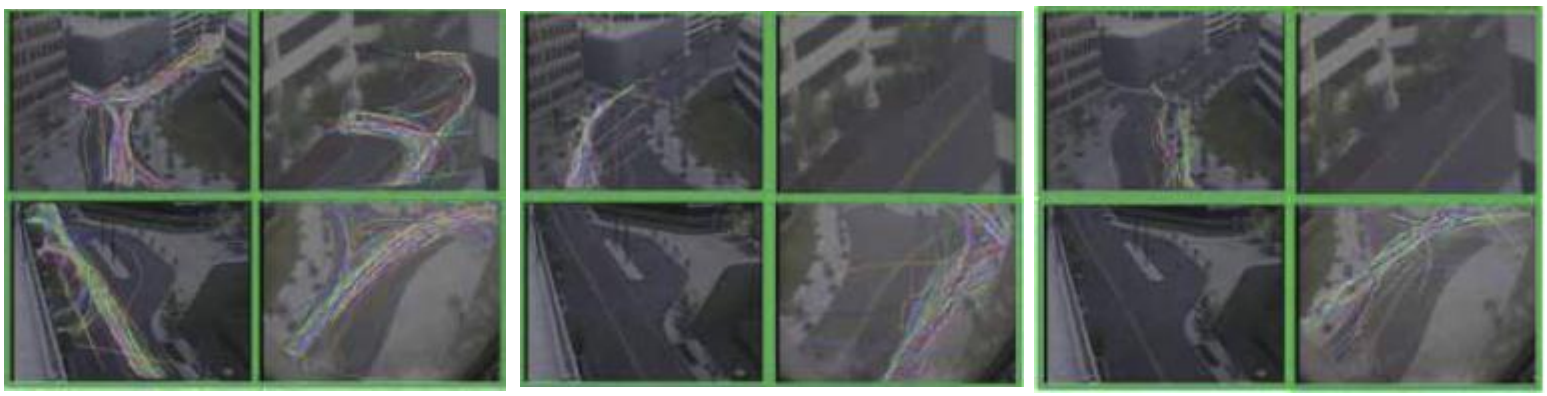
Activity analysis under different scenarios



Sparse scenes in a single camera view

- ✓ Objects can be detected and tracked
- ✓ Trajectories of objects are used for activity analysis
- ✓ Cluster trajectories into different activity categories
- ✓ Detect abnormal trajectories

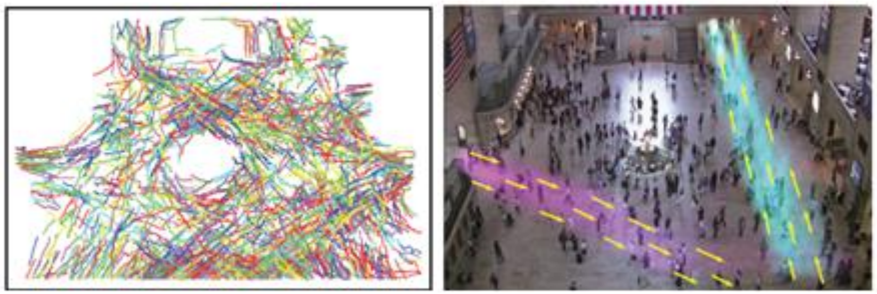
Camera 1 Camera 2 Camera 3 Camera 4 Topology



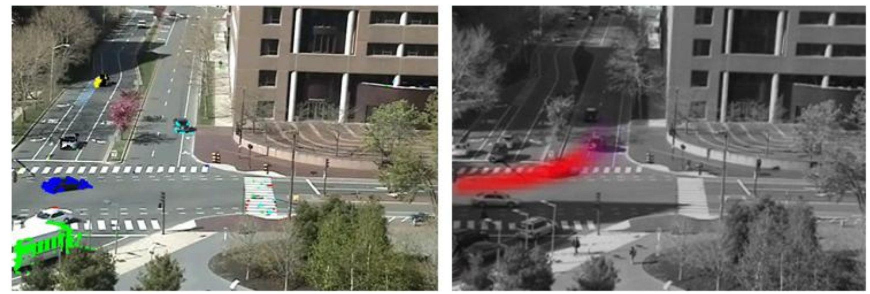
Activity 1 Activity 2 Activity 3

Sparse scenes in multiple camera views

- ✓ Objects can be detected and tracked in each of the camera views
- ✓ It is challenging to track objects across camera views
- ✓ The topology of camera views may be arbitrary and unknown
- ✓ Cluster trajectories observed in different camera views **without tracking objects across camera views** and **without knowing the topology of camera views**



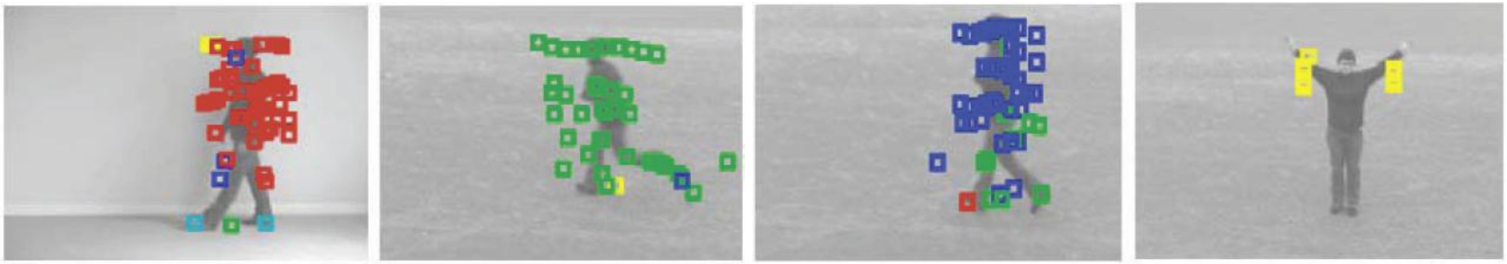
Tracklets



Local motions

Crowded Scenes

- ✓ It is challenging to detect and track objects in crowded environments
- ✓ Many different types of activities happen simultaneously in crowded scenes
- ✓ Learn the models of activities from **tracklets** (highly fragmented trajectories) or **local motions** (optical flows)



From Niebles et al. BMVC'06

Near Fields

- ✓ In far fields, objects are small in size and their activities are mainly distinguished by their positions and velocities
- ✓ In near fields, objects are in larger sizes, and more features such as shape, appearance and motions can be used for activity analysis
- ✓ Use space-time interest points as features for activity analysis

Why topic models?

- Unsupervised
  - Save labeling effort
  - Suitable for processing large scale datasets
  - Easy to transfer across different scenes
- Topic models are hierarchical Bayesian models
  - Model complex activities in a principled way
    - Jointly model simple activities and complex activities at different hierarchical levels
  - Add priors to hierarchical Bayesian models
    - Dynamically update the models of activities over time
    - Learn the models of activities across camera views

Why topic models?

- Can be extended to nonparametric Bayesian models
  - Automatically learn the number of activity categories driven by data
- Model the co-occurrence of motion features
  - Co-occurrence of motion features widely exists in many types of activities
  - No other strong constraints on the distributions of activity models
  - Can be well applied to different scenes and different types of activities

Outline

- Introduction
- **Topic Models**
- Activity analysis in sparse scenes
  - Based on trajectories
  - Single camera view
  - Multiple camera views
- Activity analysis in crowded scenes
  - Based on tracklets
  - Based on local motions
- Activity analysis in near fields

Topic models

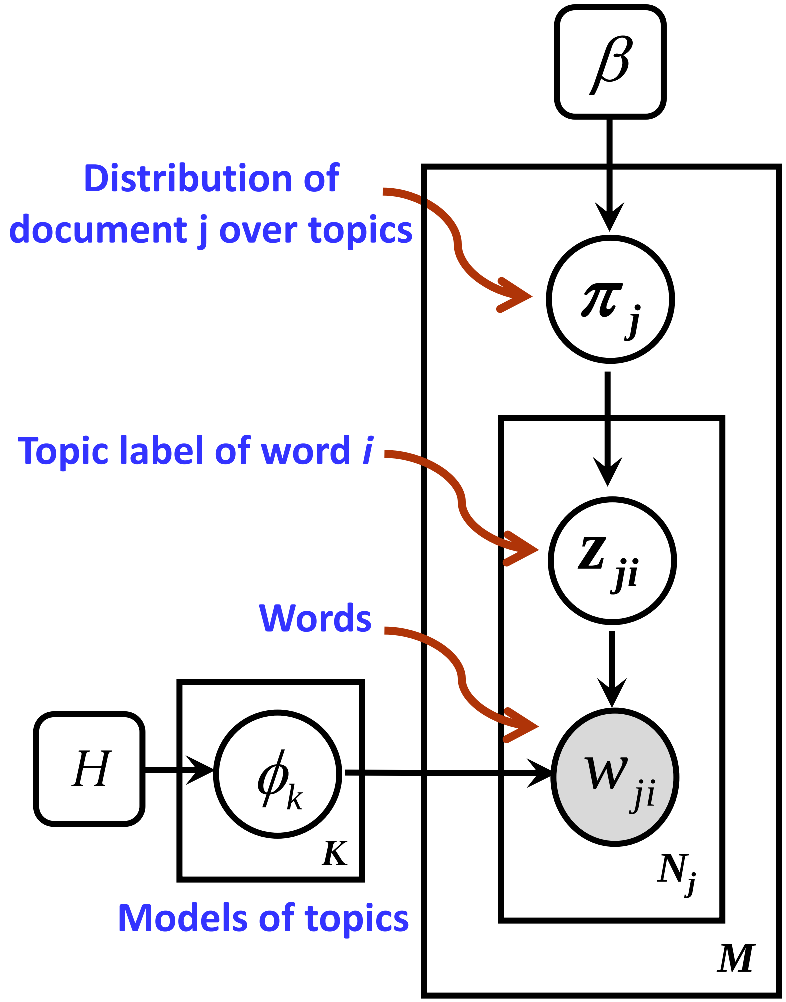
"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

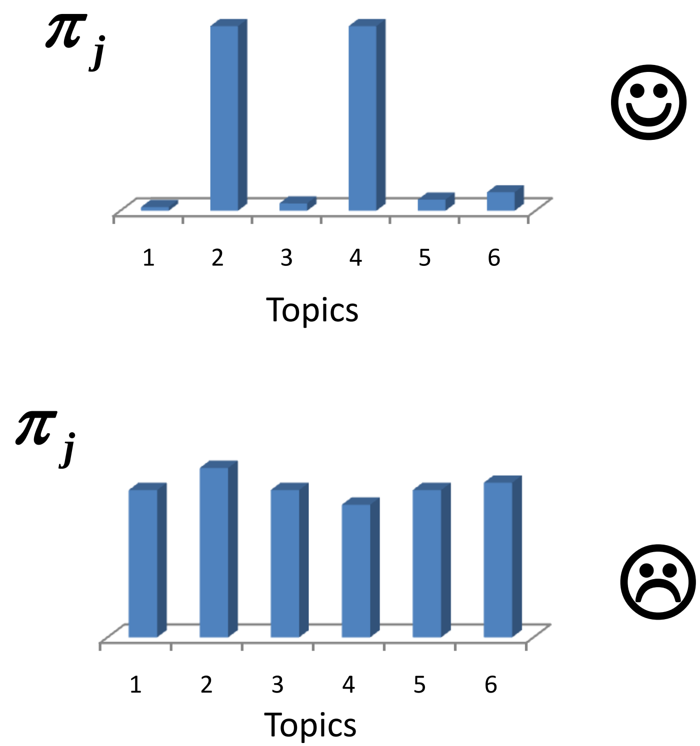
The William Randolph Hearst Foundation will give $1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be $200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive $400,000 each. The Juilliard School, where music and the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.

Figure 8: An example article from the AP corpus. Each color codes a different factor from which the word is putatively generated.

From Blei, Journal of Machine Learning Research, 2003

Latent Dirichlet Allocation





Applying topic models to activity analysis

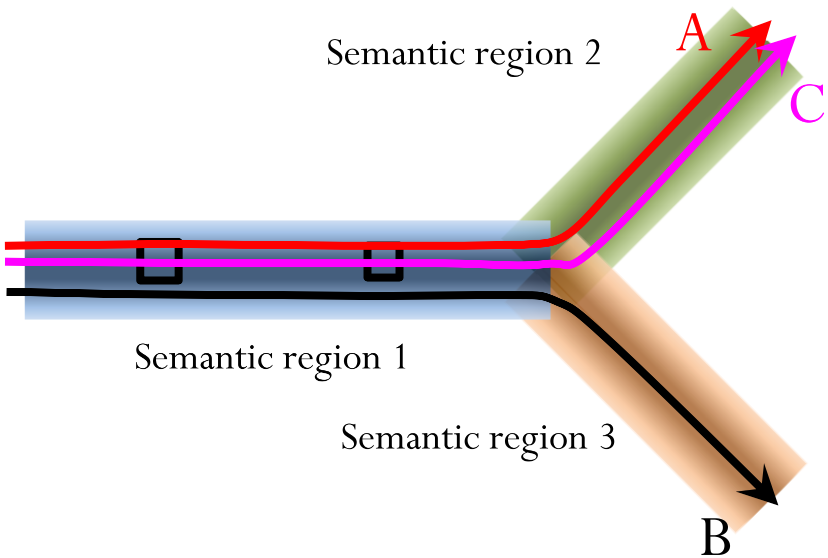
- Features = ?
- Words = ?
- Documents = ?
- Topics = ?
- How to extend the models by adding priors which capture the spatial and temporal information?

Outline

- Introduction
- Topic Models
- **Activity analysis in sparse scenes**
  - **Based on trajectories**
  - **Single camera view**
  - **Multiple camera views**
- Activity analysis in crowded scenes
  - Based on tracklets
  - Based on local motions
- Activity analysis in near fields

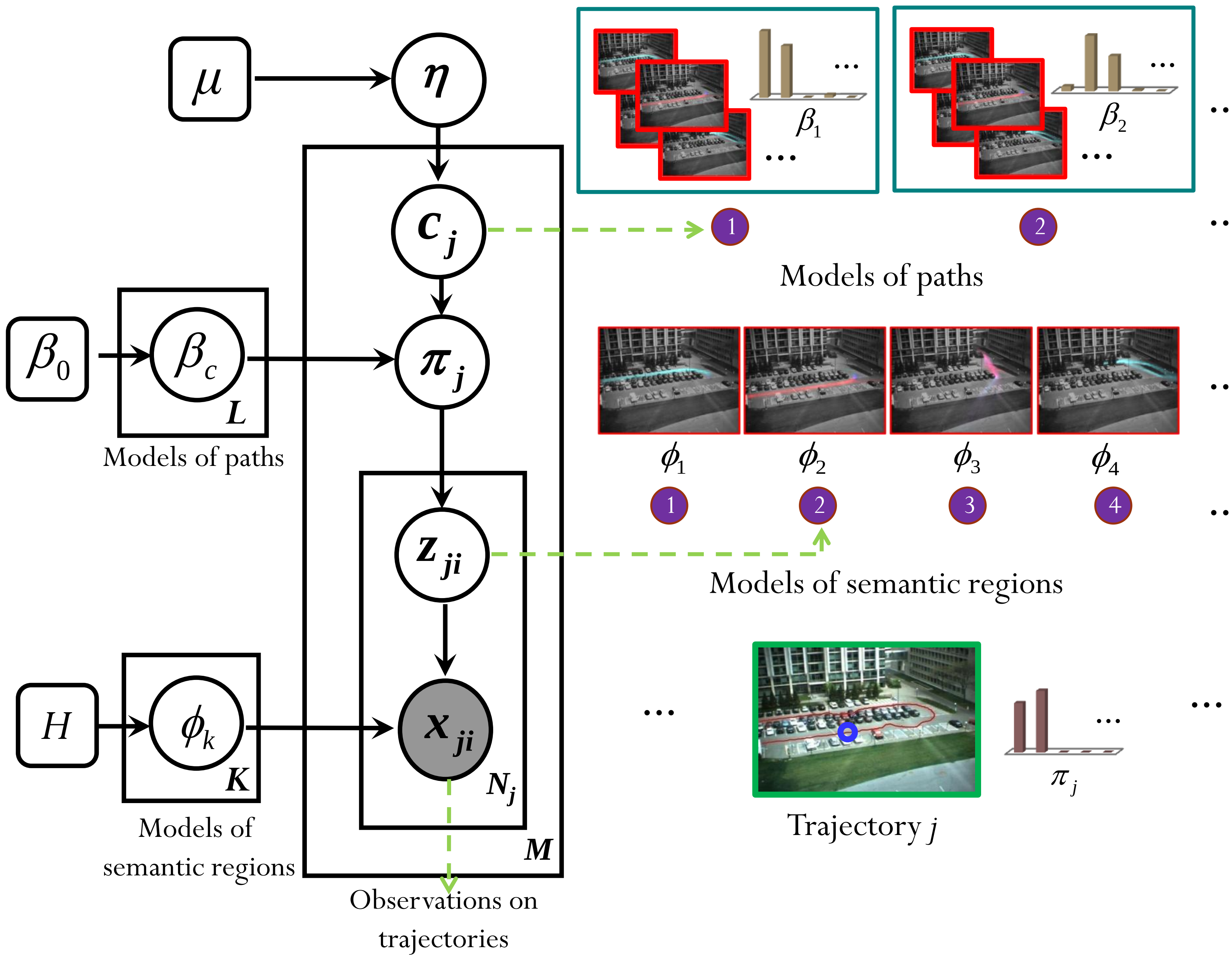
Activity analysis in sparse scenes

- Features = positions and velocities of points on trajectories
- Words = points on trajectories
- Documents = trajectories
- Topics = semantic regions (intersections of paths commonly taken by objects)
- Identity co-occurrence : two feature values are observed on the same trajectory and they are related to the same object

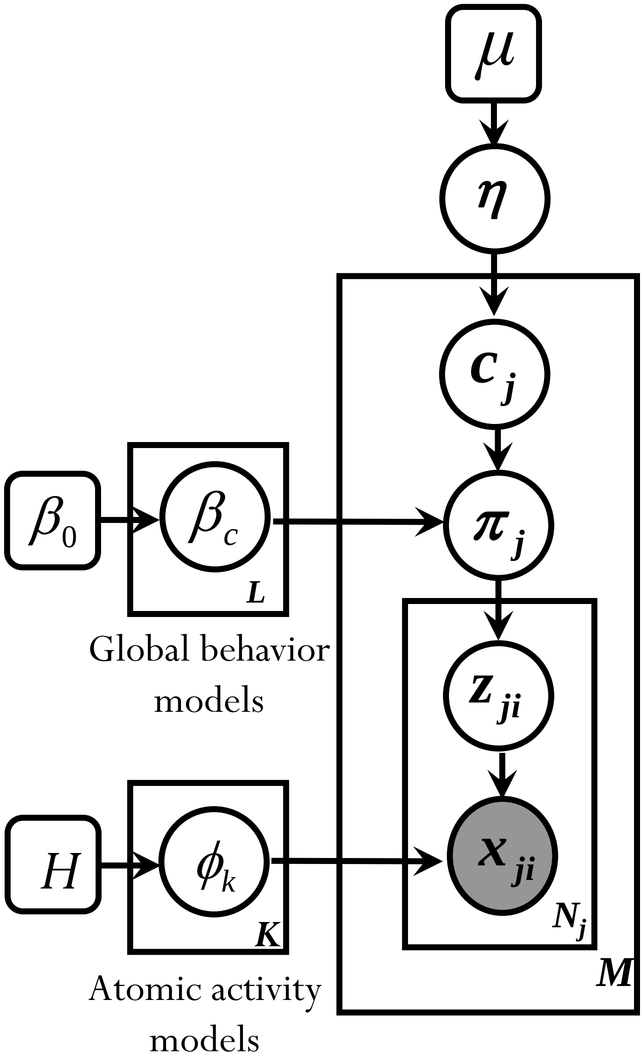


X. Wang, K. Ma, G. Ng, and E. Grimson, "Trajectory Analysis and Semantic Region Modeling Using A Nonparametric Bayesian Model," CVPR'08

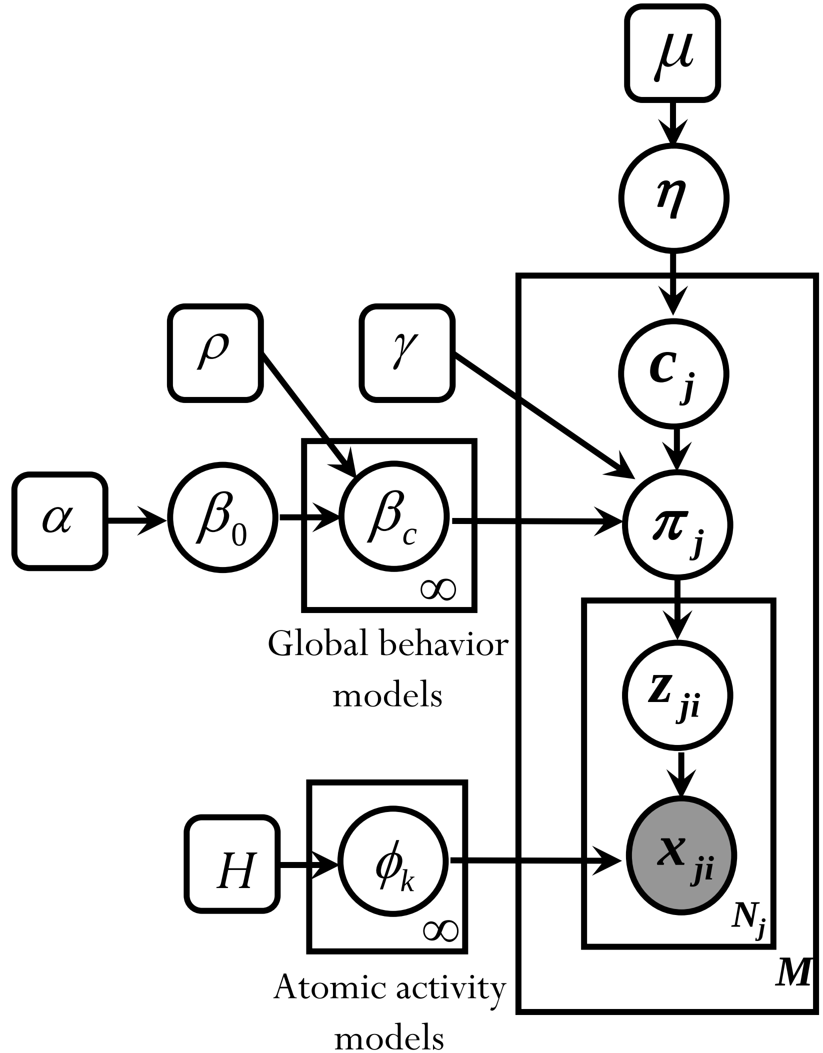
Two-level parametric topic model



Dual Hierarchical Dirichlet Processes (Dual-HDP)



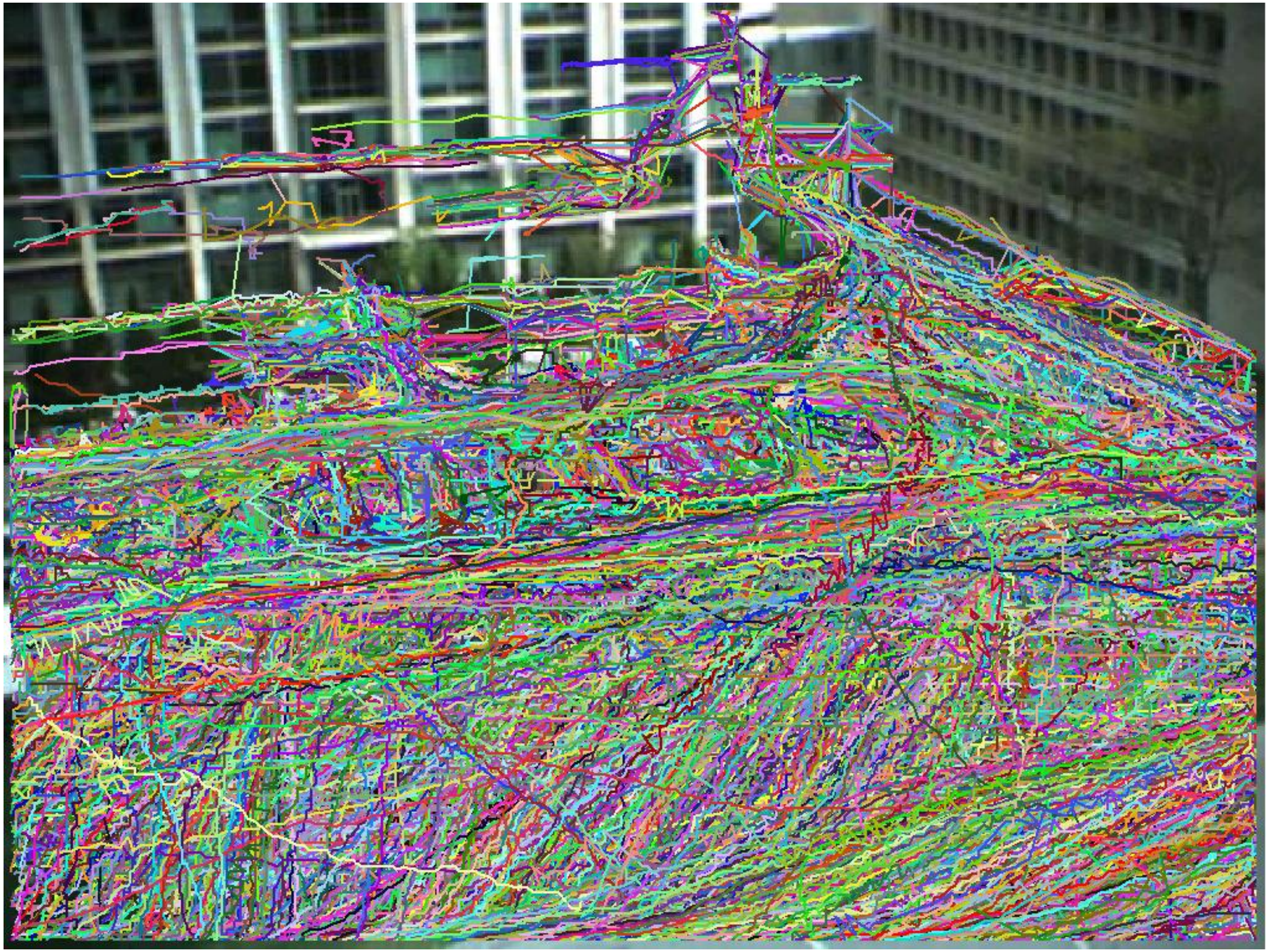
Parametric hierarchical Bayesian model



Dual-HDP

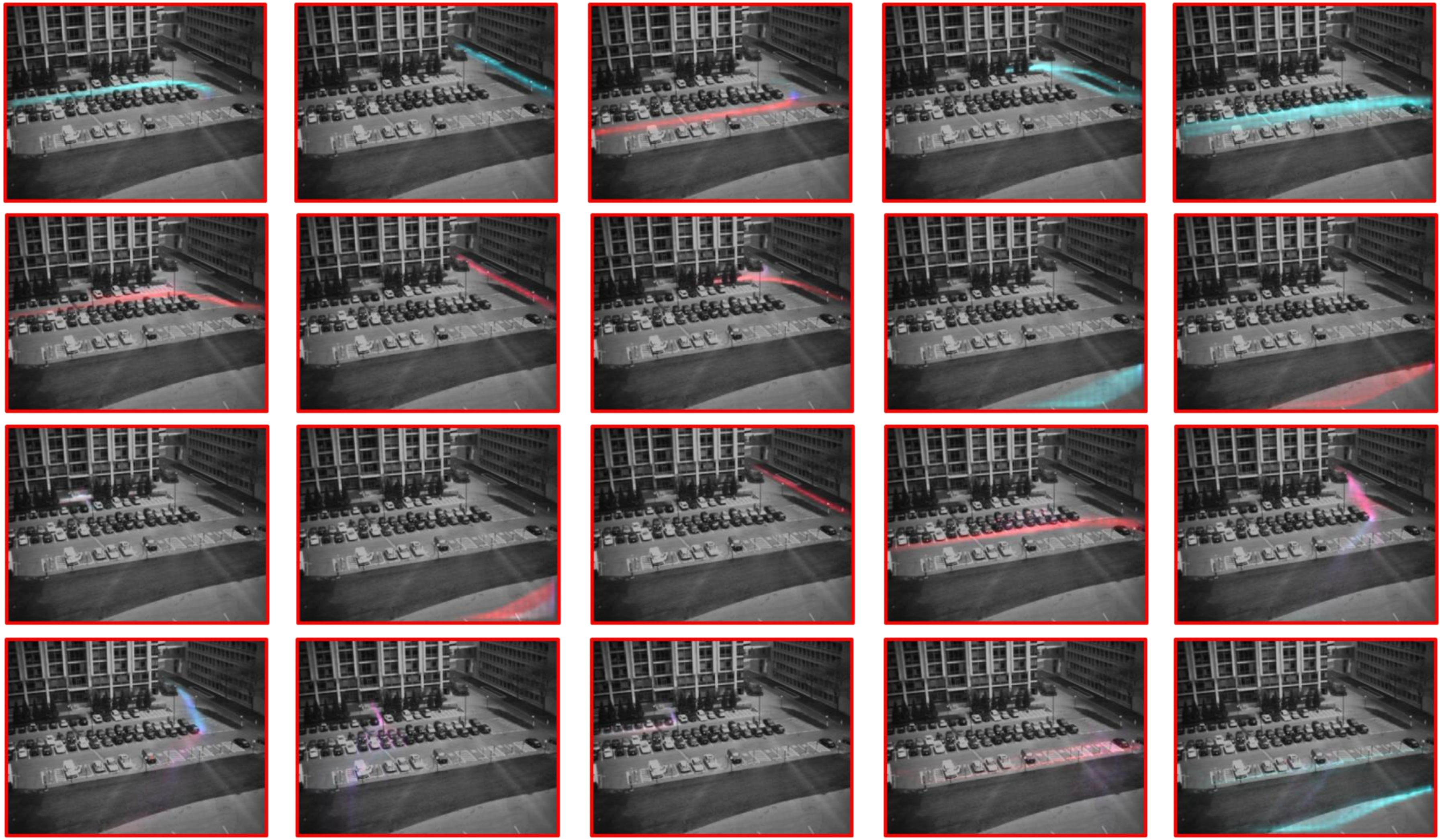
Related to the nested HDP [Rodriguez et al. 2006], Transformed HDP [Sudderth et. al IJCV'07], HDP [Teh JASA'04]

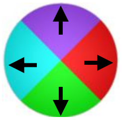
Cluster trajectories and learn models of paths



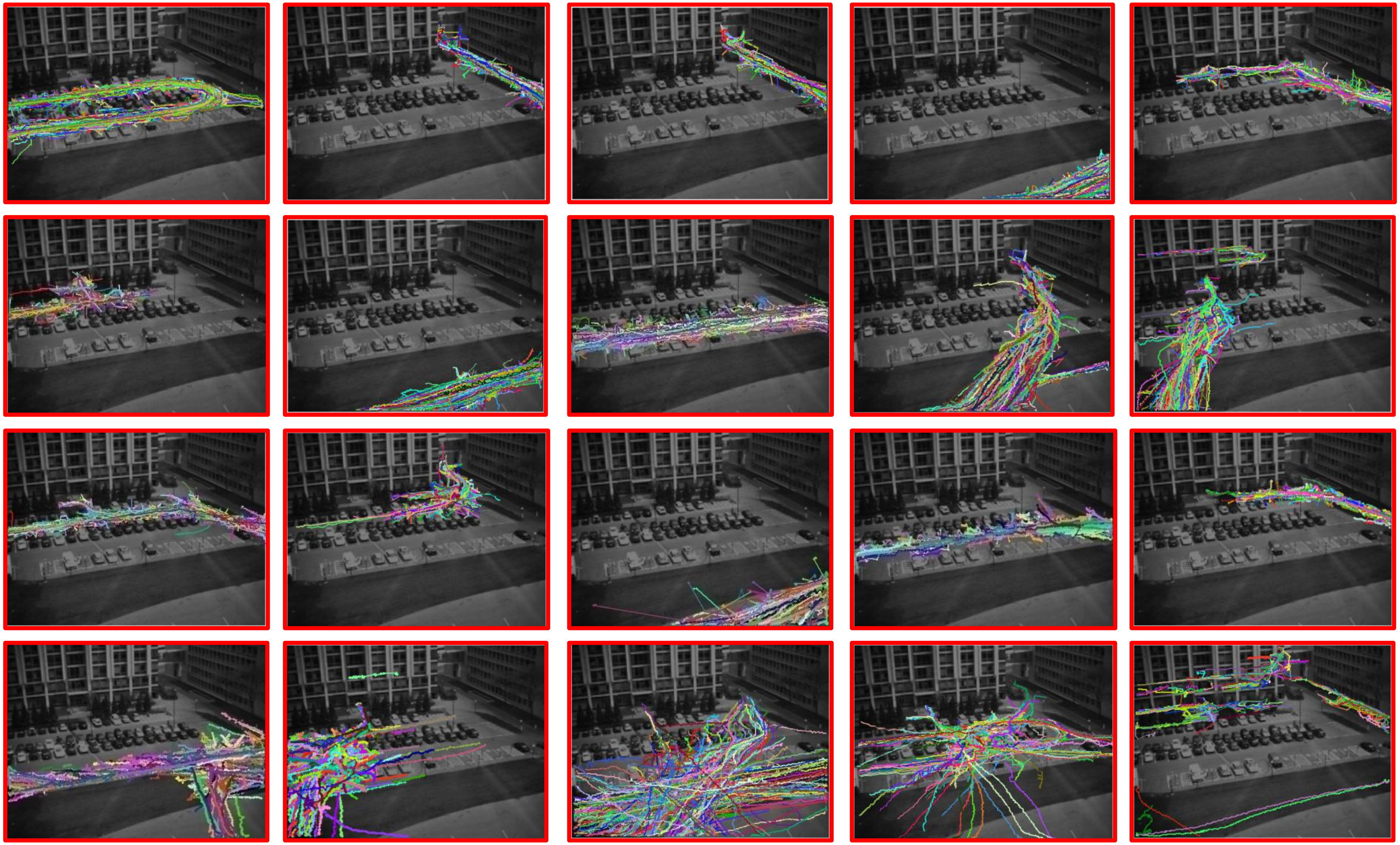
40,453 trajectories for our experiments

Models of semantic regions





Clusters of trajectories



Outlier trajectories


Top 1-20

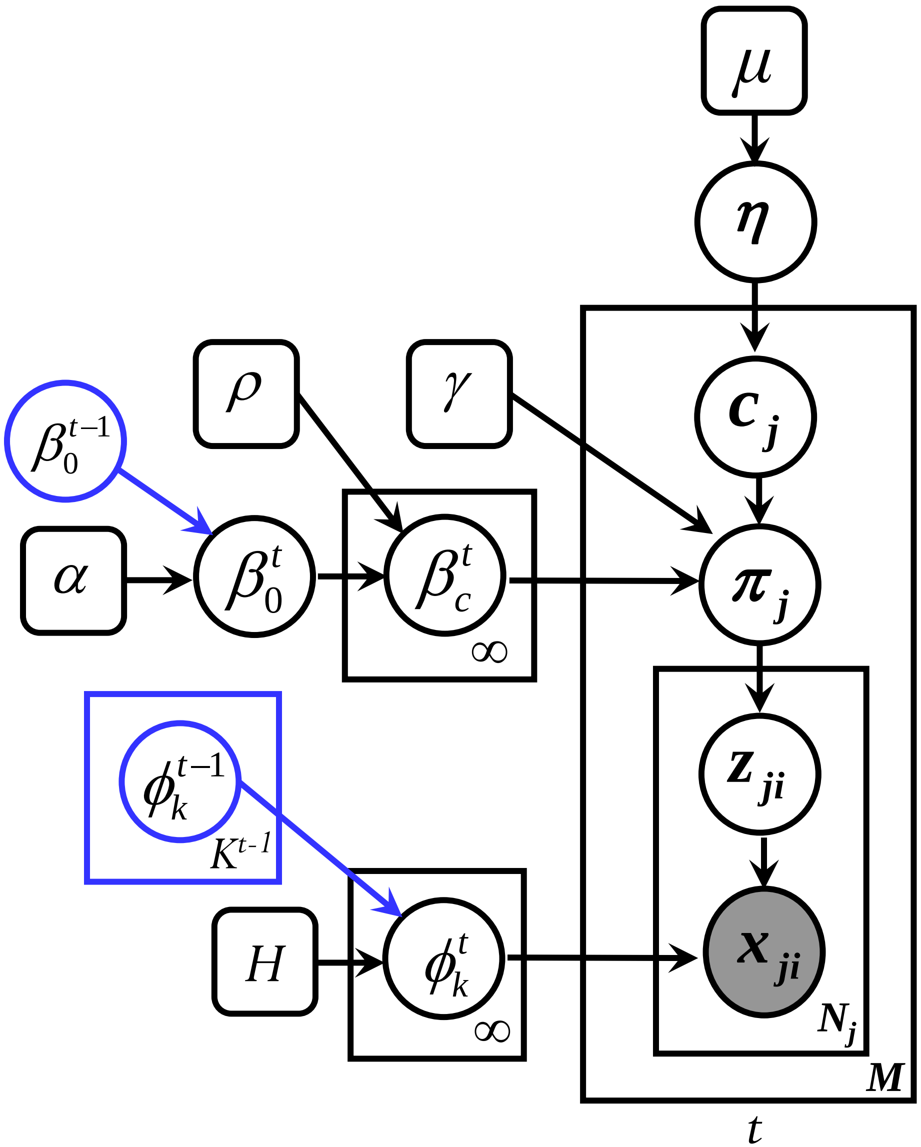

Top 21-40

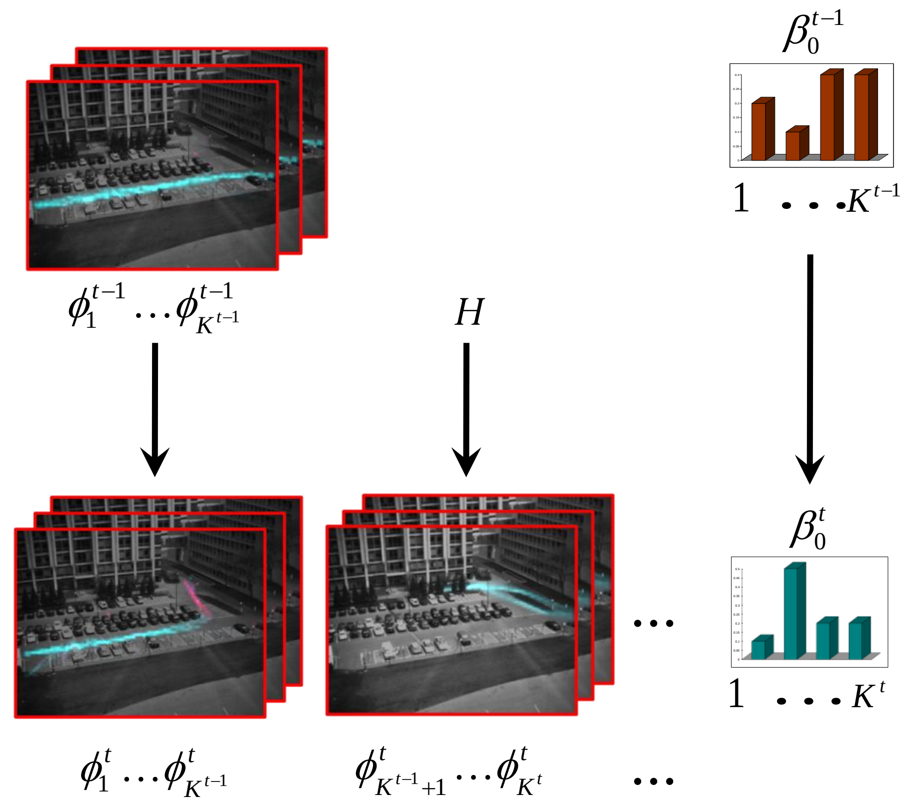

Top 41-60


Top 61-80


Top 81-100

Dynamic Dual-HDP





- Models are dynamically updated
- The information of data before t is included in ϕ_k^{t-1} and β_0^{t-1}

X. Wang, K. Ma, G. Ng, and E. Grimson, "Trajectory Analysis and Semantic Region Modeling Using Nonparametric Bayesian Models," IJCV'11

Dynamic models of semantic regions in a parking lot

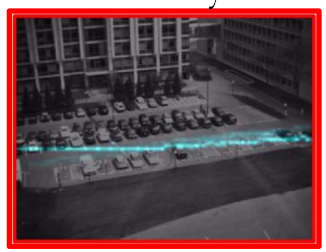

2am-3am May 15


7am-8am May 15

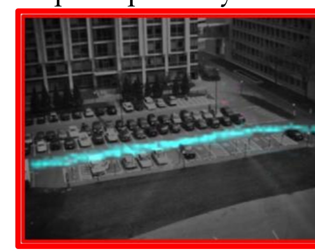

1pm-2pm May 15


7pm-8pm May 15


2am-3am May 16

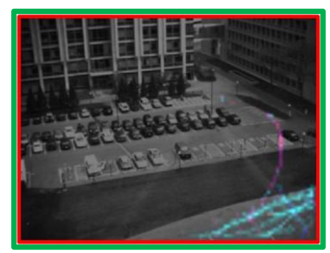

7am-8am May 16


1pm-2pm May 16


7pm-8pm May 16


2am-3am May 15


7am-8am May 15


1pm-2pm May 15


7pm-8pm May 15


2am-3am May 16

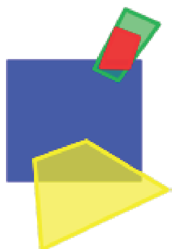

7am-8am May 16


1pm-2pm May 16


7pm-8pm May 16

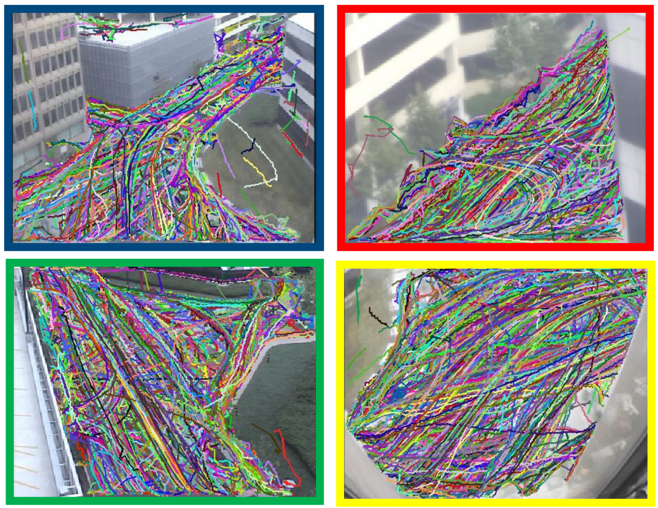
Cluster trajectories in multiple camera views

- Correspondence free: doesn't track object across camera views
- No camera calibration
- Unsupervised
- The topology of camera views is unknown and arbitrary (overlapping or non-overlapping)
- **Add smoothness prior according to the temporal co-occurrence of trajectories observed in different camera views**



Topology of camera views

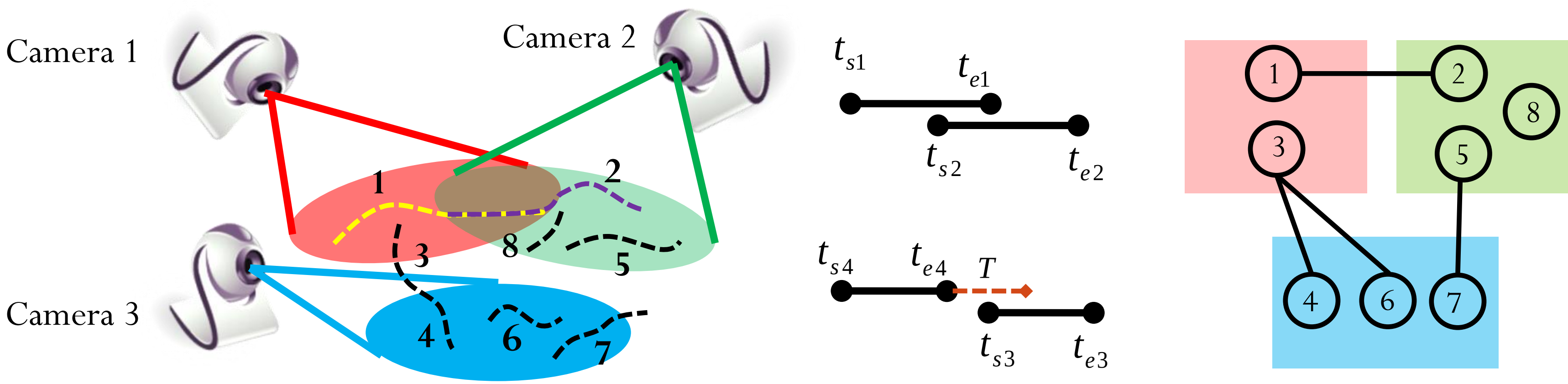
Four camera views



Trajectories observed in four camera views

Build a Trajectory Network

Identity co-occurrence + temporal co-occurrence



(a) Trajectories in three camera views

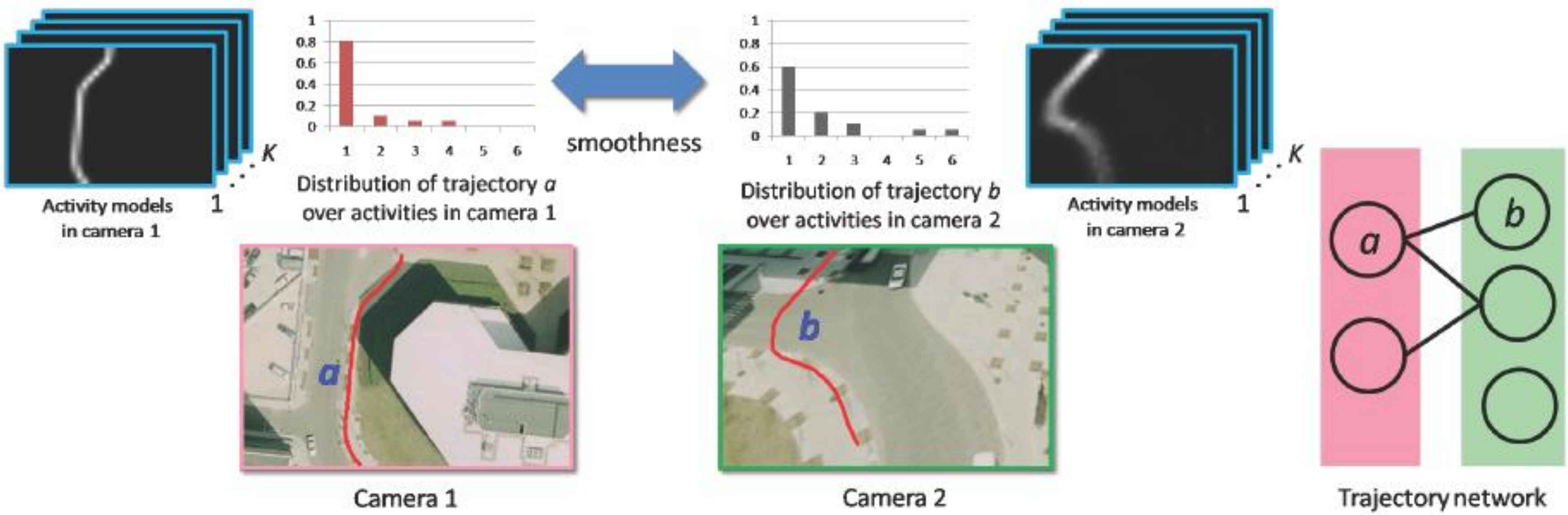
(b) Temporal extents of trajectories

(c) The network connecting trajectories

X. Wang, K. Tieu, and E. Grimson, "Correspondence-Free Activity Analysis and Scene Modeling in Multiple Camera Views," *TPAMI'10*.

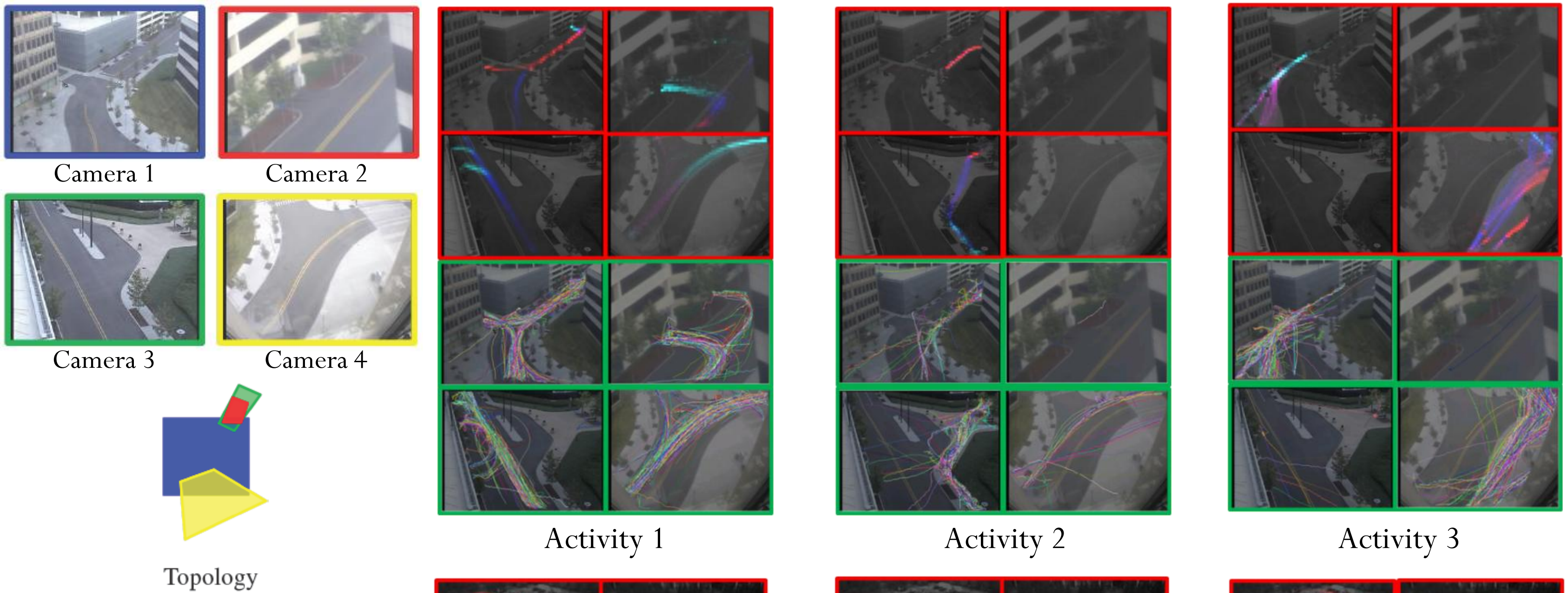
Model

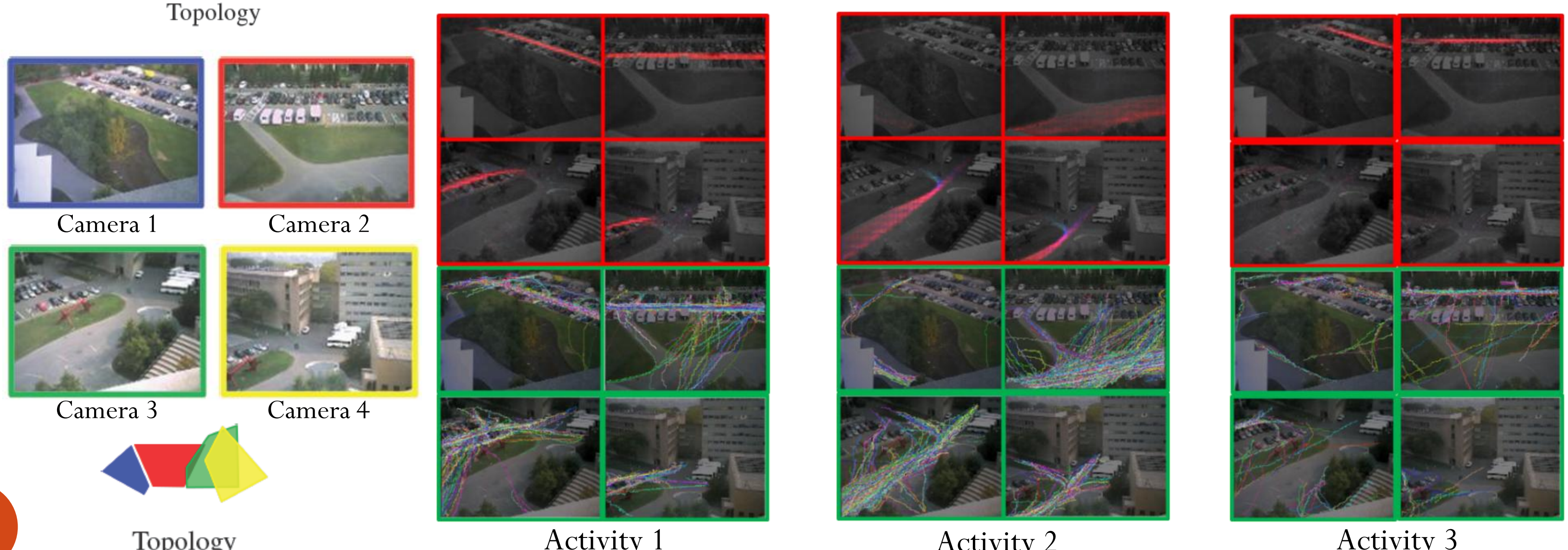
- Codebook is the concatenation of the local codebooks of all the camera views. Feature value is (*camera_id*, *location*, *moving_direction*).
- A semantic region has a joint distribution in all camera views.
- If two trajectories are connected by an edge on the network, there is a smoothness constraint on their distribution over semantic regions.



An example to describe the high level picture of our model

Models of paths and clusters of trajectories



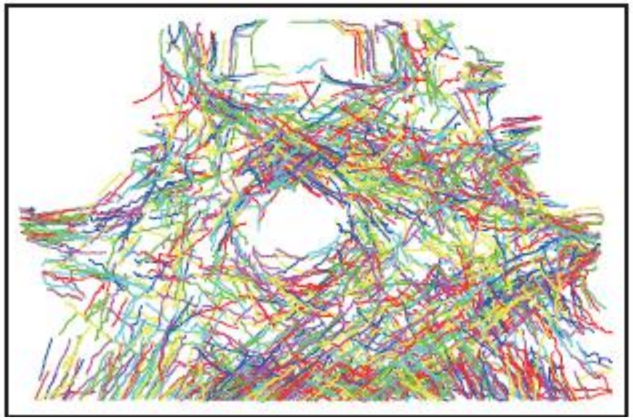


Outline

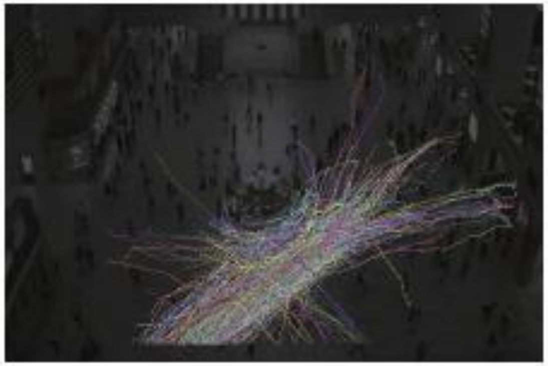
- Introduction
- Topic Models
- Activity analysis in sparse scenes
  - Based on trajectories
  - Single camera view
  - Multiple camera view
- **Activity analysis in crowded scenes**
  - **Based on tracklets**
  - **Based on local motions**
- Activity analysis in near fields

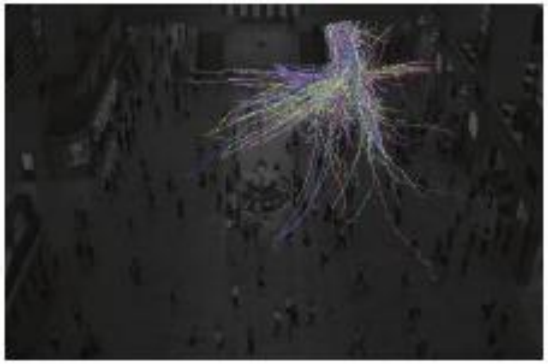
Learning the models of activities from tracklets

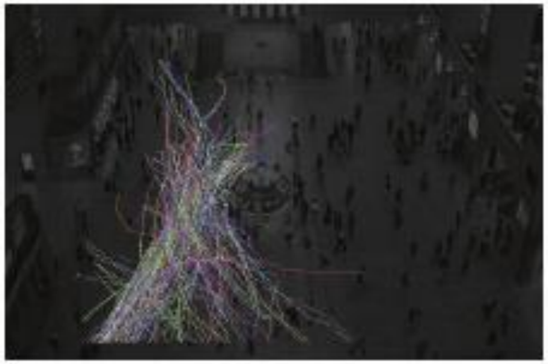
- Tracklets: fragments of trajectories obtained by weak trackers. They are short and very noisy.
- Trajectories extracted from the videos of the New York Grand Central Station: 95% trajectories are highly fragmented



Clustering results using topic models



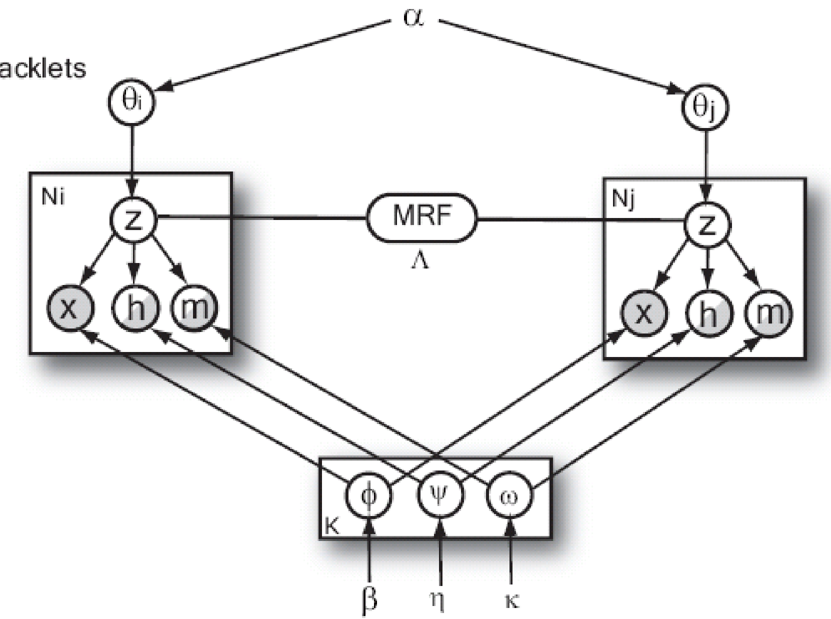




Random field topic model

- MRF models the dependency between tracklets based on their spatial and temporal consistency and velocity similarity
- Model the sources and sinks

B. Zhou, X. Wang, and X. Tang, "Random Field Topic for Semantic Region Analysis," *CVPR* 2011.

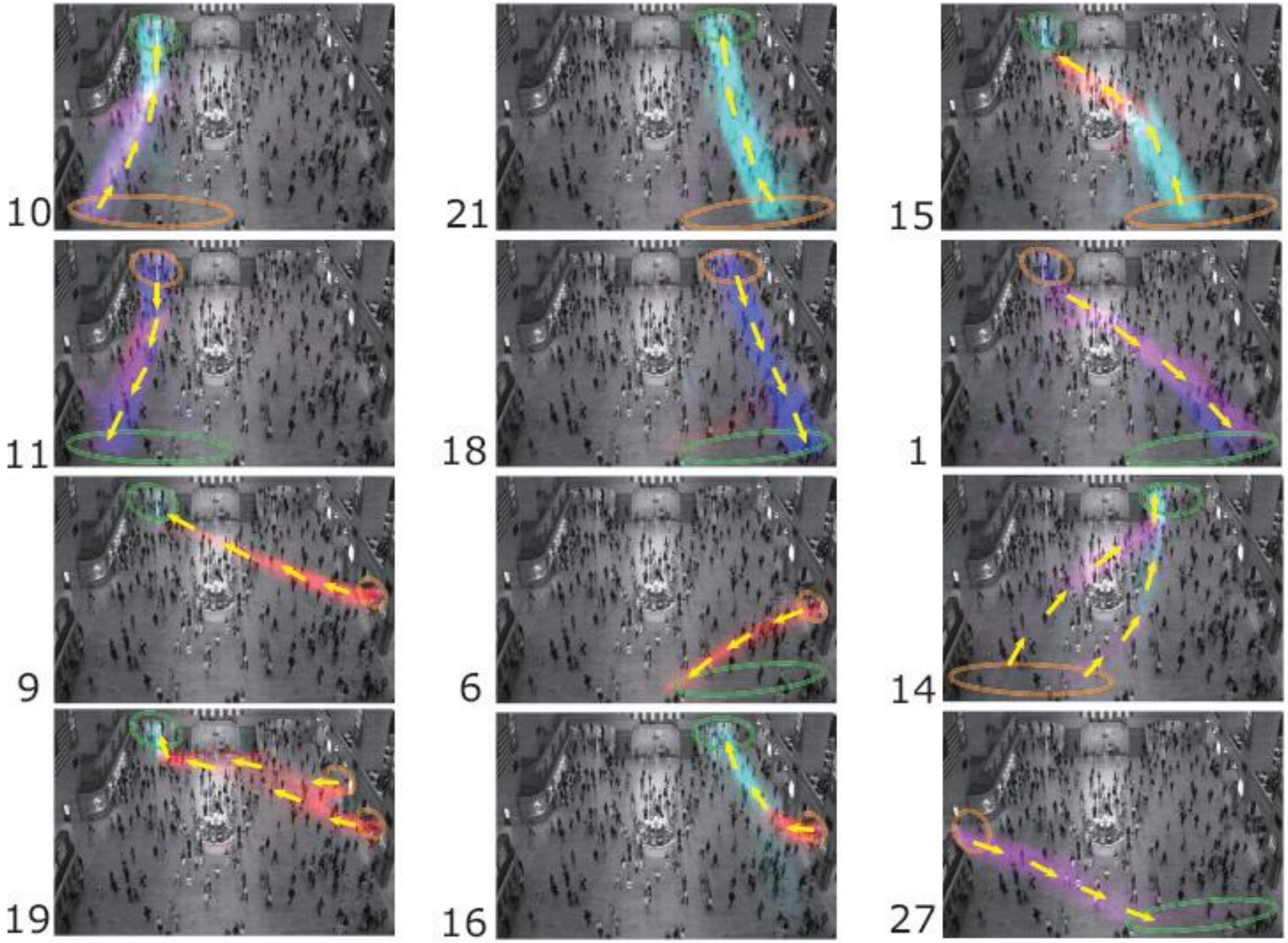

α
N tracklets
θi
θj
Ni
Nj
z
MRF
Λ
x
h
m
φ
ψ
ω
K
β
η
κ

(A)

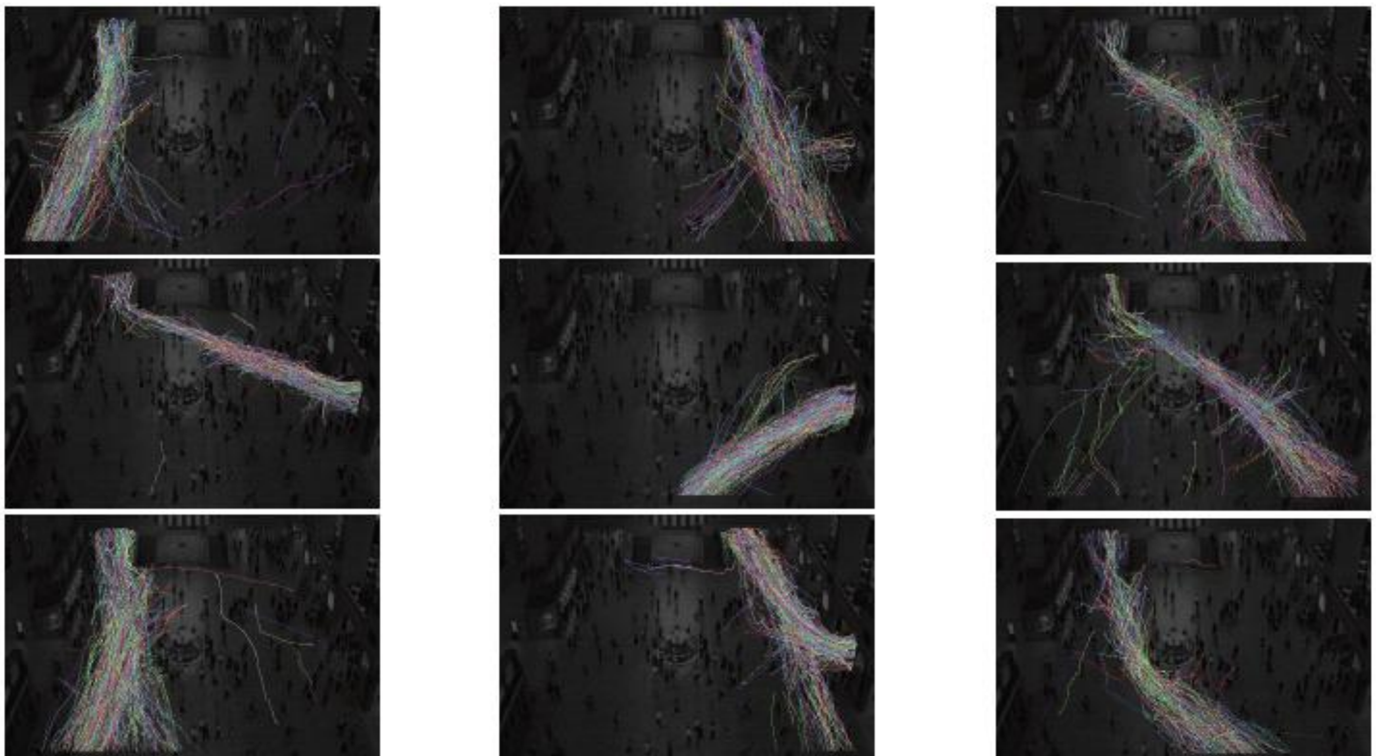

Tracklet:
complete
only source
only sink
no source&sink
MRF:
pairwise
tree
Source
Sink

(B)

Learned Models of Paths



Learned Models of Paths



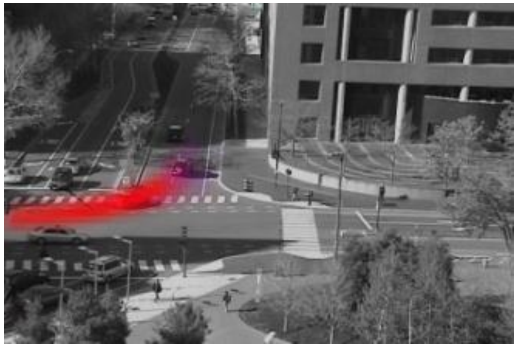
Activity Analysis Based on Moving Pixels

X. Wang, X. Ma, and E. Grimson, "Unsupervised Activity Perception in Crowded and Complicated Scenes Using Hierarchical Bayesian Models," *TPAMI'09*.

Activity Analysis Based on Moving Pixels

- Features = **positions and velocities of moving pixels**
- Words = **moving pixels**
- Documents = **short video clips**
- Topics = **atomic activities**
- **Temporal co-occurrence**: if two feature values are related to the same atomic activities, they often co-occur in the same video clips and have strong temporal correlation

Moving pixels in a short video clip



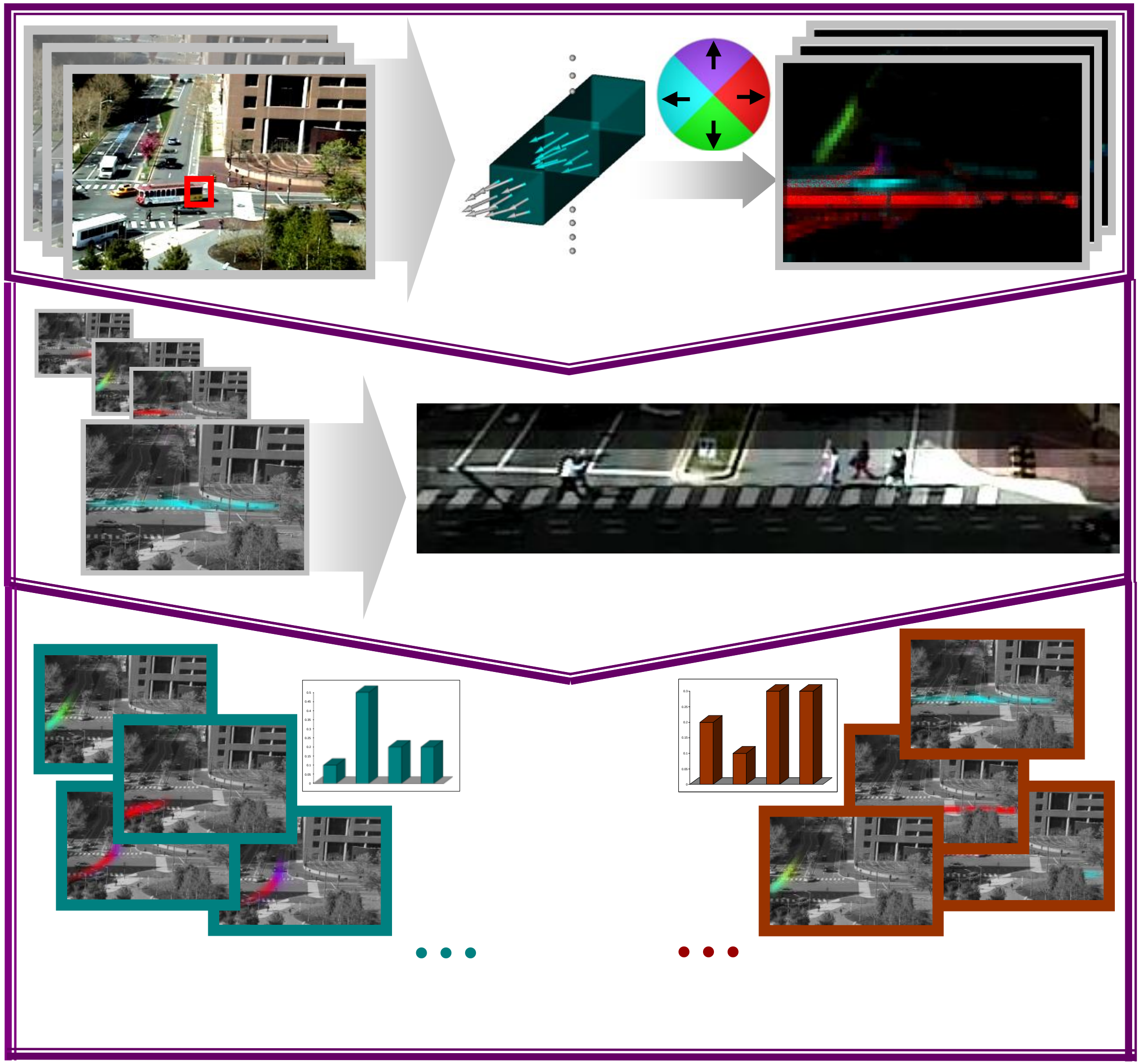
Spatial distribution of an atomic activity

High level picture of our approach

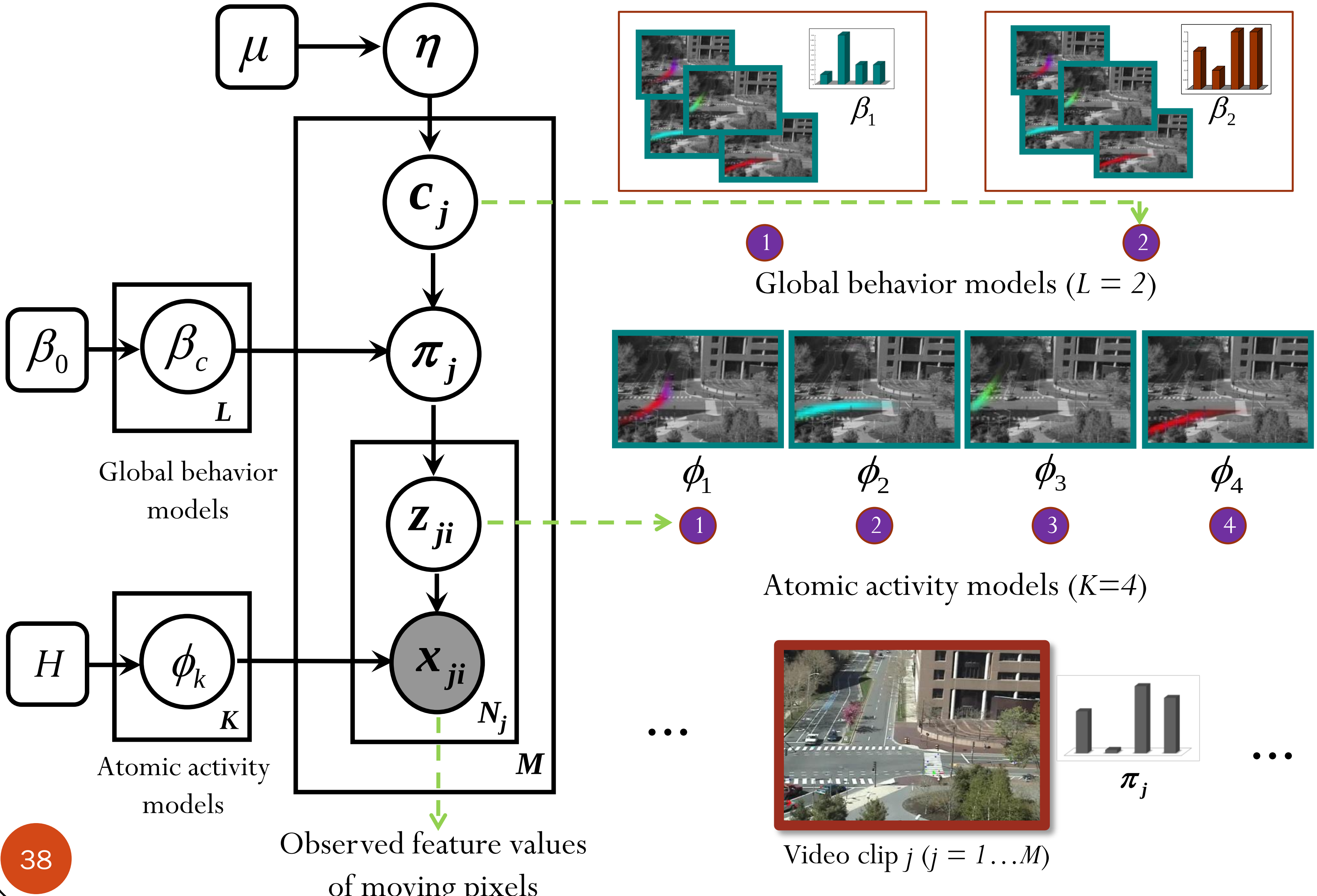
Motion Features
(a)

Atomic activities modeled as distributions over the feature codebook
(b)

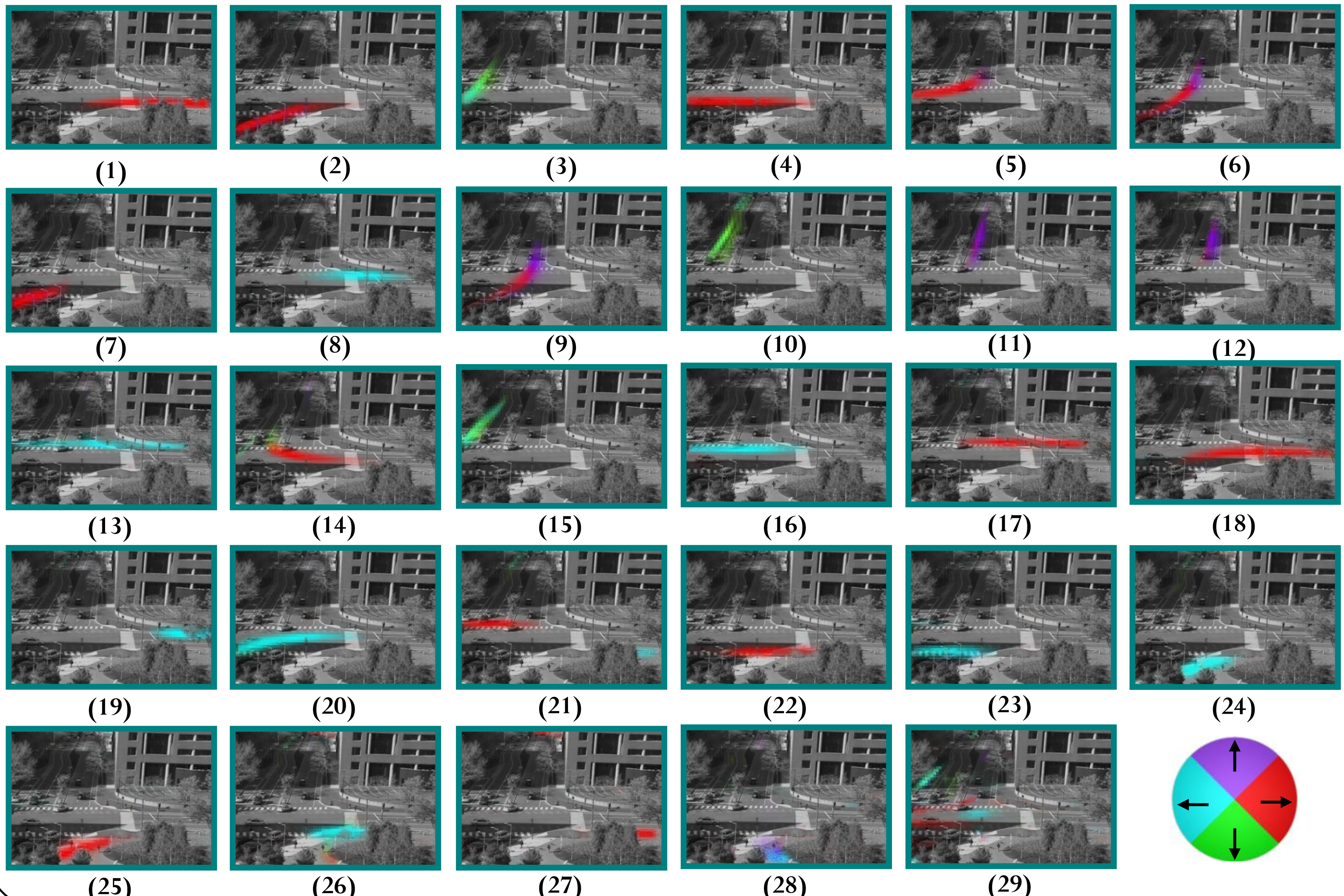
Global behaviors modeled as distributions over atomic activities
(c)

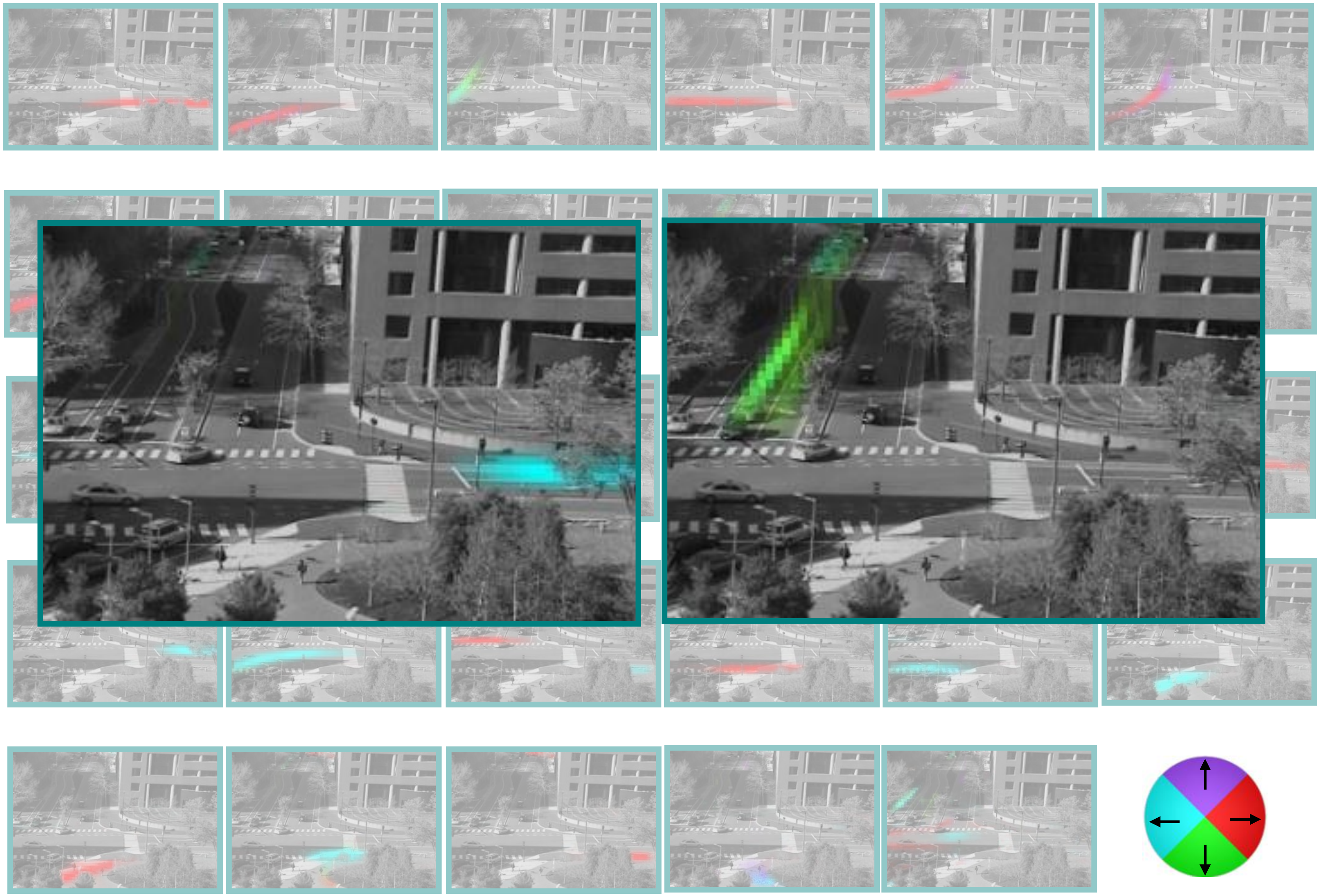


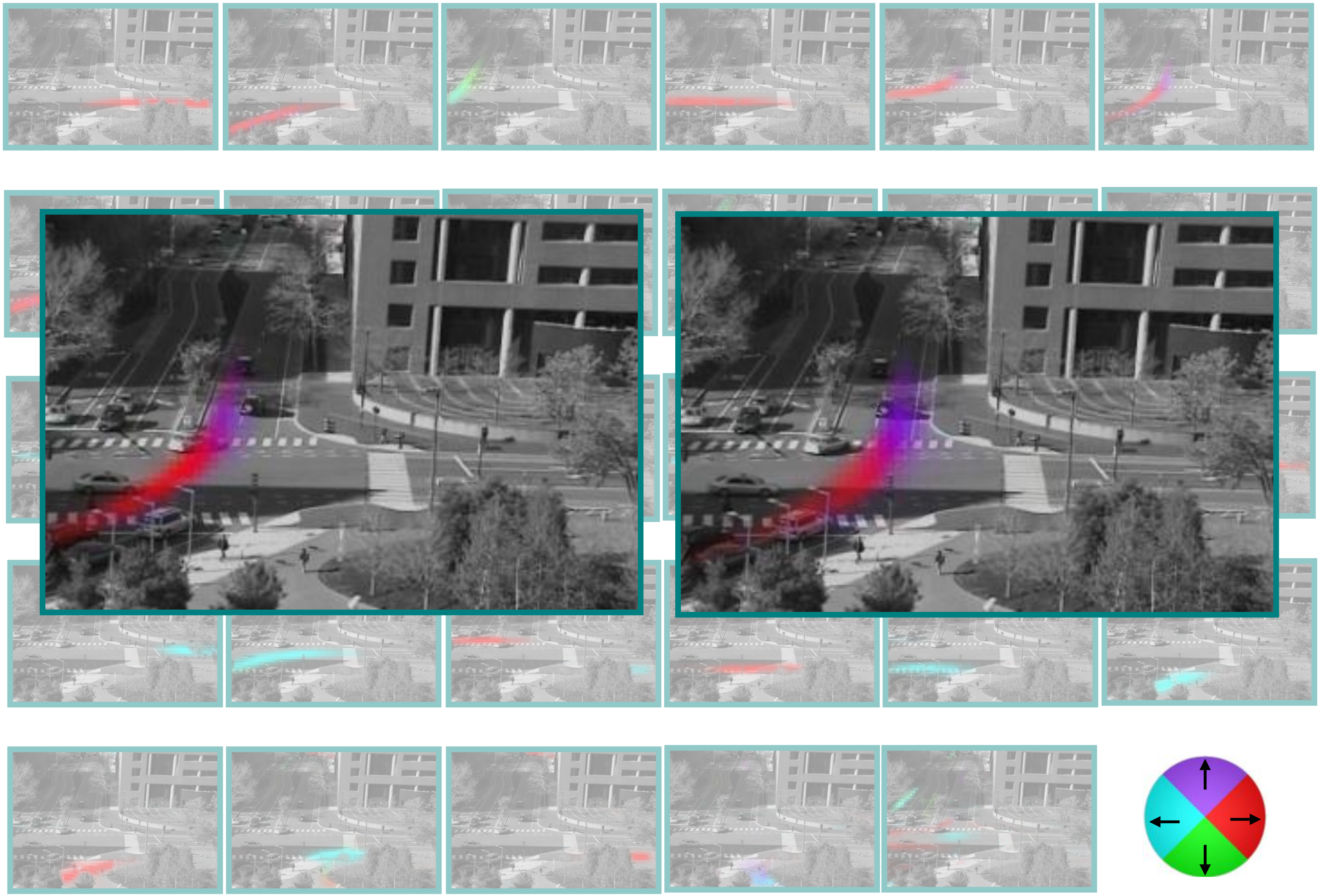
Parametric hierarchical Bayesian model



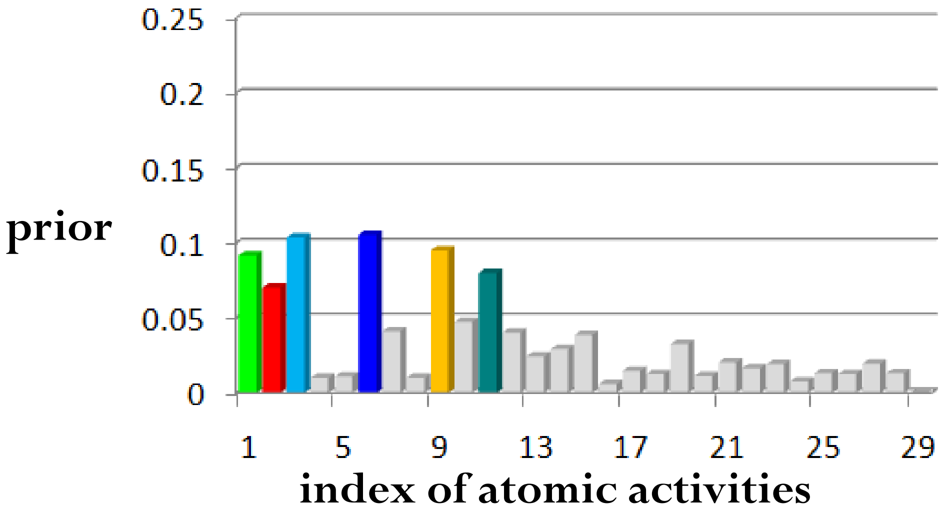
Learned atomic activities from a traffic scene

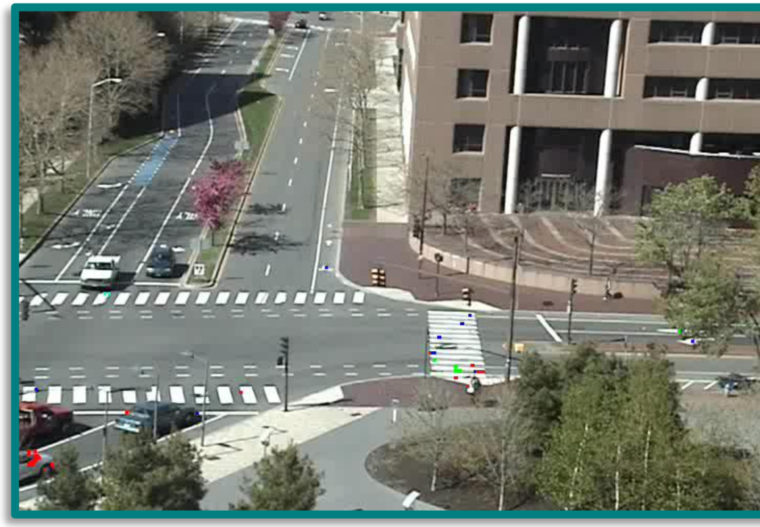




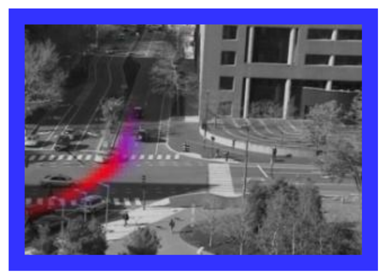


Global behavior I: green light for south/north traffic

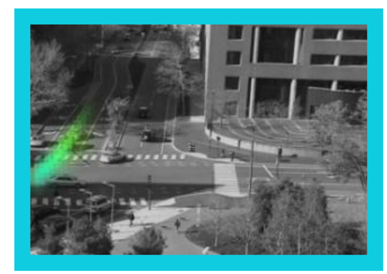




vehicles northbound



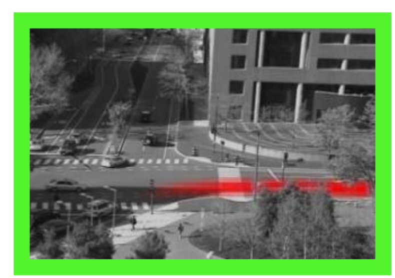
vehicles northbound



vehicles southbound

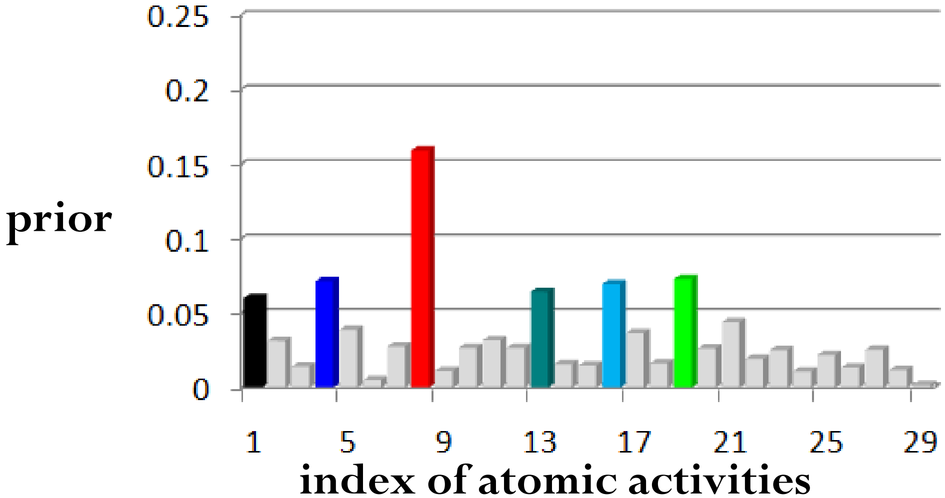
vehicles incoming northbound

vehicles incoming southbound



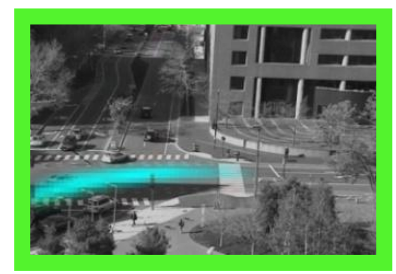
vehicles outgoing eastbound

Global behavior II: green light for east/west traffic



vehicles incoming westbound

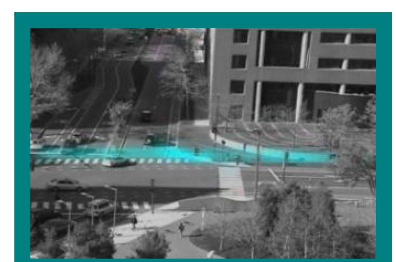
vehicles outgoing westbound



vehicles outgoing southbound

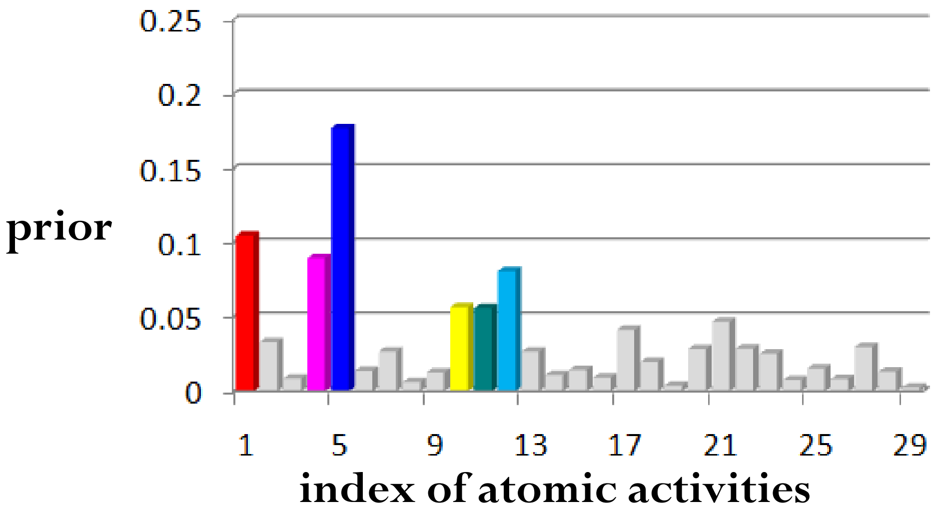
vehicles incoming eastbound

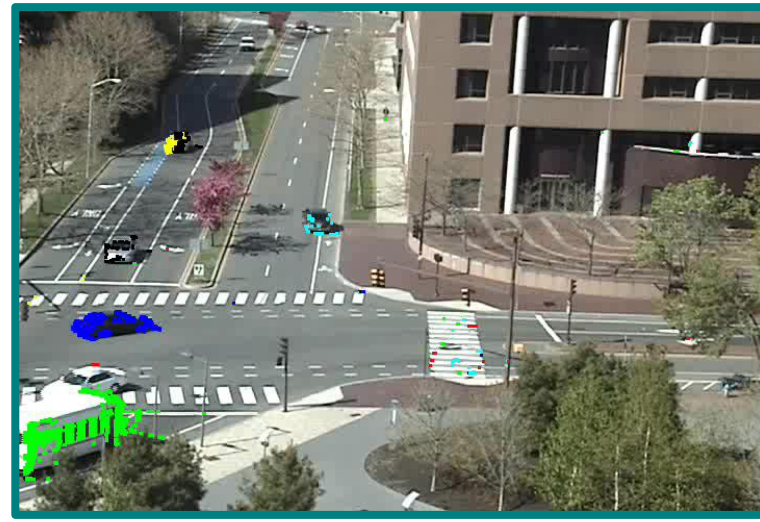
vehicles outgoing eastbound

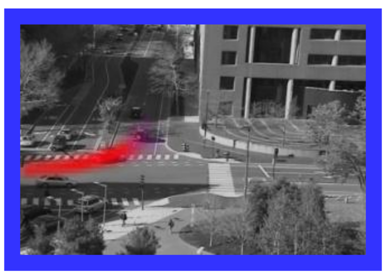


pedestrians westbound

Global behavior III: left turn signal for east/west traffic






vehicles turning left eastbound


vehicles outgoing northbound

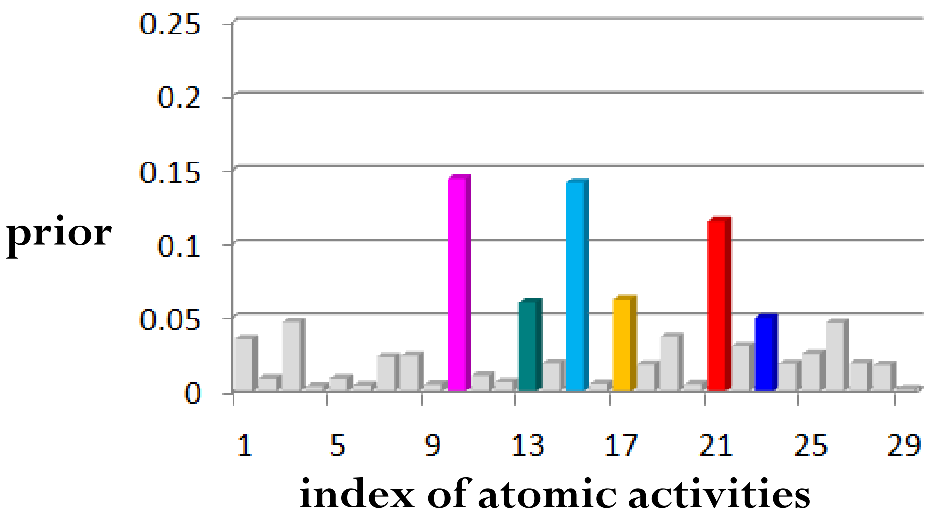

vehicles outgoing northbound

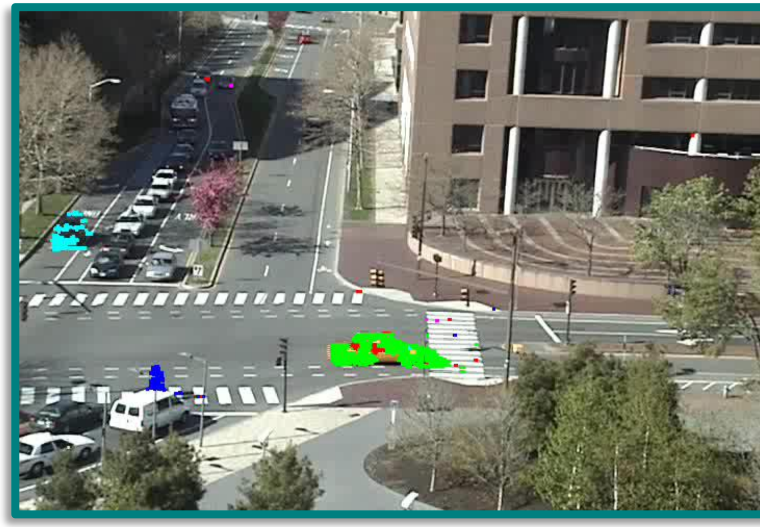

vehicles incoming eastbound


vehicles outgoing eastbound


vehicles stopping southbound

Global behavior IV: walk sign





pedestrians incoming eastbound

pedestrians outgoing eastbound

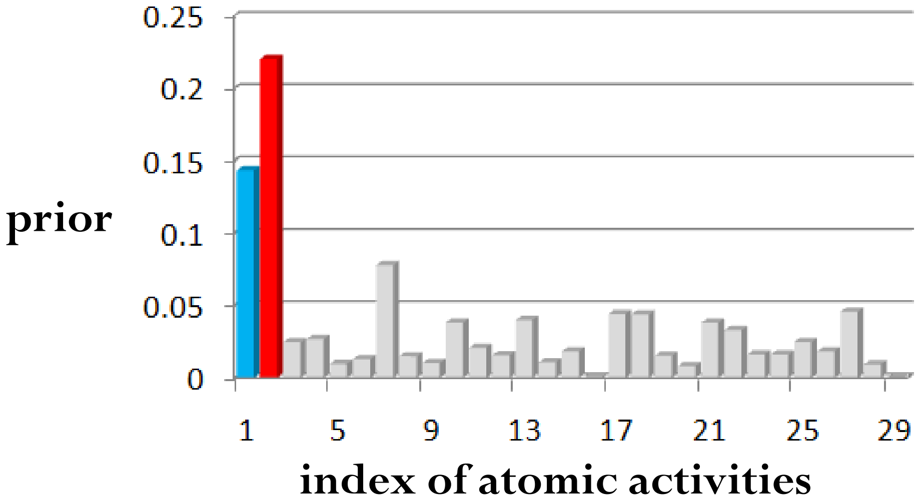
pedestrians westbound

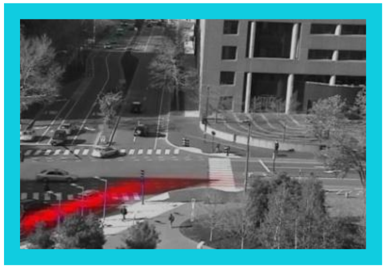
pedestrians westbound

vehicles stopping

vehicles stopping

Global behavior V: northbound right turns





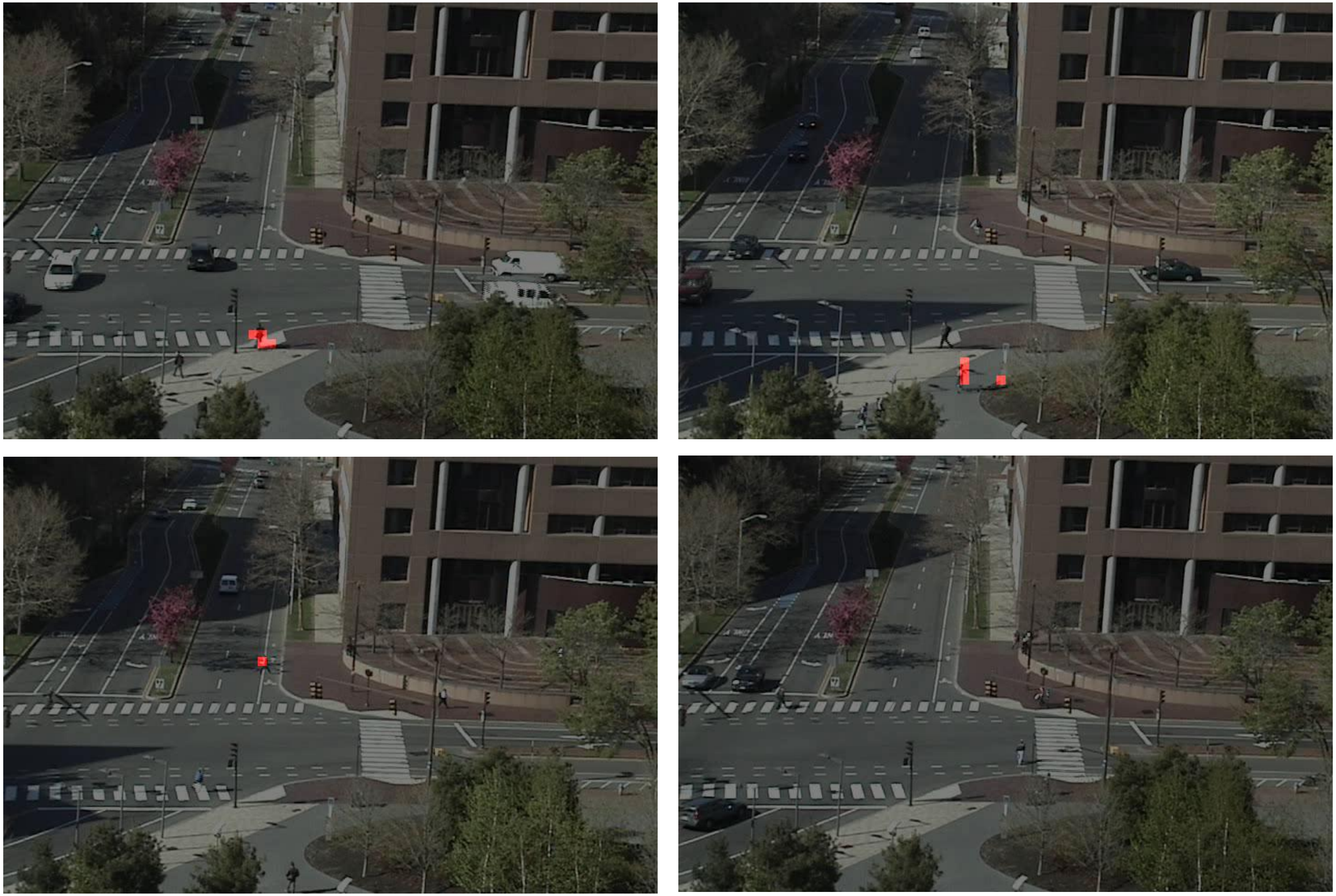
vehicles incoming northbound

vehicles outgoing eastbound

Temporal video segmentation

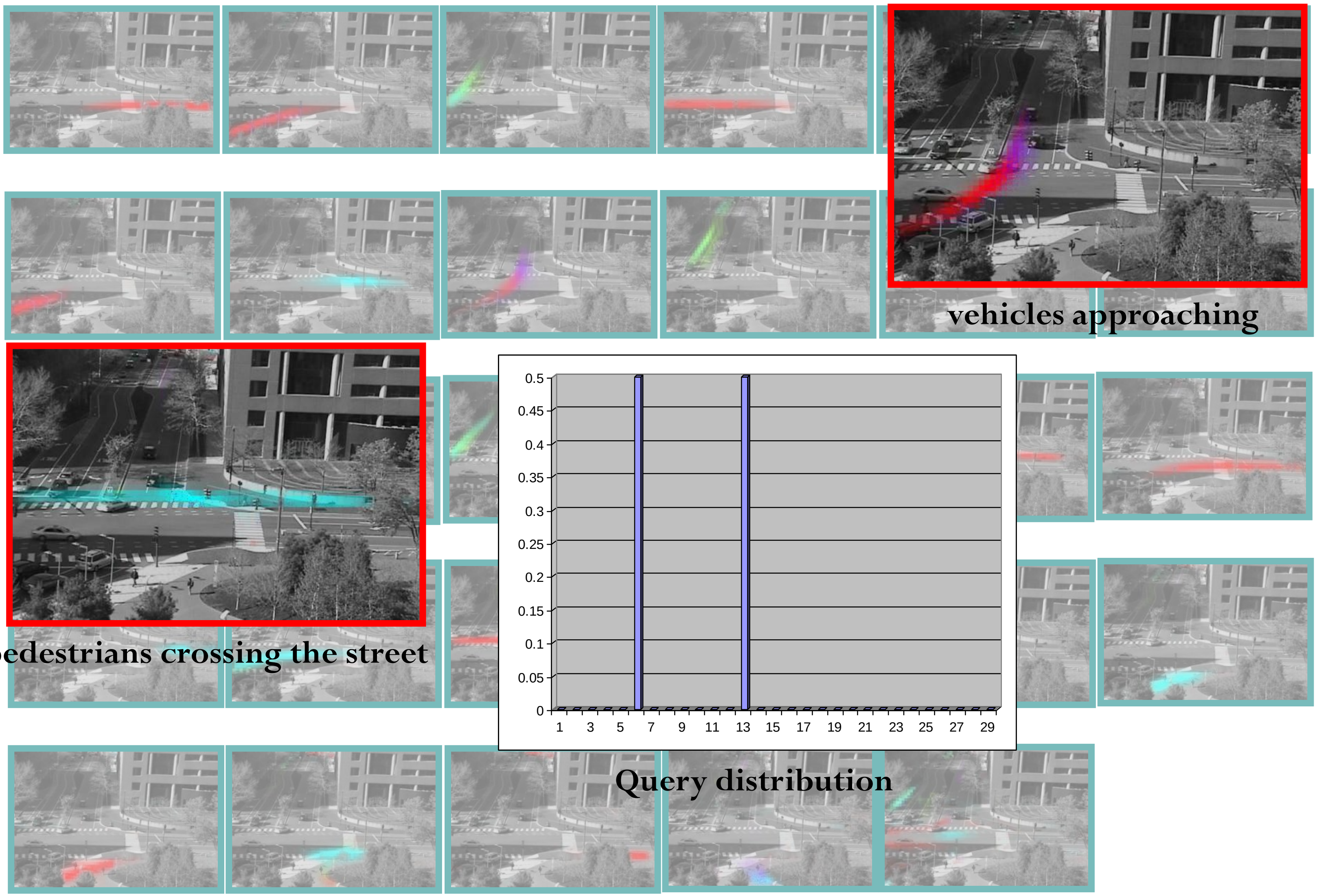
(magenta)	green light for east/west traffic
(green)	green light for south/north traffic
(cyan)	left turn signal for east/west traffic
(yellow)	walk sign
(blue)	northbound right turns

Abnormality detection results



Top four abnormal video clips

Interaction query



Top four retrieved jay-walking examples

More works

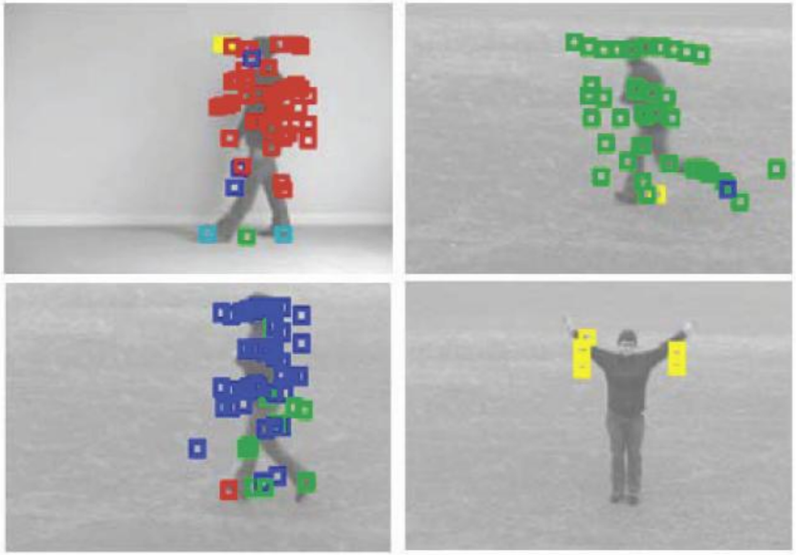
- Modeling the temporal dependencies of global behaviors and atomic activities (Hospedales et al. ICCV'09)
- Modeling the temporal duration of atomic activities (Varadarajan et al. BMVC'10, Emonet et al. ECCV'11)
- Modeling the temporal variations of the atomic activities and global behaviors over time (Faruquie, BMVC'09)
- Weakly supervised topic model for rare (abormal) and subtle behavior detection (Hospedales, et al. TPAMI'11)

Outline

- Introduction
- Topic Models
- Activity analysis in sparse scenes
  - Based on trajectories
  - Single camera view
  - Multiple camera view
- Activity analysis in crowded scenes
  - Based on tracklets
  - Based on local motions
- **Activity analysis in near fields**

Activity analysis in near fields

- Features = **visual descriptors of space-time interest points**
- Words = **space-time interest points**
- Documents = **video sequences**
- Topics = **actions**



J. Niebles, H. Wang, and L. Fei-fei, "Unsupervised learning of human action categories using spatial-temporal words," BMVC'06

Conclusions and Discussions

- Topic models capture the co-occurrence of features and can be applied activity analysis under different contexts
- They are extendable by adding different types of priors
  - Dynamically update the models of activities
  - Activity analysis across multiple camera views
- How to better capture the spatial and temporal relationships of "words" and "documents"?
- How to apply topic models to very large camera networks?
- How to jointly solve low-level object detection/tracking and high-level activity modeling under the hierarchical Bayesian model?

Thank you!