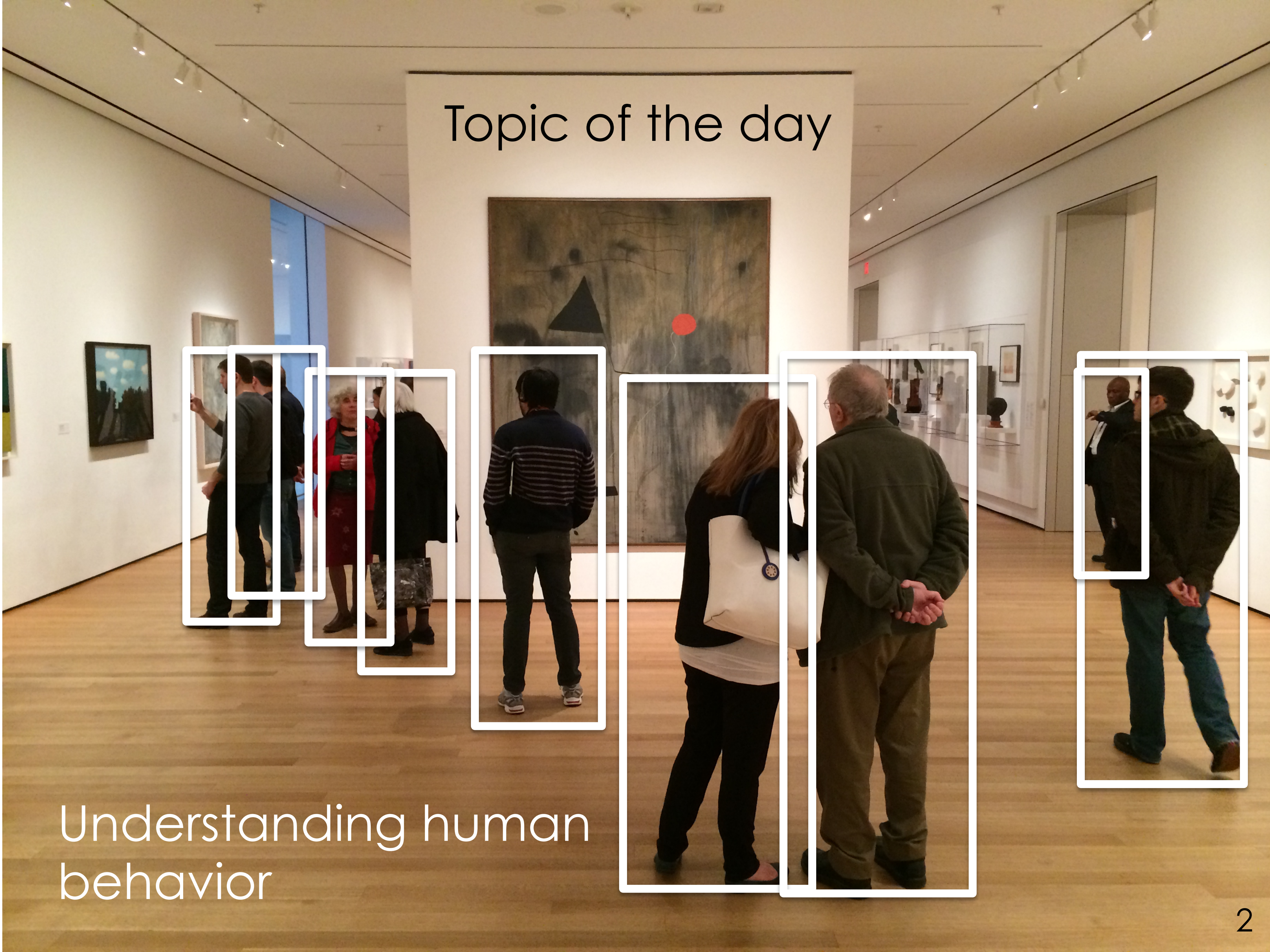
CS131
Tracking people

A. Alahi

November 19th 2014

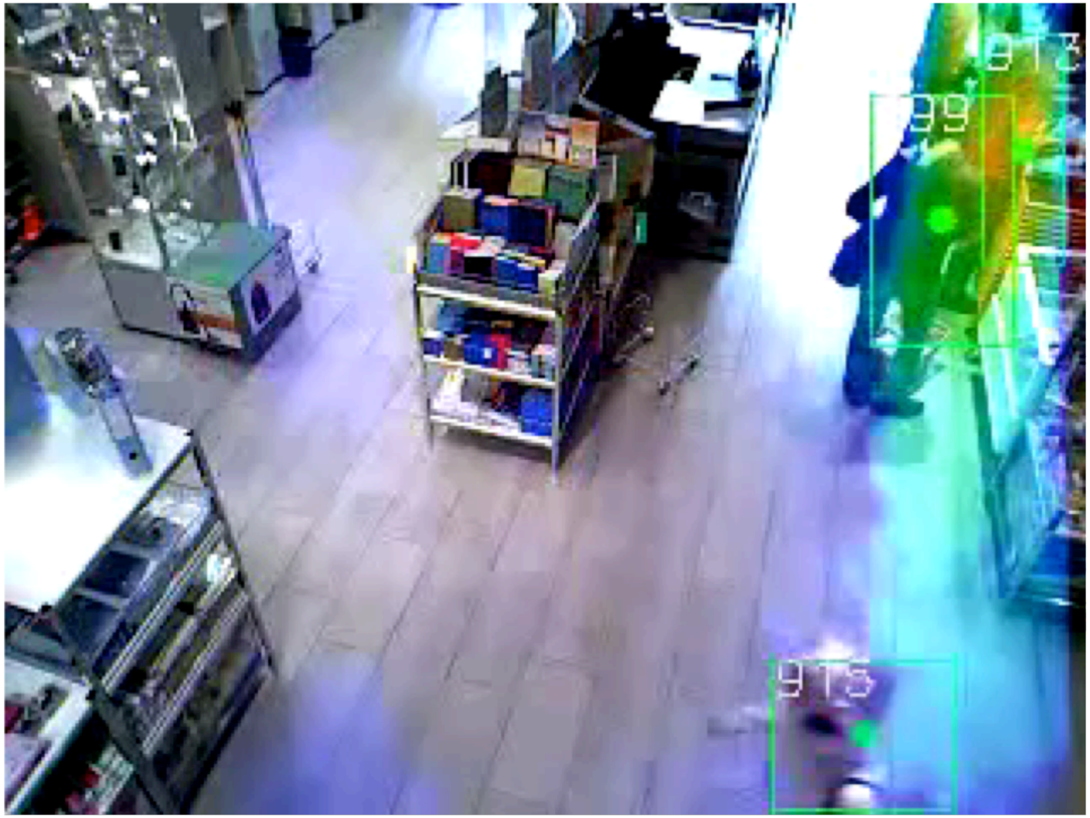
STANFORD UNIVERSITY


Topic of the day
Understanding human
behavior

Motivation:
Elderly monitoring

Courtesy of Ph.D Guido Pusiol

Motivation:
Path-to-purchase understanding



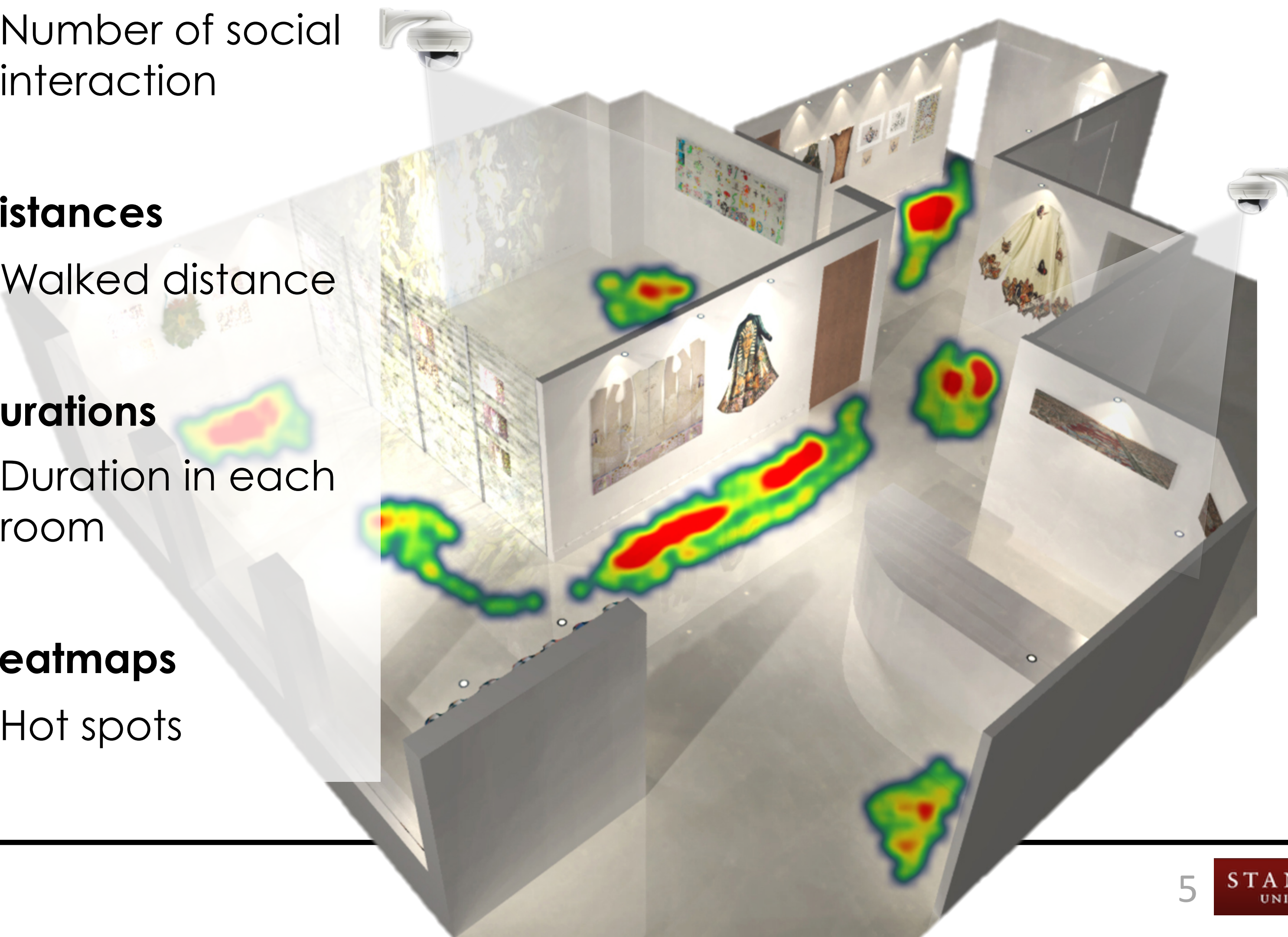
Motivation: Space analytics

Interaction
- Number of social interaction

Distances
- Walked distance

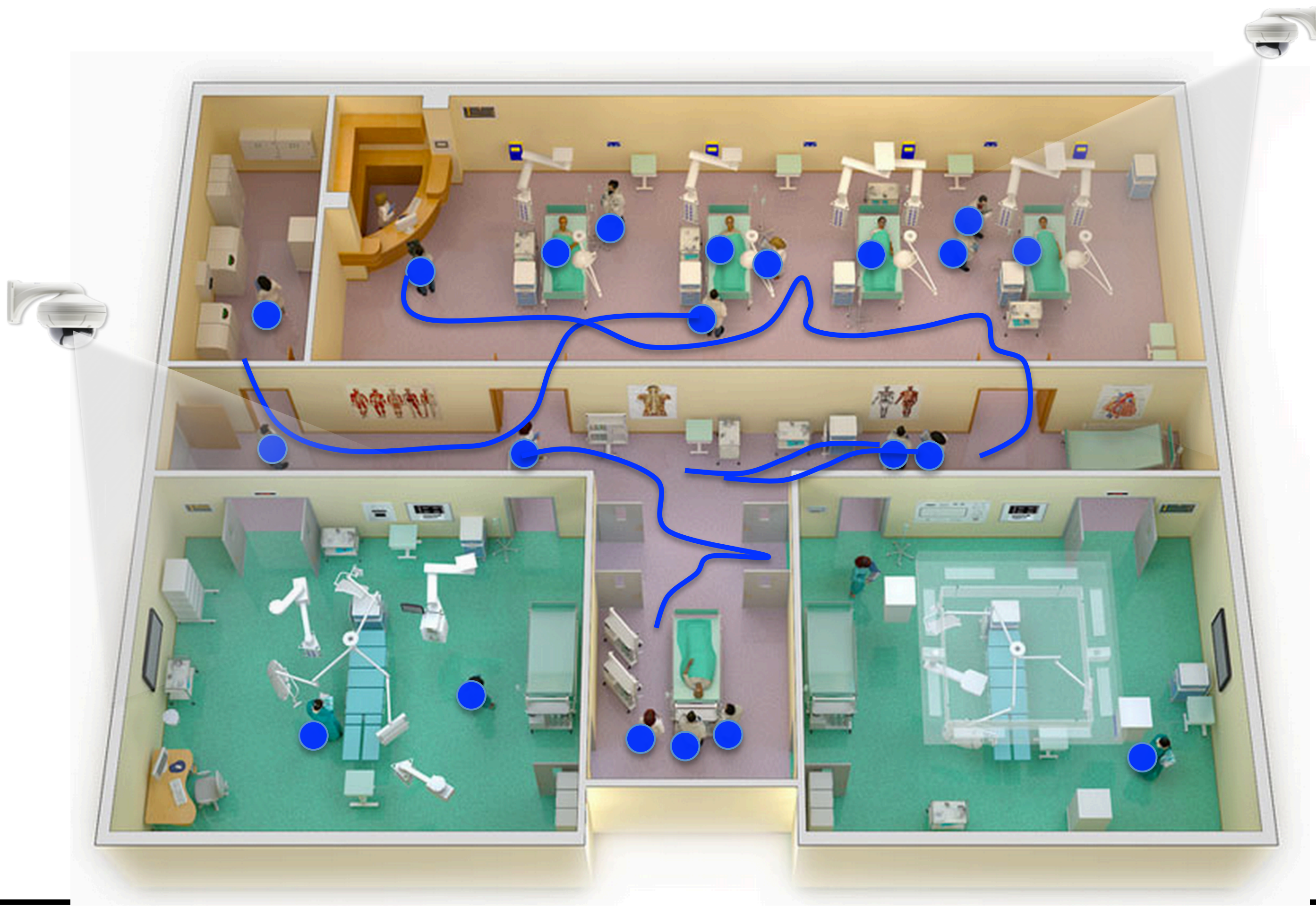
Durations
- Duration in each room

Heatmaps
- Hot spots



Motivation:
Performance analysis

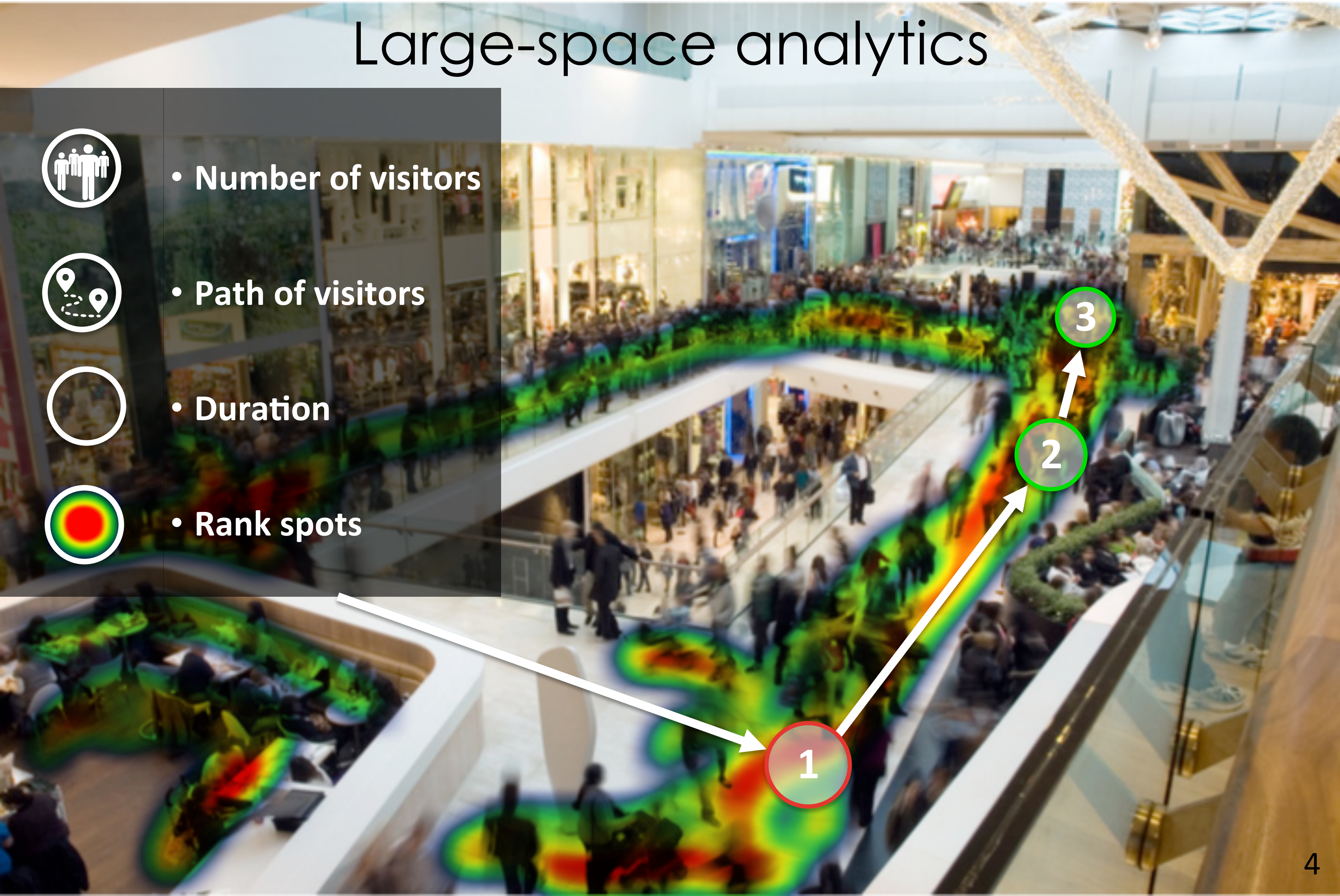
Motivation:
Behavior monitoring



Motivation:
Large-space analytics

Motivation:
Large-space analytics

- Number of visitors
- Path of visitors
- Duration
- Rank spots



Motivation:
Large-space analytics

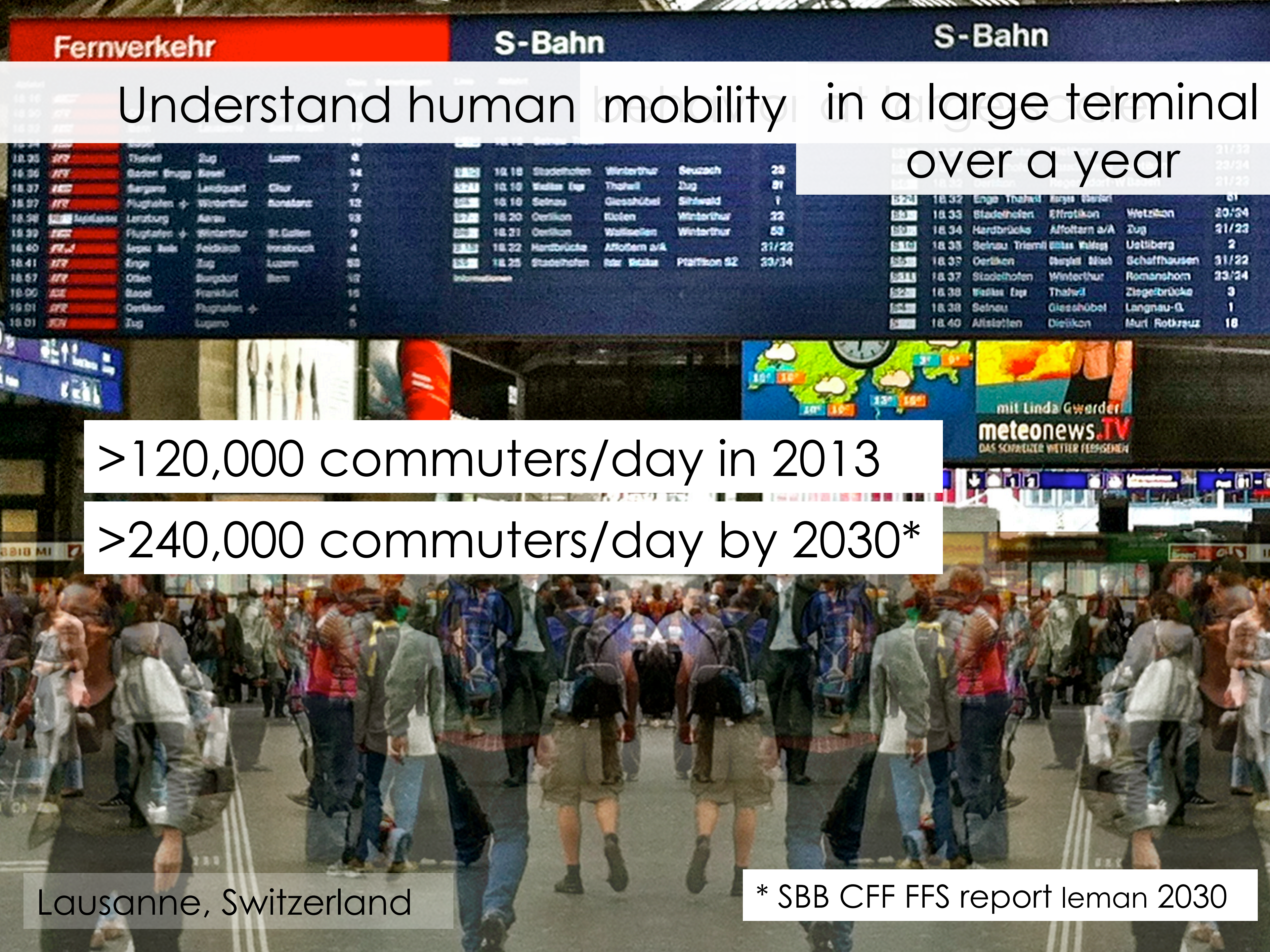
- Number of visitors
- Path of visitors
- Duration
- Rank spots

- Maximize productivity
- Increase revenues
- Optimize staff
- Justify infrastructure change

Understand human mobility in a large terminal over a year

>120,000 commuters/day in 2013

>240,000 commuters/day by 2030*



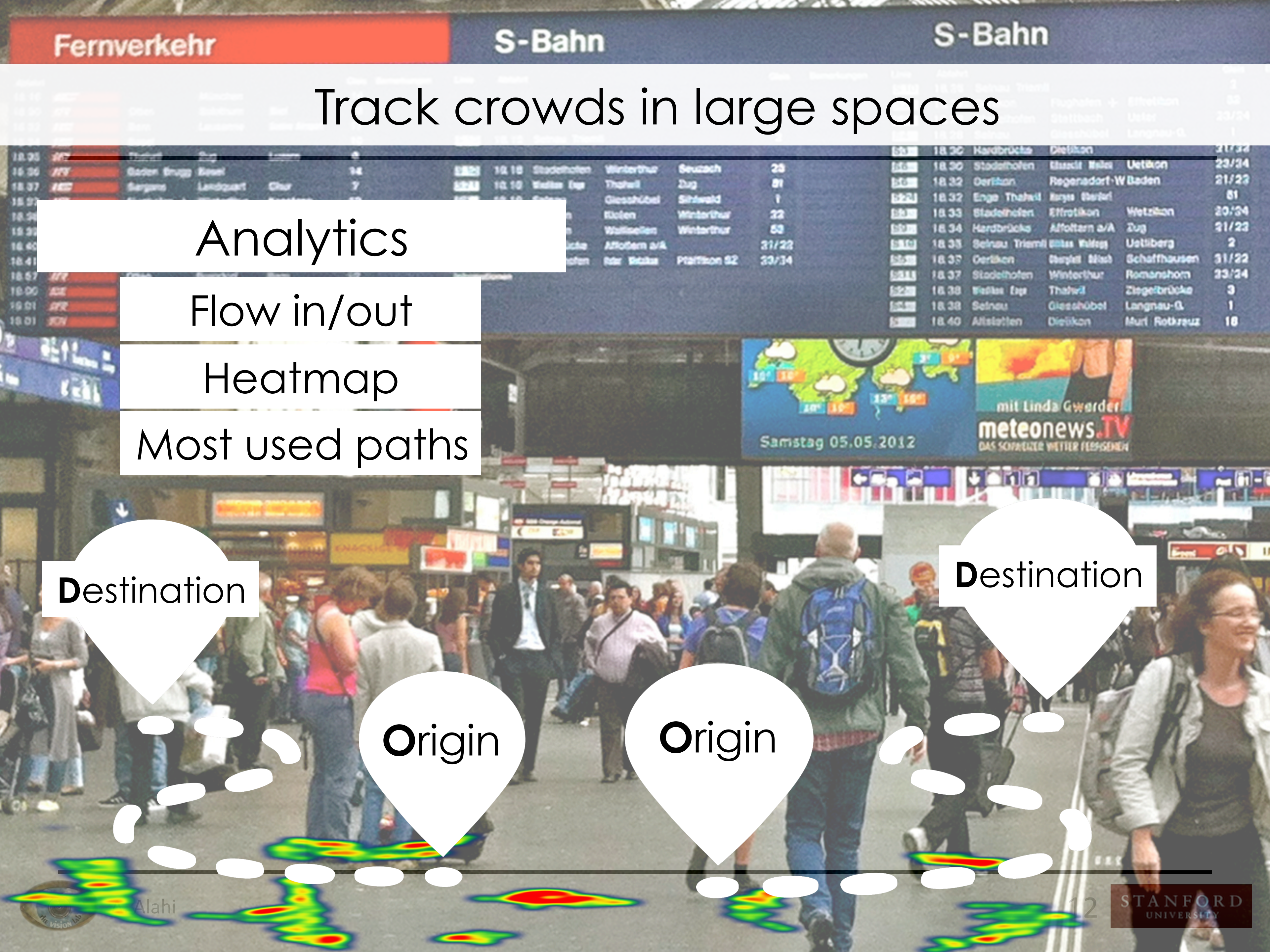
Lausanne, Switzerland

* SBB CFF FFS report leman 2030

Track crowds in large spaces

Analytics

- Flow in/out
- Heatmap
- Most used paths



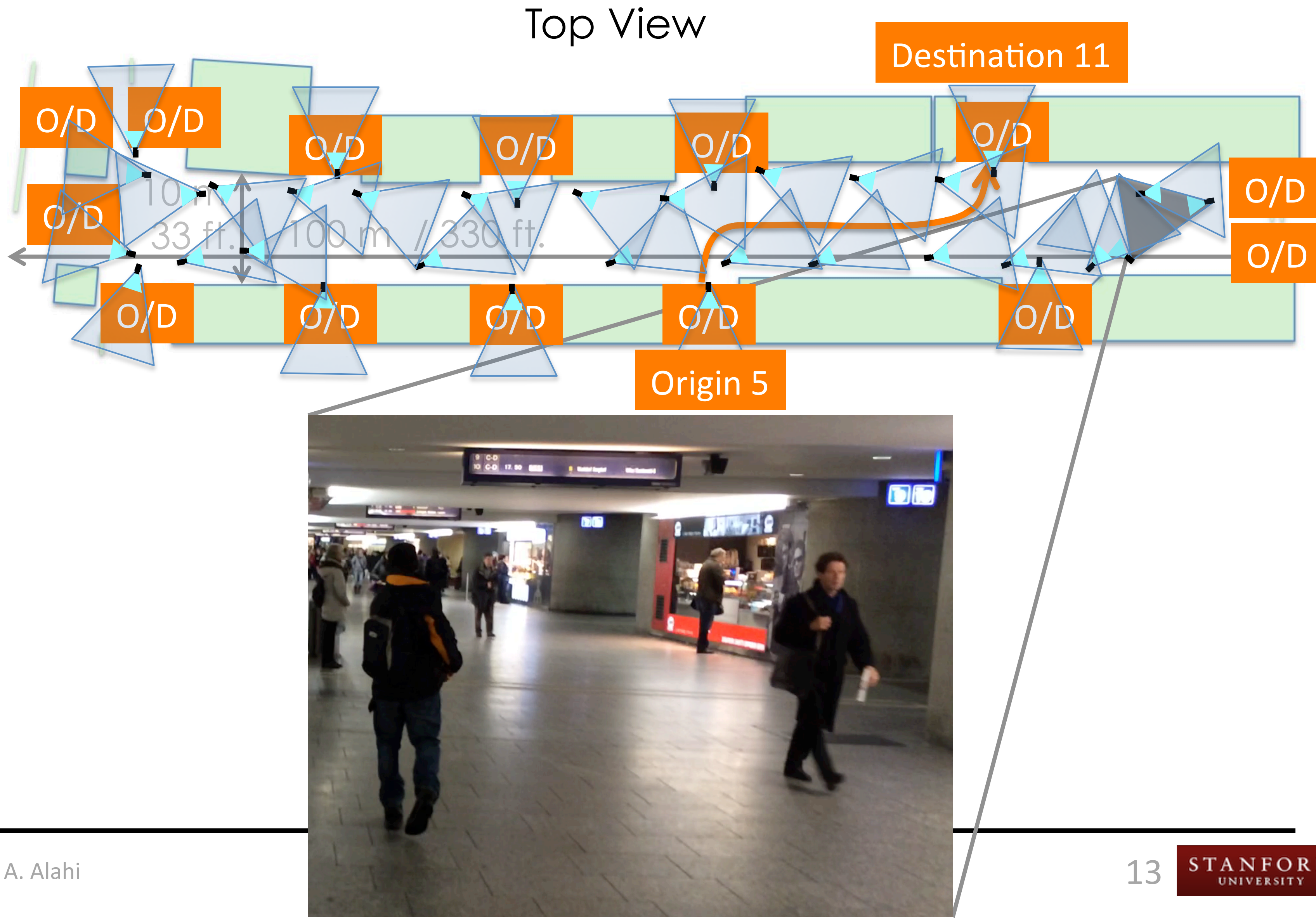
Destination

Origin

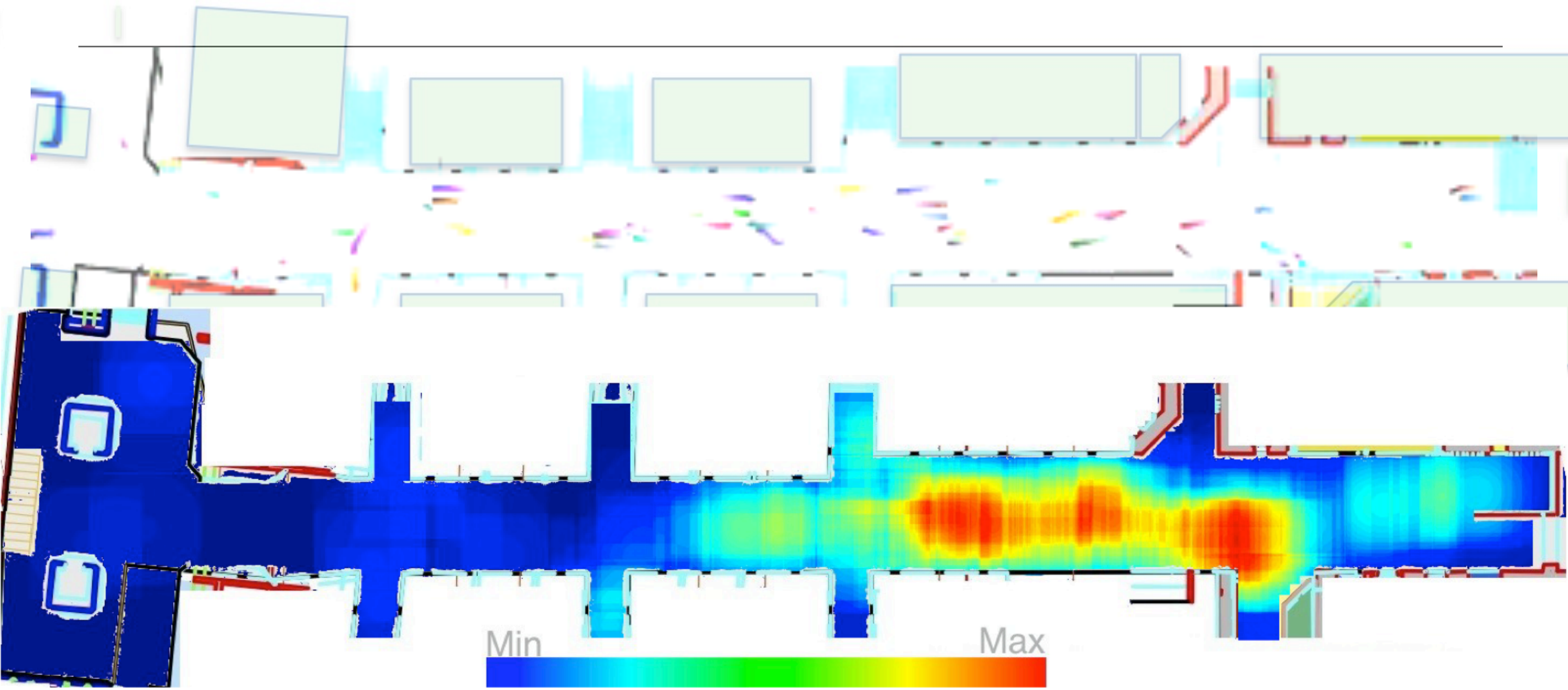
Origin

Destination

A corridor with 14 Origin/Destination (O/D)



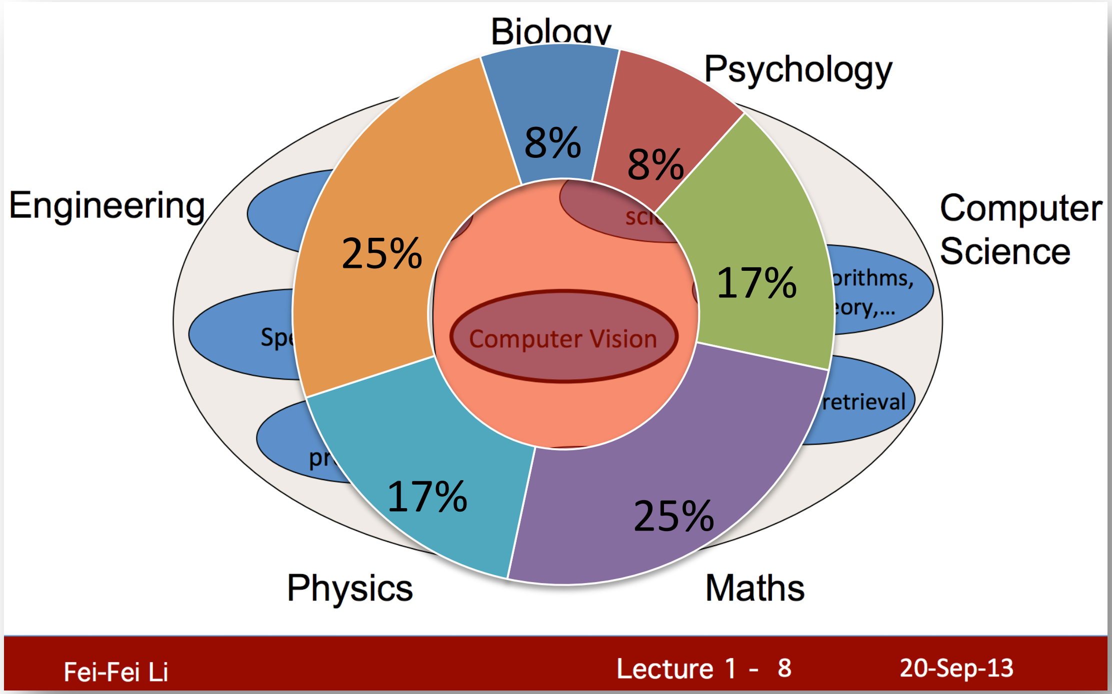
Collect long-term trajectories



# People	Av. duration	Av distance	Density (up to)	# Paths (O/D)
42 million	1 minute	100m	1 pedestrian/m^2	196

Tracking people

- Thousands scientific publications about tracking people
- What is related to tracking people?



Tracking people

- Focus of today's lecture:
  - Tracking people with
    - Video streams (vs static images)

      ⇒Prior from the scene

      ⇒Calibration data

      ⇒Mapping from image plane to real-world

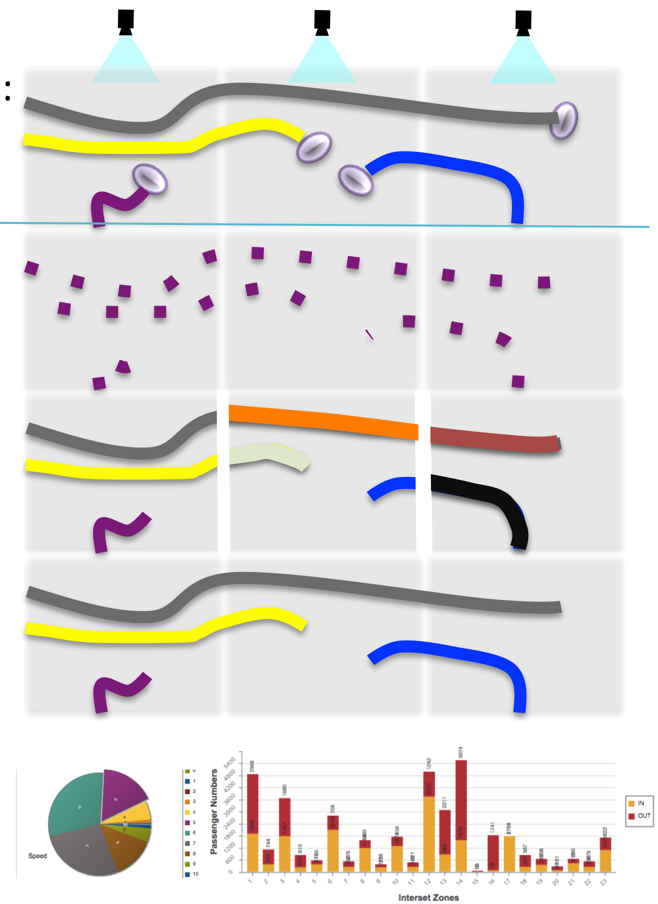
      ⇒Model temporal variation/changes

Outline:

From Foreground Extraction **To** Tracking 42 million Pedestrians

Network of 3 sensors:
4 people walking
(top view)

>100 GB/day



1. Detection intra-sensor — 5GB/day
2. Tracklet Generation — 4 GB/day
3. Tracklet Association — 5 GB/day
4. Statistic Generation — 0.05 GB/day

Outline:

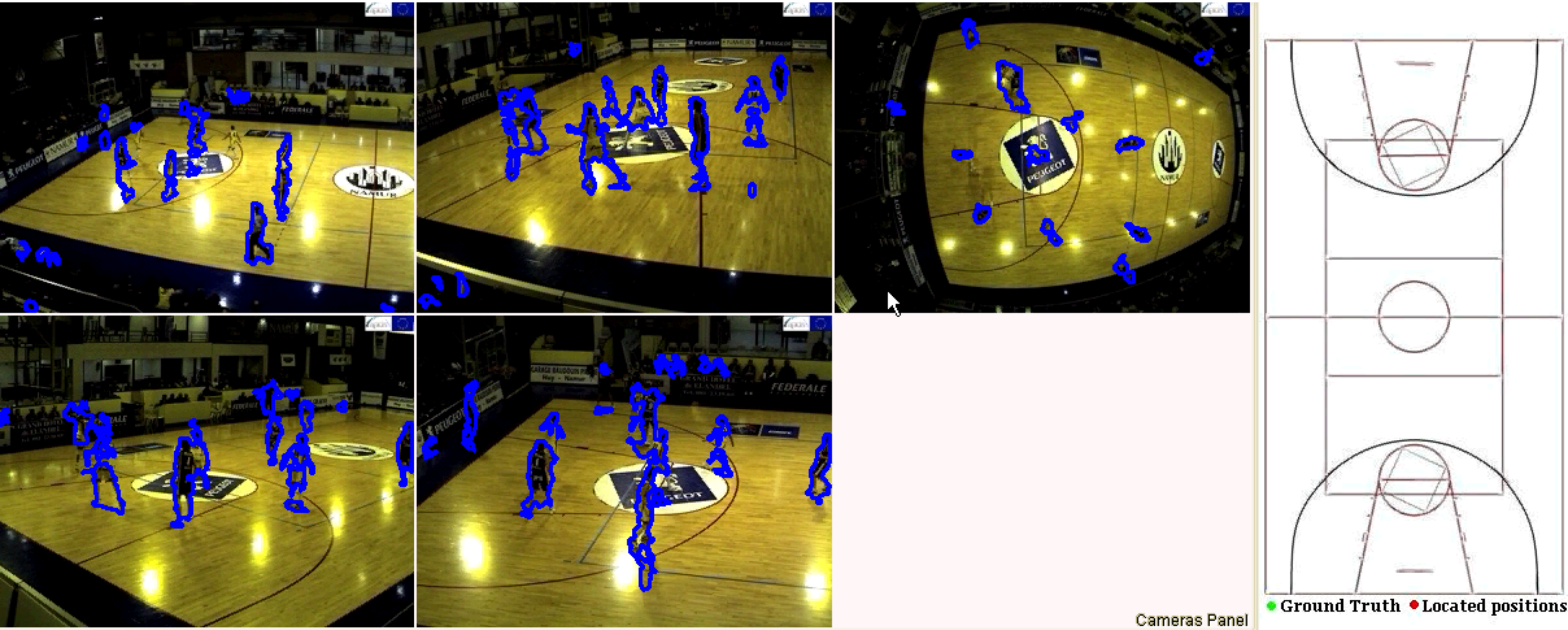
From Foreground Extraction **To** Tracking 42 million Pedestrians

I. Detection
   I. Foreground extraction
   II. Pedestrian localization

II. Tracklet Generation
   I. Data association problem
   II. Matching appearance cues

III. Tracklet Association
   I. Modeling Social Affinities

I. Detection: Foreground extraction



- Severely degraded foreground silhouettes
- Spatially dense distribution
- Strong occlusions

I. Detection: Foreground extraction

$$F(x,y,t+1) = \begin{cases} 1 \text{ if } \;|\; I(x,y,t) - \mathbf{B}(x,y) \;|\; > T \\ 0 \;\; \text{otherwise} \end{cases}$$

- Frame differencing

$$\mathbf{B}(x,y) = I(x,y,t-1)$$

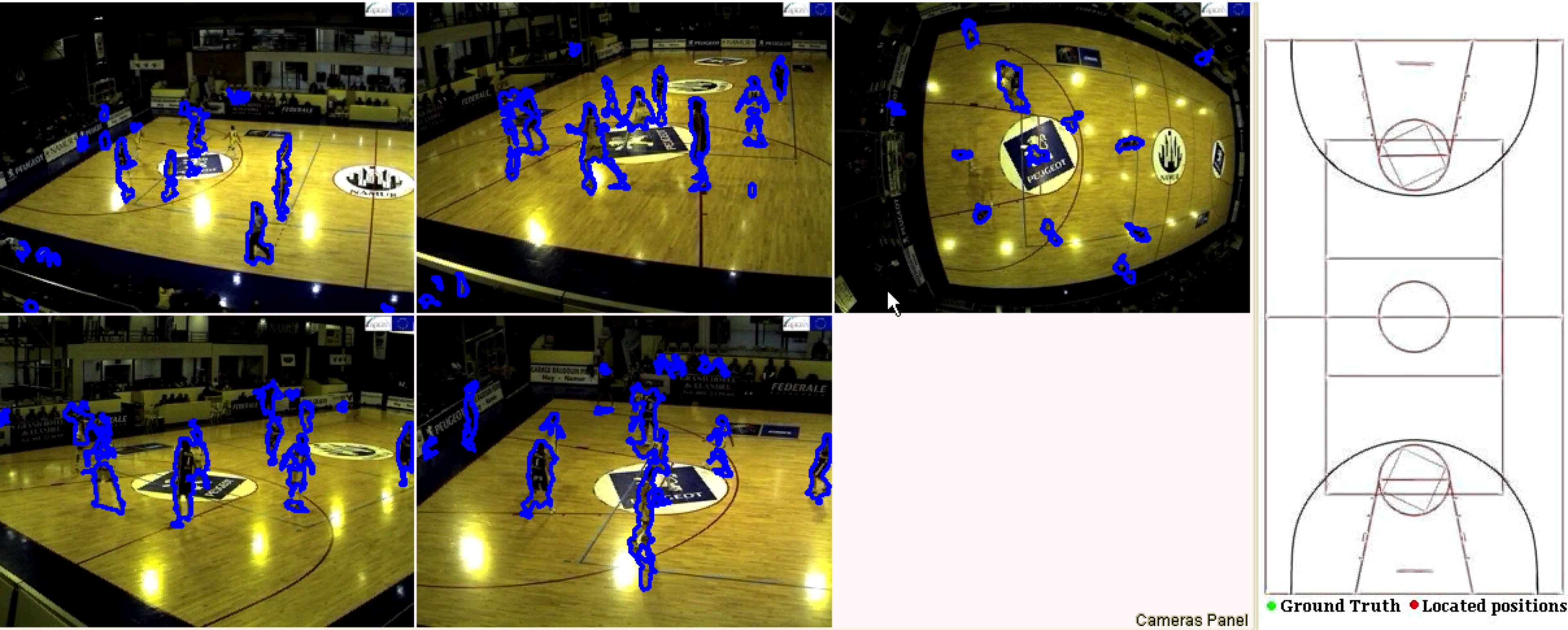
- Mean filter

$$B(x,y) = 1/N \;\; \Sigma_{i=1\ldots N}\, I(x,y,t-i)$$

- Gaussian averaging
- GMM
- …

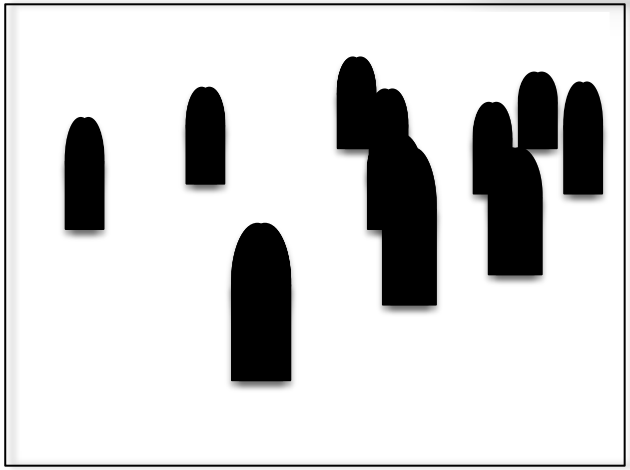
=> Library of 32 algorithms (*BGS library*)

I. Detection: Foreground extraction



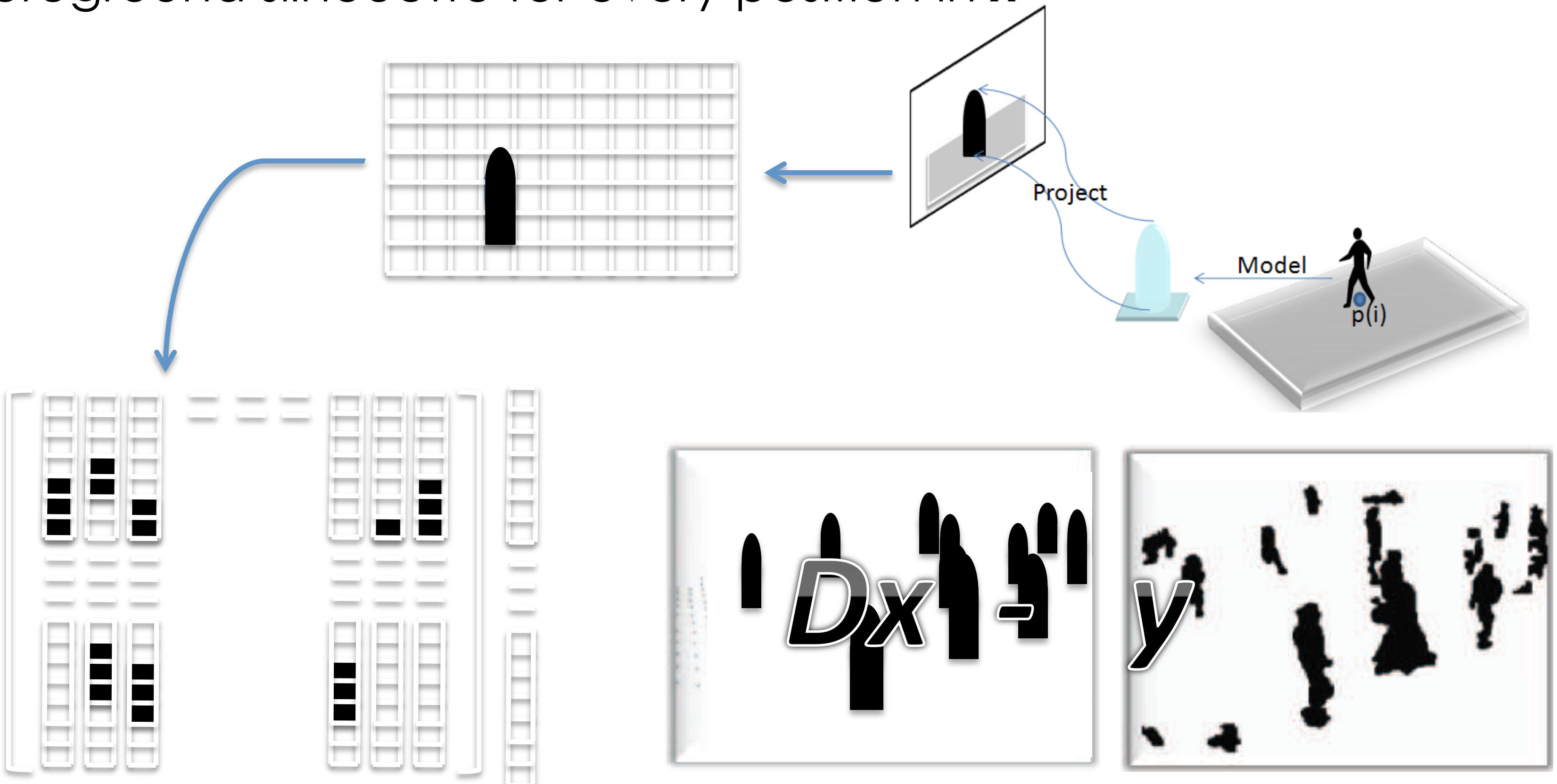
- Severely degraded foreground silhouettes
- Spatially dense distribution
- Strong occlusions

I. Detection: Pedestrian localization



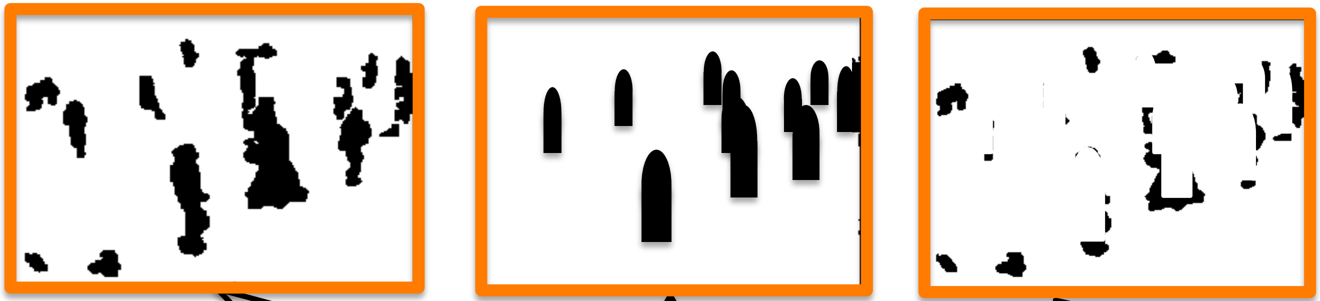
I. Detection: Calibrated Camera

- Create a dictionary ***D*** of atoms approximating the ideal foreground silhouette for every position in ***x***



Dictionary ***D*** ***x***

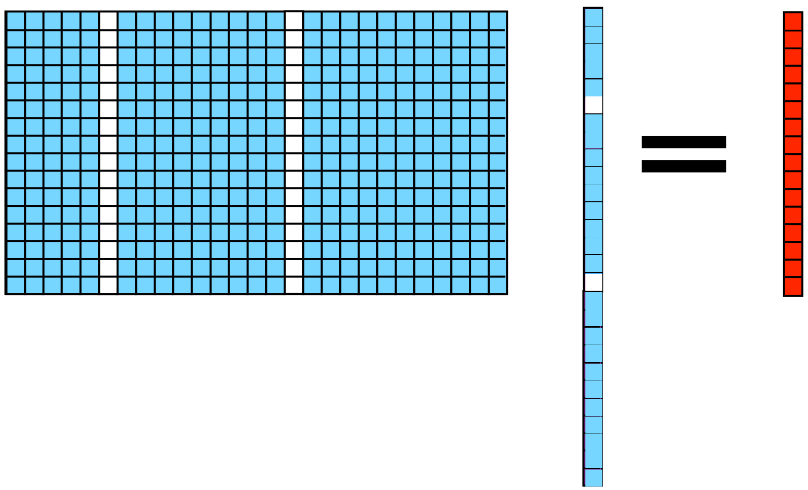
I. Detection: Sparsity driven framework



- Inverse problem: $y = Dx + n$

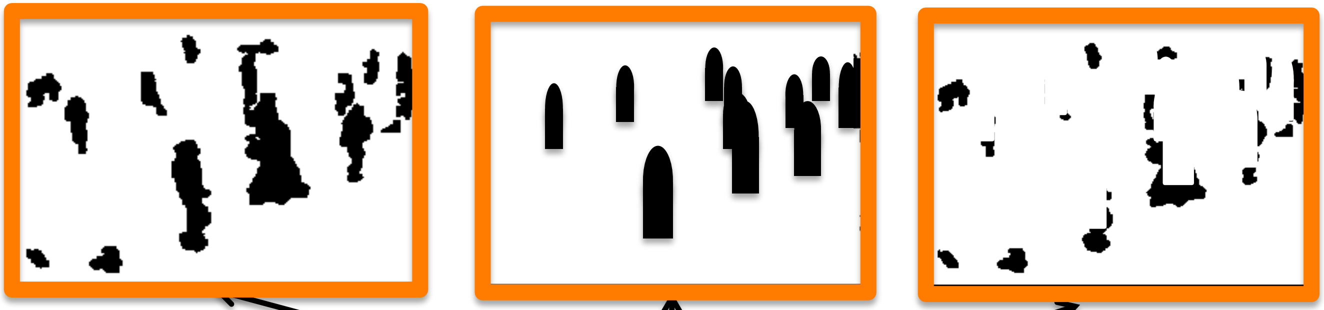
I. Detection: Greedy approach

$$D \quad x = y$$



[1] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," IEEE Transactions on signal processing,1993.

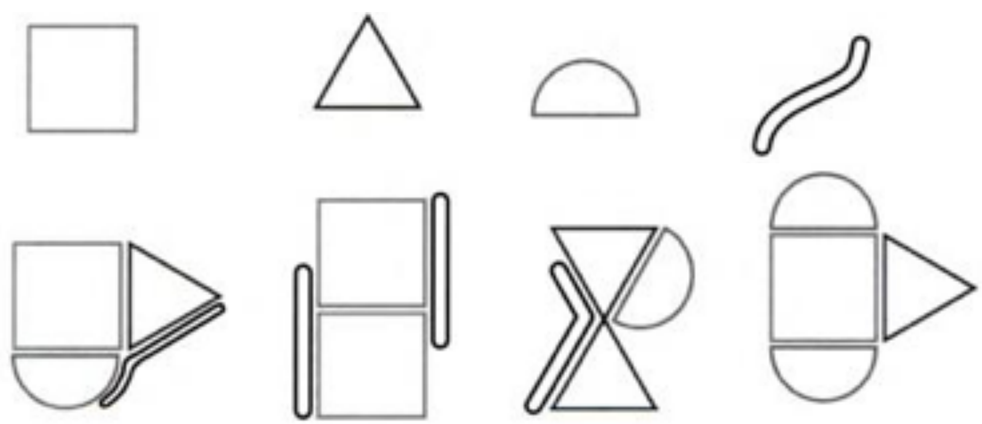
I. Detection: Sparsity driven framework



- Inverse problem: $y = Dx + n$
- Sparsity prior: $\min \|x\|_0$ s. t. $y = Dx + n$

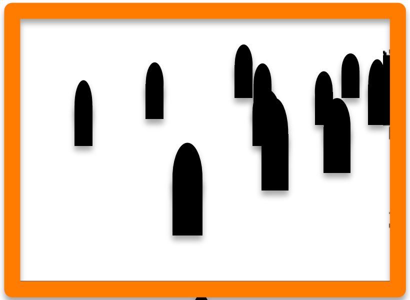
I. Detection: In praise of sparsity

"Creation is based on small number of primary, indivisible elements that combine with one another according to a few simple patterns." [1]



[1] B. Elahi, "Spirituality is a science: Foundation of Natural Spirituality"

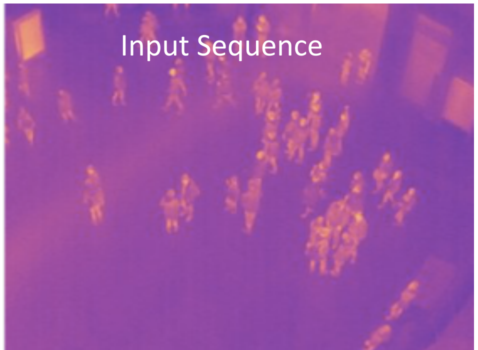
I. Detection: Sparsity driven framework


- Inverse problem: $y = Dx + n$
- Sparsity prior: $\min \|x\|_0$ s. t. $y = Dx + n$
- Basis Pursuit [1]: $\min \|x\|_1$ s. t. $y = Dx$
- BPDN: $\min \|x\|_1$ s. t $\|y - Dx\| \leq n$
- Lasso: $\min \|y - Dx\|_2$ s. t $\|x\|_1 \leq \varepsilon$

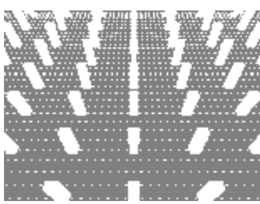
[1] S. S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” SIAM Journal on Scientific Computing

I. Detection: Pedestrian localization

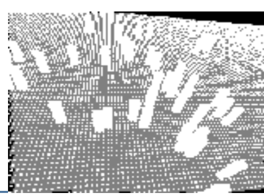


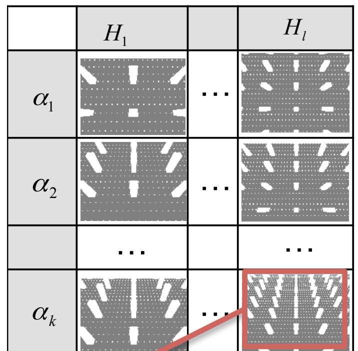
- Foreground extraction

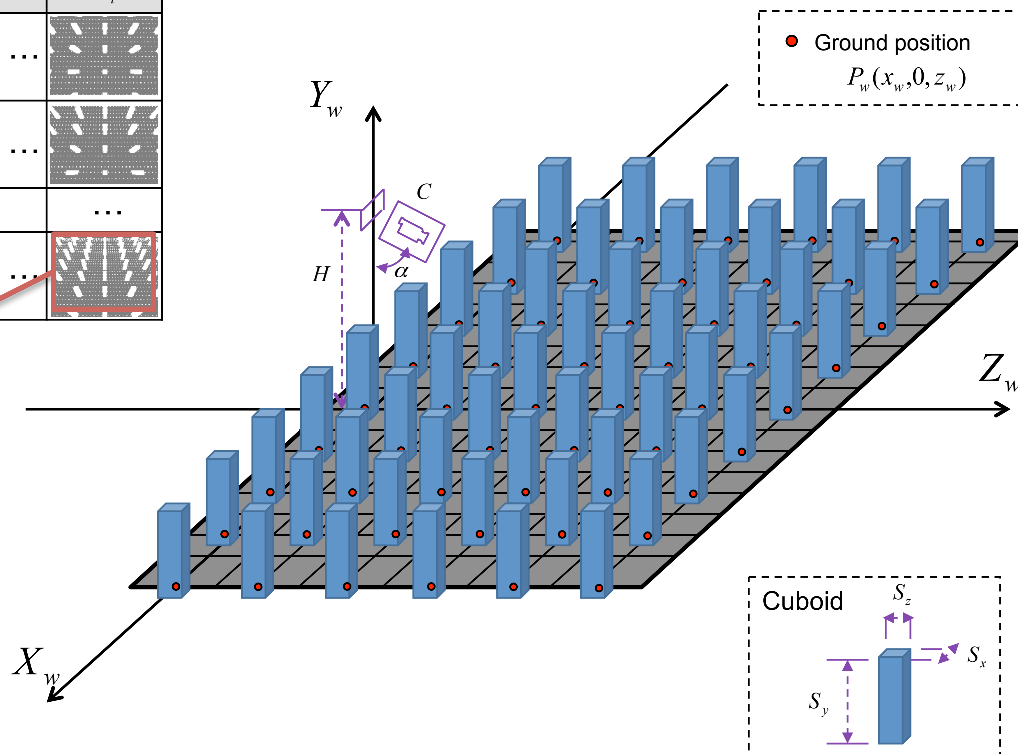
- Used Dictionary



- Localization



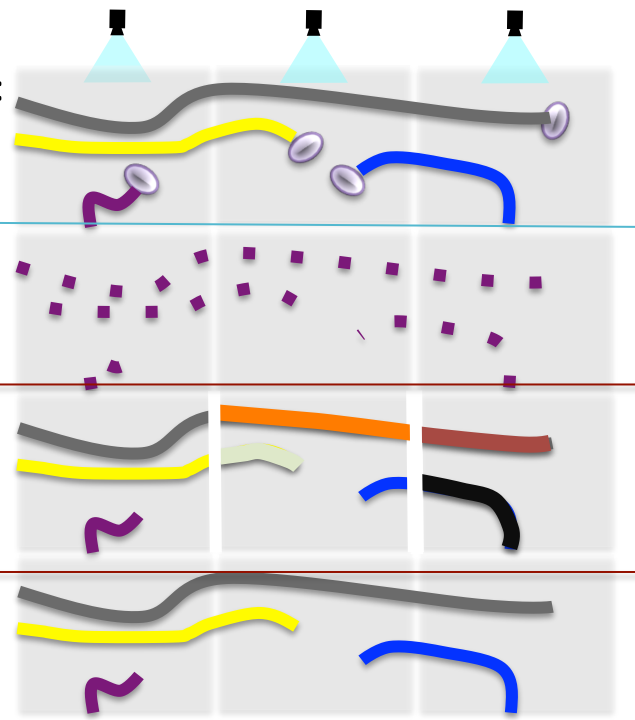




Outline:

From Foreground Extraction **To** Tracking 42 million Pedestrians

Network of 3 sensors:
4 people walking
(top view)



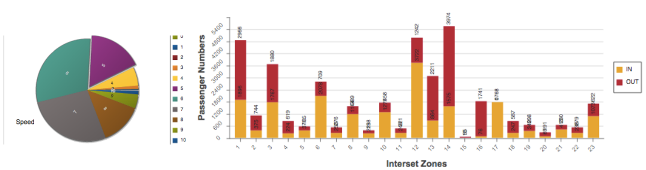
>100 GB/day

1 Detection intra-sensor — 5GB/day

2 Tracklet Generation — 4 GB/day

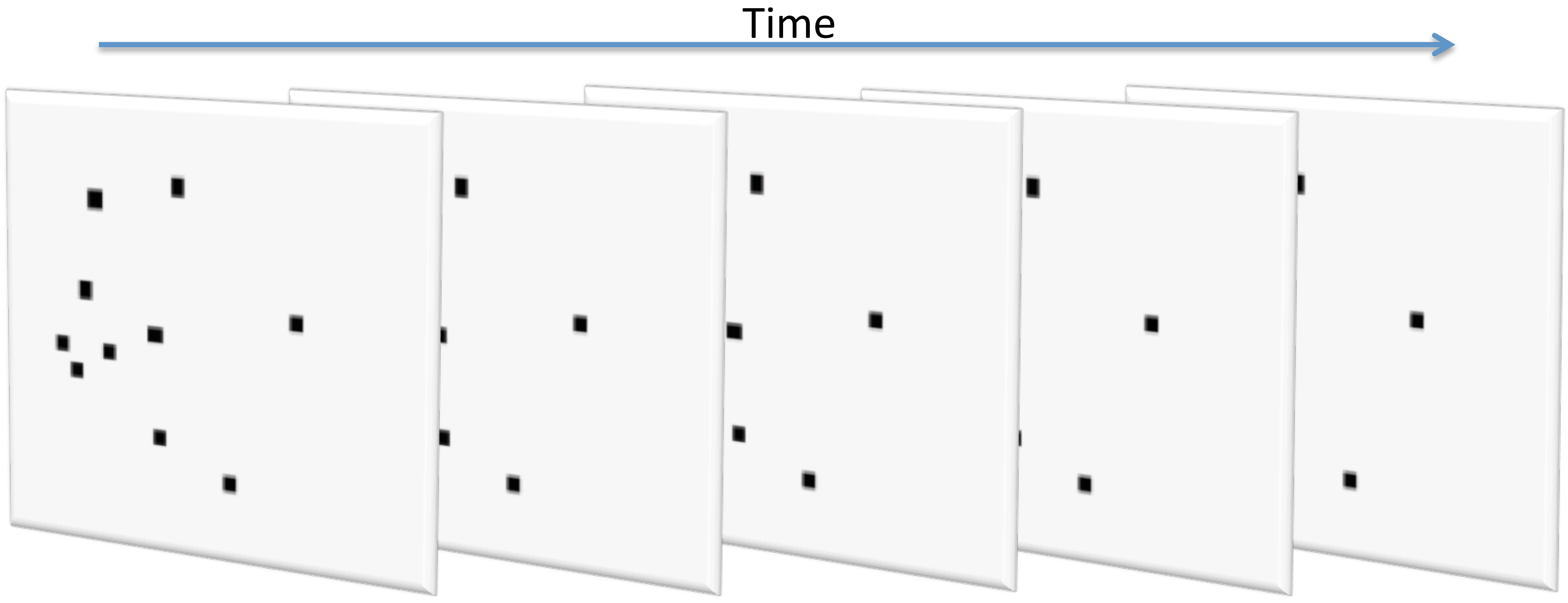
3 Tracklet Association — 5 GB/day

4 Statistic Generation — 0.05 GB/day

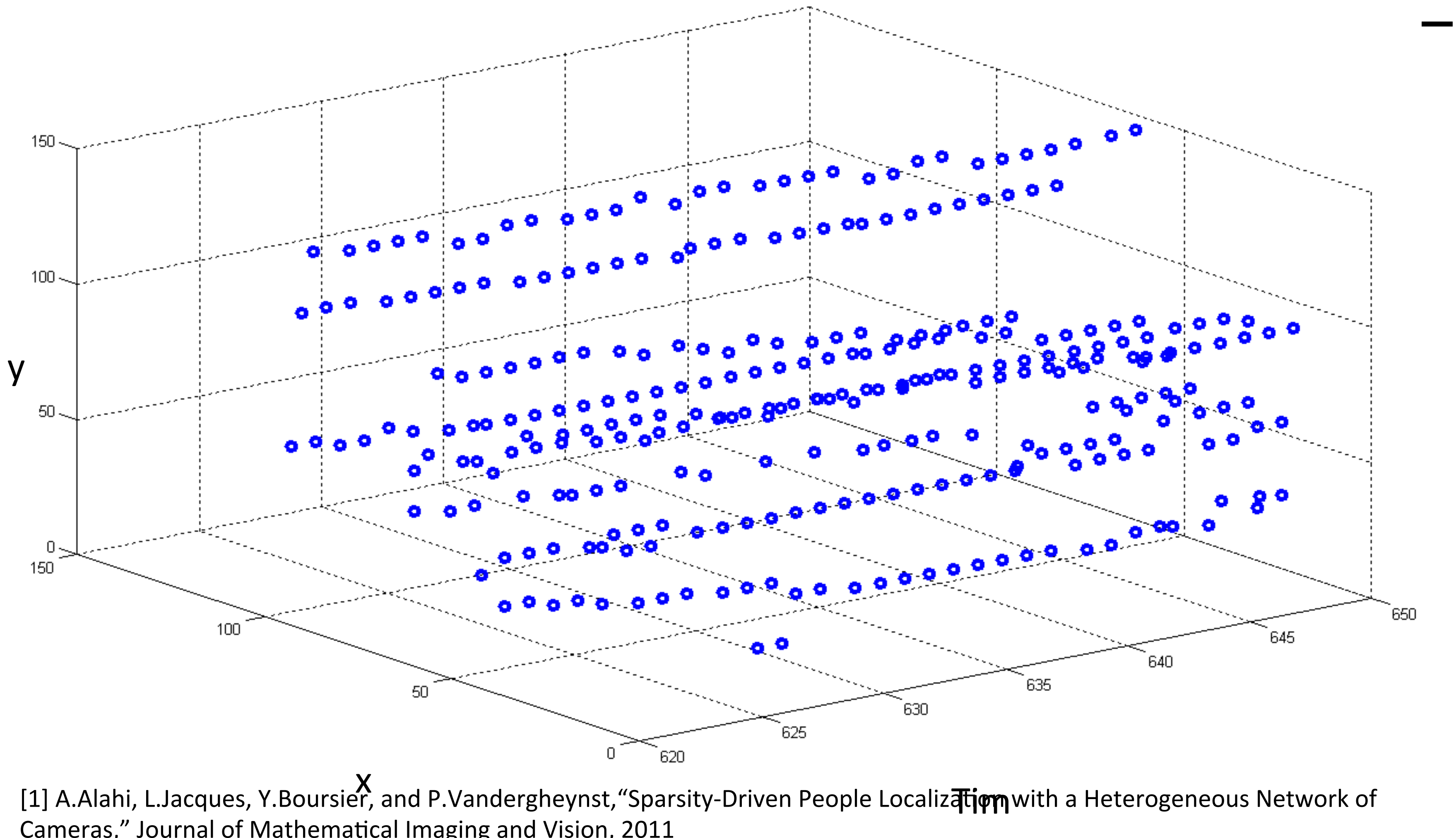


II. Tracklet generation: Data Association Problem

- Create a Directed Acyclic Graph G= (N,E) where
  - N= The detected ground plane points across time
  - E= The connectivity cost between the detections (based on motion/appearance model)

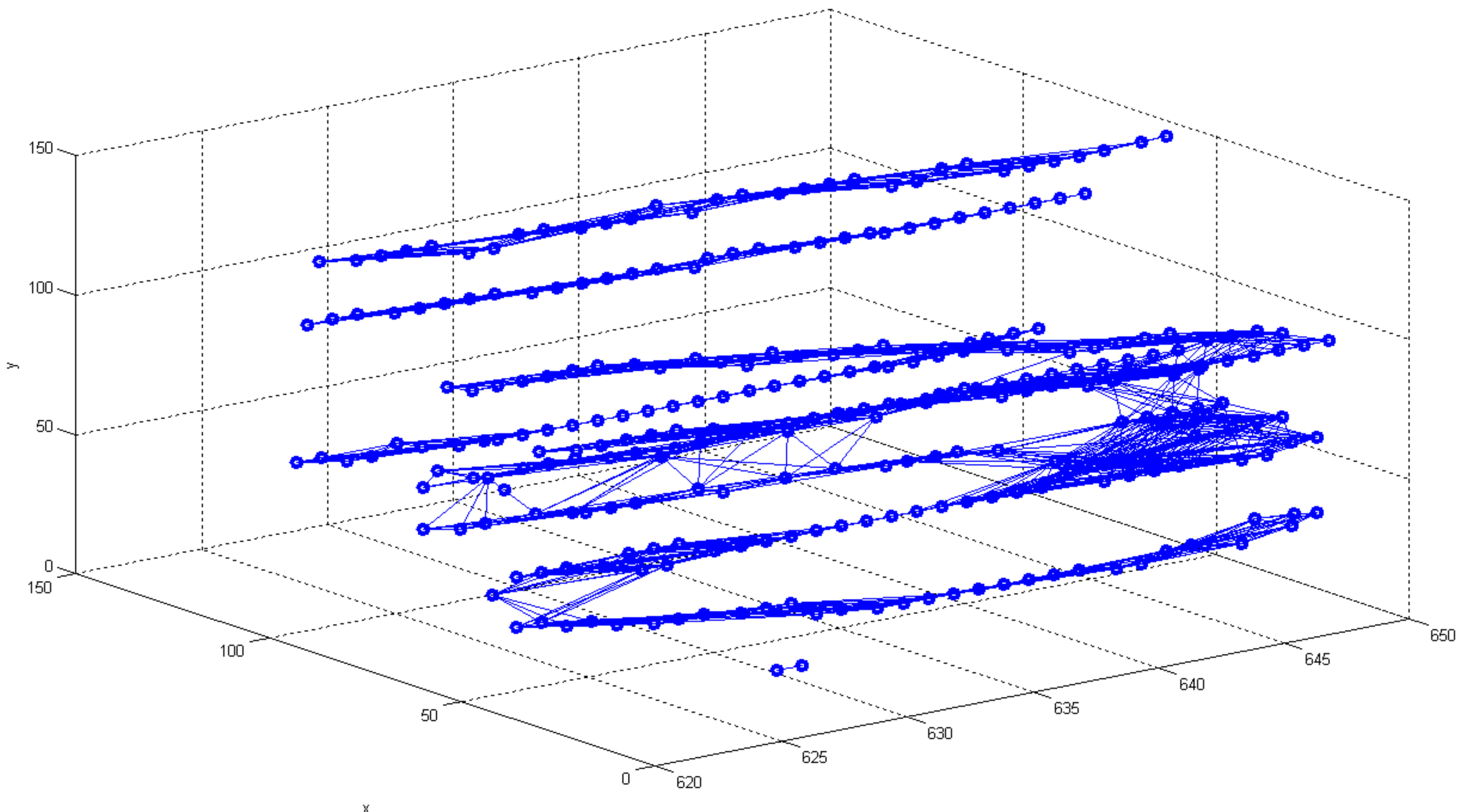


II. Tracklet generation: Data Association Problem



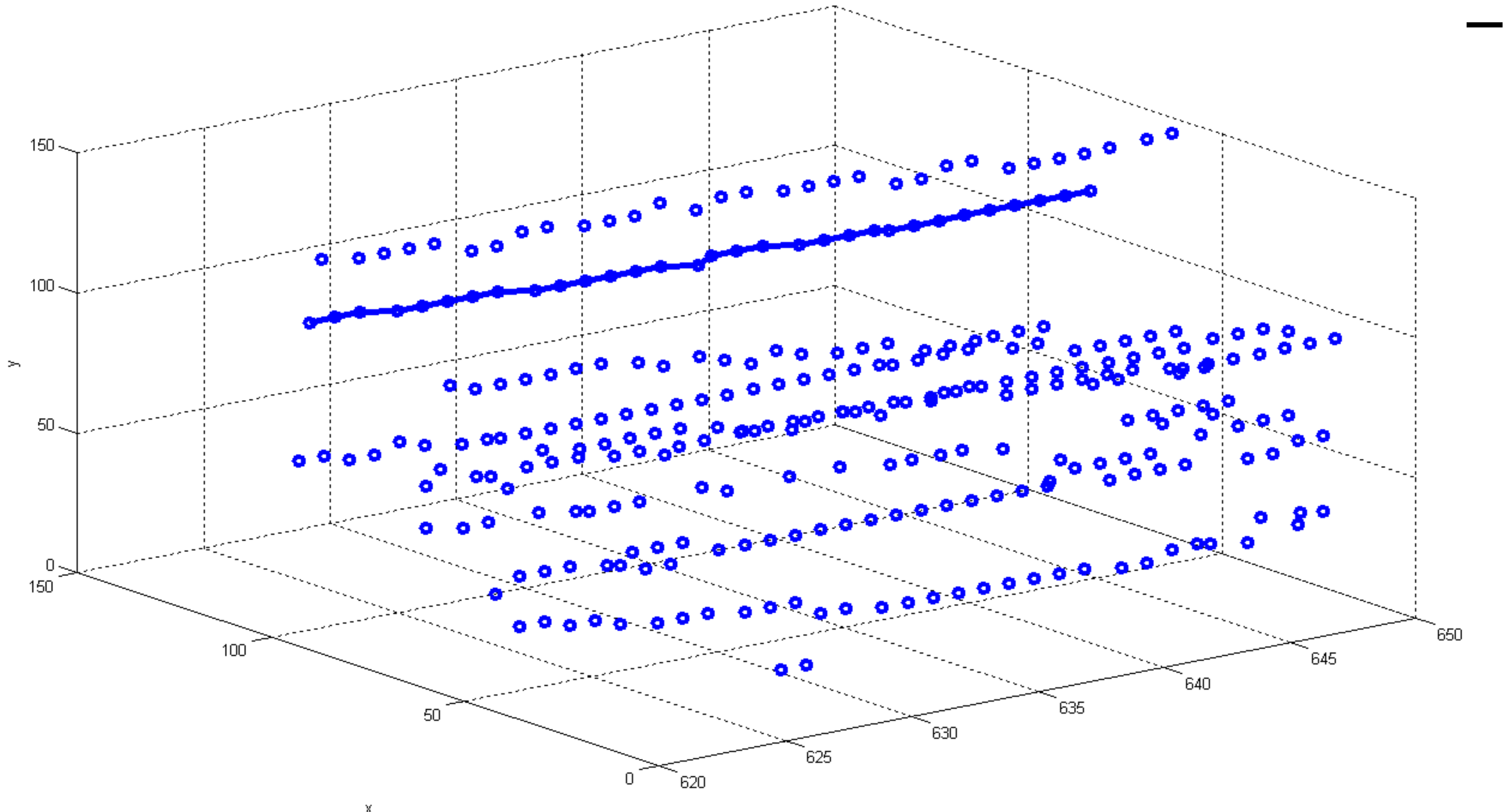
[1] A.Alahi, L.Jacques, Y.Boursier, and P.Vandergheynst,"Sparsity-Driven People Localization with a Heterogeneous Network of Cameras," Journal of Mathematical Imaging and Vision, 2011
[2] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multi-Camera People Tracking With a Probabilistic Occupancy Map", PAMI 2008

II. Tracklet generation: Create a DAG



[1] A.Alahi, L.Jacques, Y.Boursier, and P.Vandergheynst,"Sparsity-Driven People Localization with a Heterogeneous Network of Cameras," Journal of Mathematical Imaging and Vision, 2011
[2] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multi-Camera People Tracking With a Probabilistic Occupancy Map", PAMI 2008

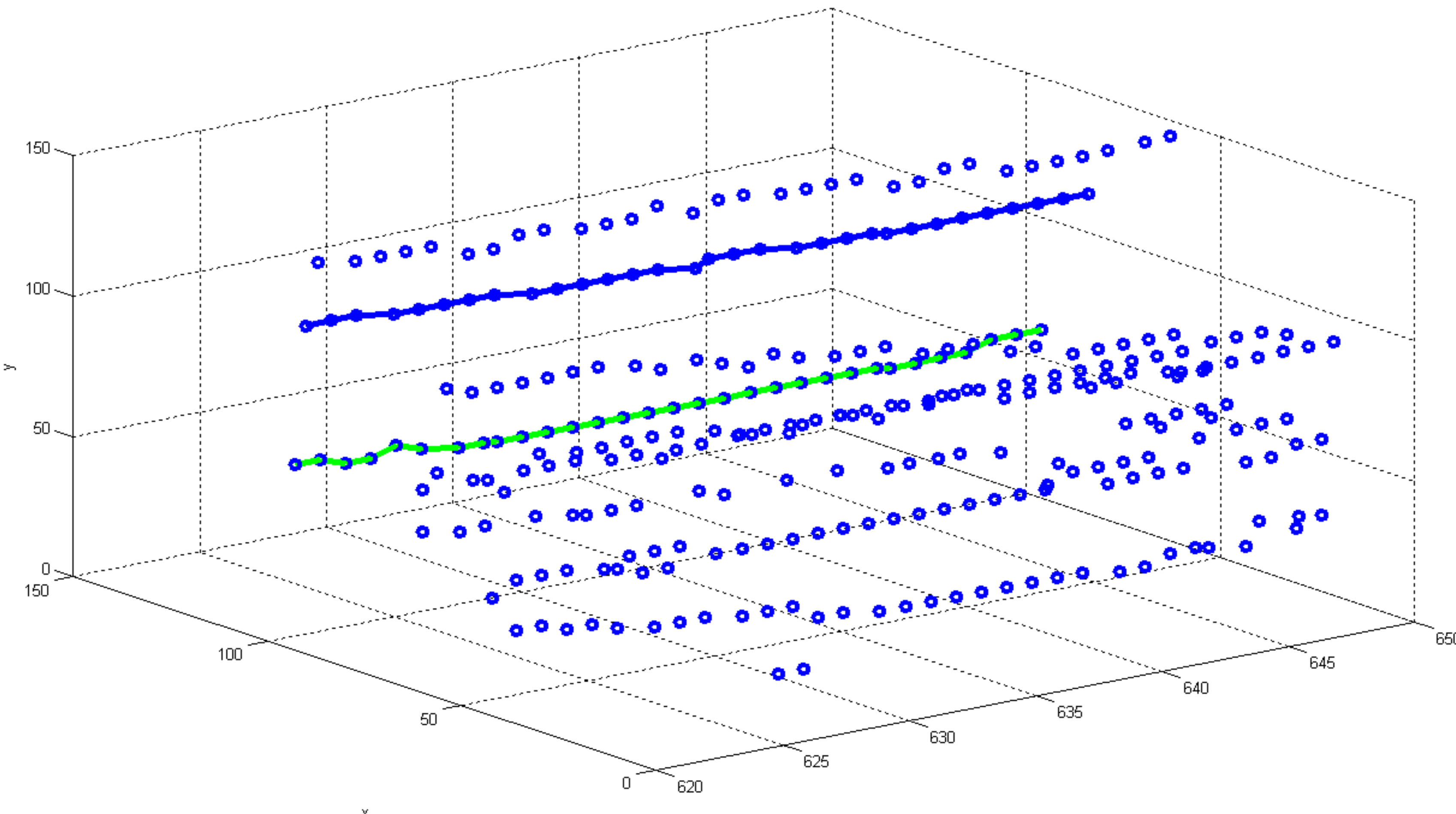
II. Tracklet generation: Select longest shortest path with smallest cost



[1] A.Alahi, L.Jacques, Y.Boursier, and P.Vandergheynst,"Sparsity-Driven People Localization with a Heterogeneous Network of Cameras," Journal of Mathematical Imaging and Vision, 2011

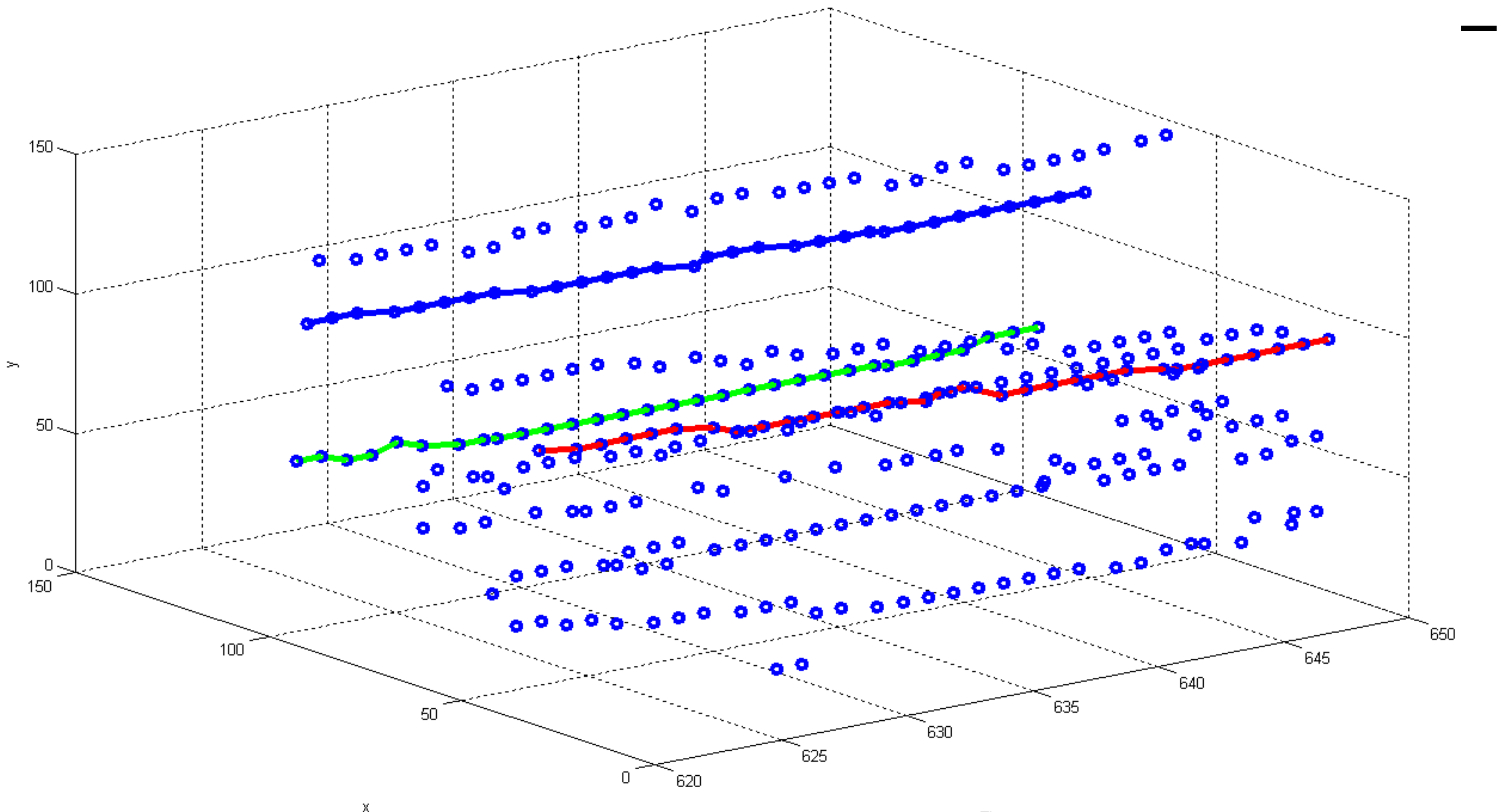
[2] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multi-Camera People Tracking With a Probabilistic Occupancy Map", PAMI 2008

II. Tracklet generation: Iterate



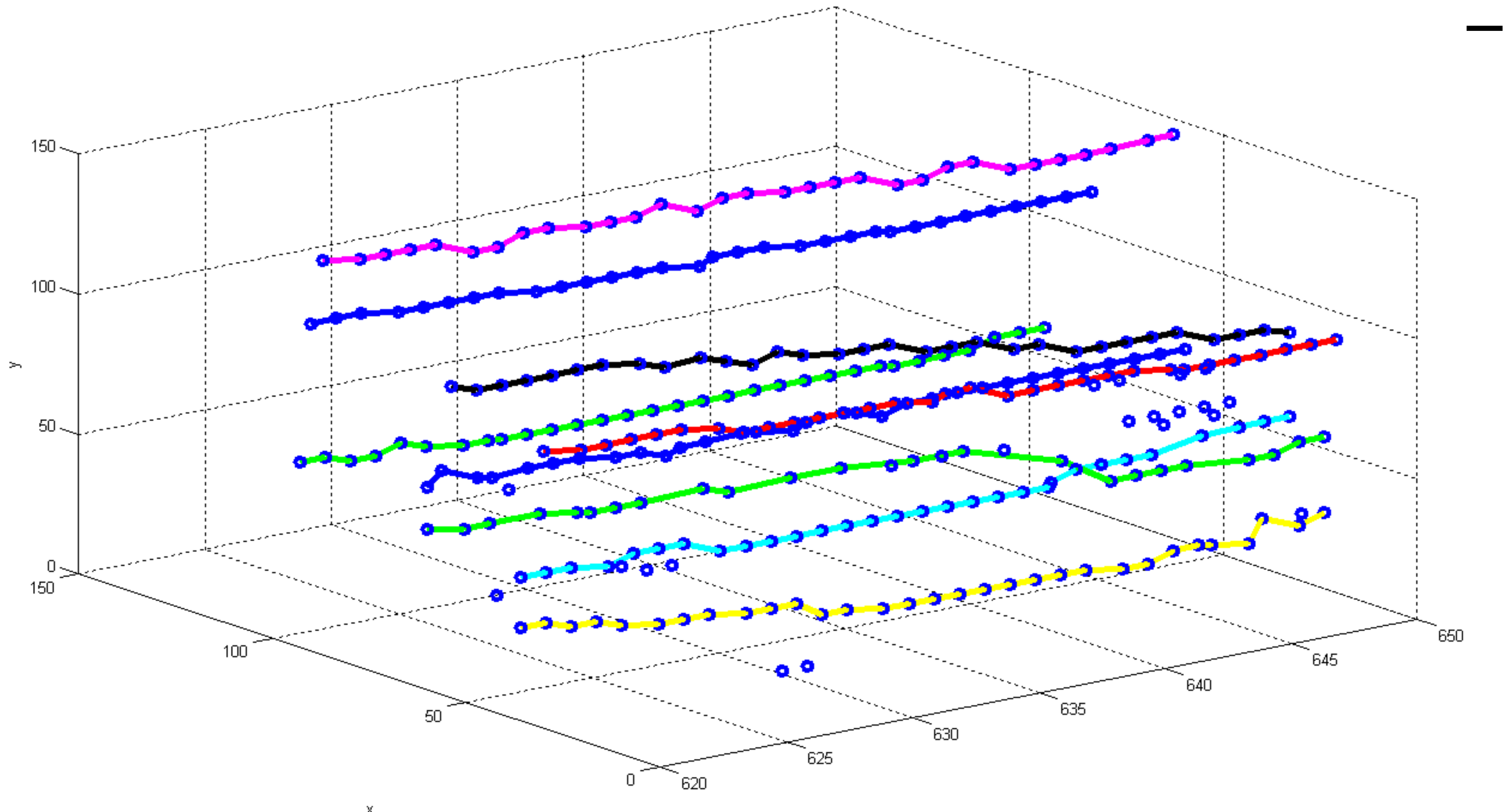
[1] A.Alahi, L.Jacques, Y.Boursier, and P.Vandergheynst,"Sparsity-Driven People Localization with a Heterogeneous Network of Cameras," Journal of Mathematical Imaging and Vision, 2011
[2] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multi-Camera People Tracking With a Probabilistic Occupancy Map", PAMI 2008

II. Tracklet generation: Iterate



[1] A.Alahi, L.Jacques, Y.Boursier, and P.Vandergheynst,“Sparsity-Driven People Localization with a Heterogeneous Network of Cameras,” Journal of Mathematical Imaging and Vision, 2011
[2] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, “Multi-Camera People Tracking With a Probabilistic Occupancy Map”, PAMI 2008

II. Tracklet generation: Till no more paths



[1] A.Alahi, L.Jacques, Y.Boursier, and P.Vandergheynst,"Sparsity-Driven People Localization with a Heterogeneous Network of Cameras," Journal of Mathematical Imaging and Vision, 2011
[2] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multi-Camera People Tracking With a Probabilistic Occupancy Map", PAMI 2008

II. Tracklet Generation: Edge cost

- Motion model with social interactions
- Appearance model

II. Tracklet generation: Modeling social interactions

$$\mathbf{F}_i = \mathbf{F}_i^{Goal} + \boxed{\mathbf{F}_i^{Avoidance}} + \mathbf{F}_i^{Attraction} + \mathbf{F}_i^{Scene}$$

$$\mathbf{F}_i^{Avoidance} = \sum_{j \in P \setminus i} \mathbf{f}_{j \to i}^{Avoidance},$$

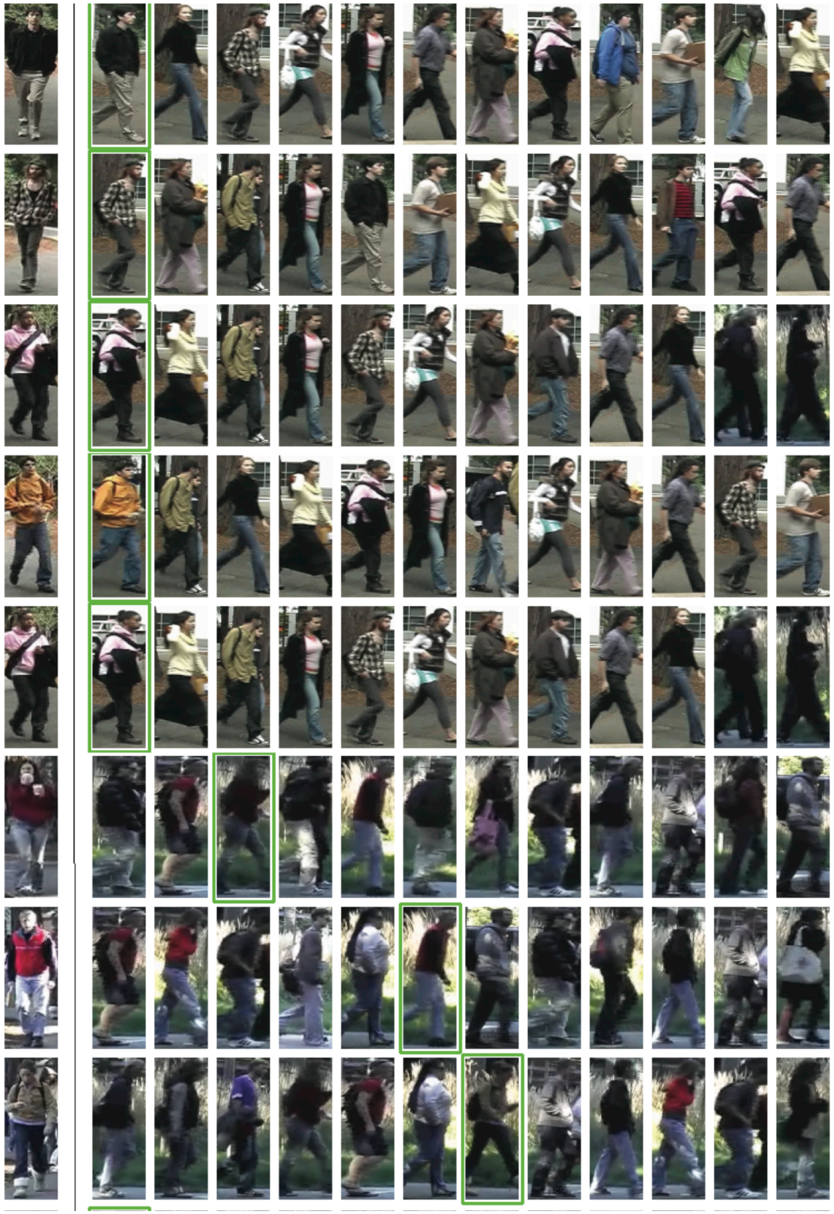
where

$$\mathbf{f}_{j \to i}^{Avoidance} = \alpha e^{\frac{d_p - d_{ij}}{\beta}} \mathbf{n}_{j \to i}$$

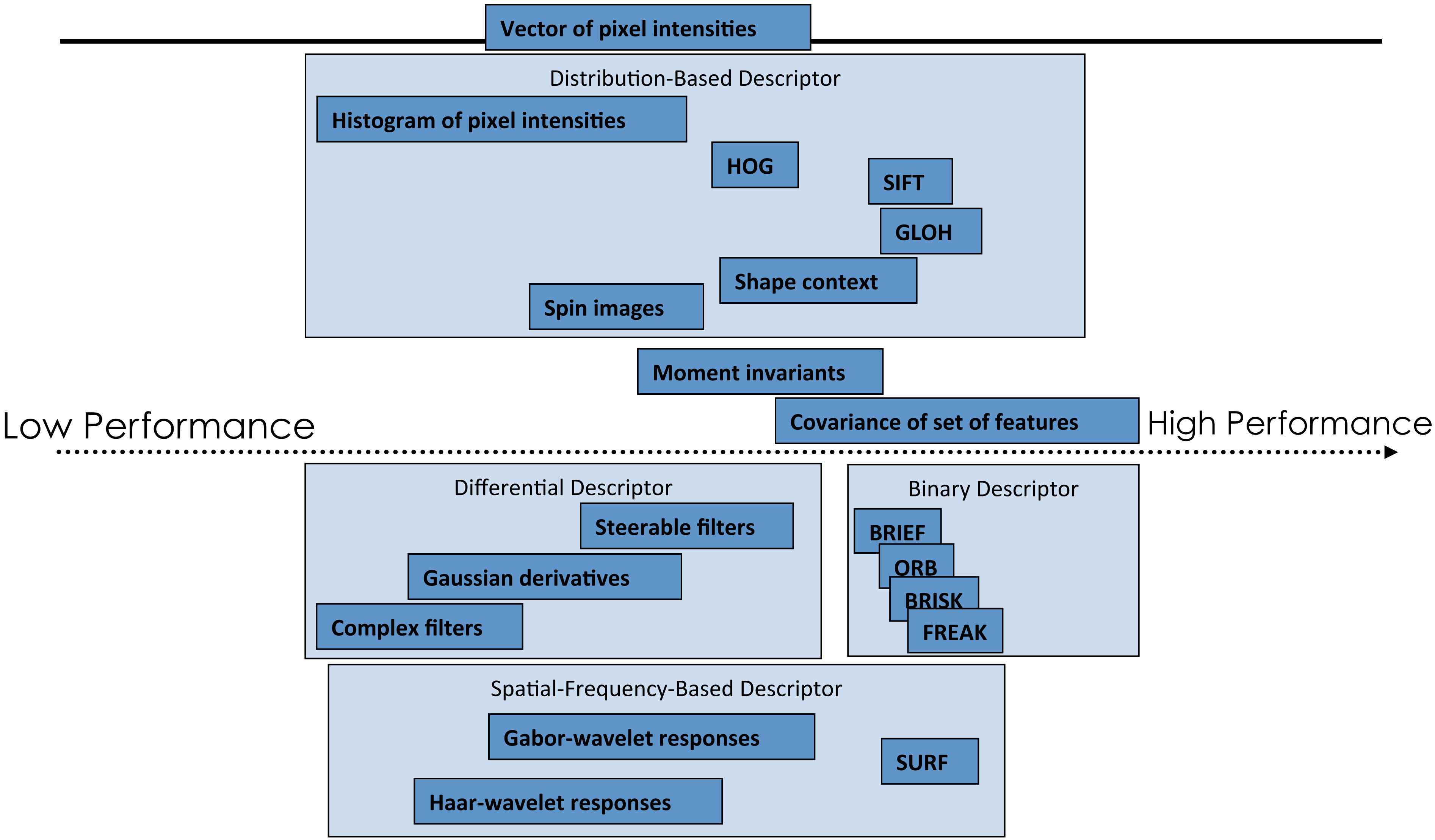
$$\frac{d}{dt}\mathbf{v} = \frac{\mathbf{F}_i}{m},$$

[1] Laura Leal-Taixé, Bodo Rosenhahn, "Pedestrian interaction in tracking: the social force model and global optimization methods", Springer 2012

II. Tracklet Generation: Modeling appearance cues

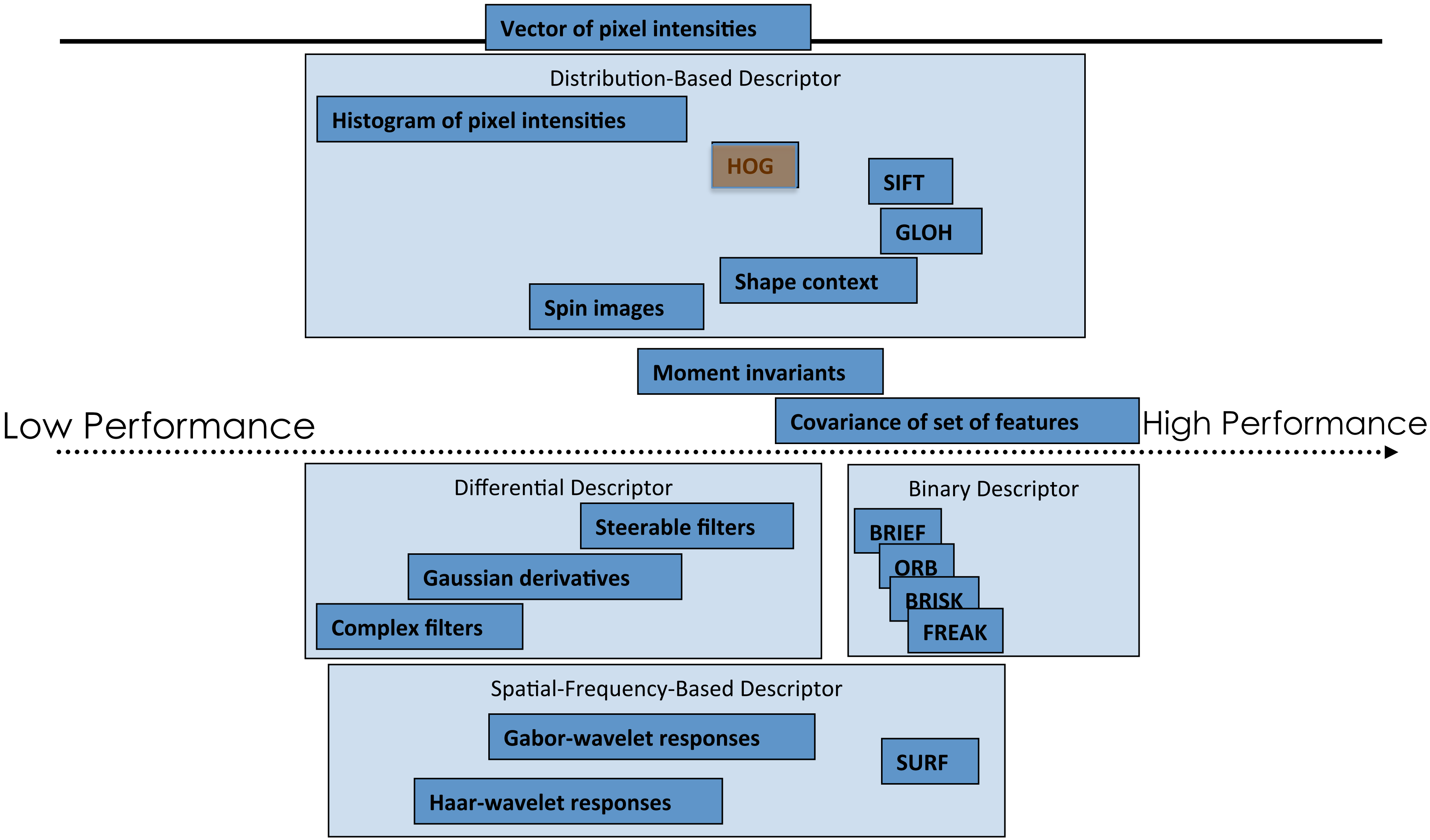


II. Tracklet Generation: An arm-race of image descriptors



[1] Gabriel, P., Hayet, J., Piater, J., Verly, J.: Object tracking using color interest points
[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection,",
[3] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification,".

II. Tracklet Generation: An arm-race of image descriptors



[1] Gabriel, P., Hayet, J., Piater, J., Verly, J.: Object tracking using color interest points
[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection,",
[3] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification,".

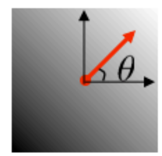
II. Tracklet Generation: HOG

Image gradient

- The gradient of an image: $\nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]$

$\nabla f = \left[\frac{\partial f}{\partial x}, 0\right]$

$\nabla f = \left[0, \frac{\partial f}{\partial y}\right]$



$\nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]$

The gradient points in the direction of most rapid increase in intensity

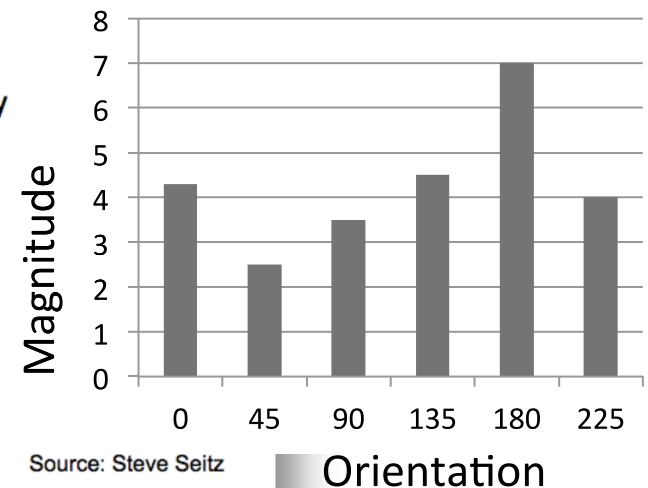
The gradient direction is given by $\boxed{\theta = \tan^{-1}\left(\frac{\partial f}{\partial y} / \frac{\partial f}{\partial x}\right)}$

- how does this relate to the direction of the edge?

The *edge strength* is given by the gradient magnitude

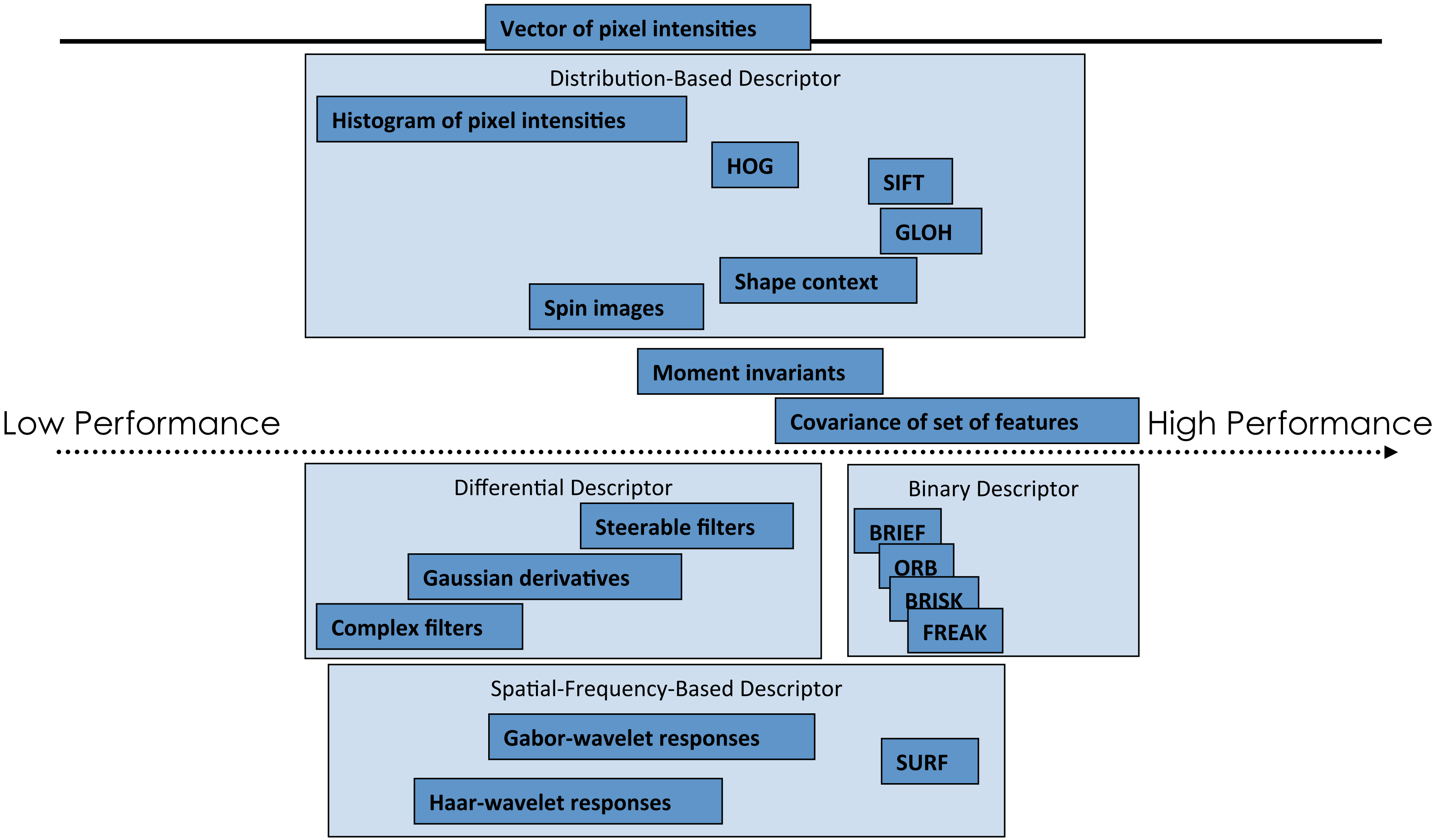
$$\boxed{\|\nabla f\| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}}$$

Fei-Fei Li　Lecture 5 - 18　25-Sep-13



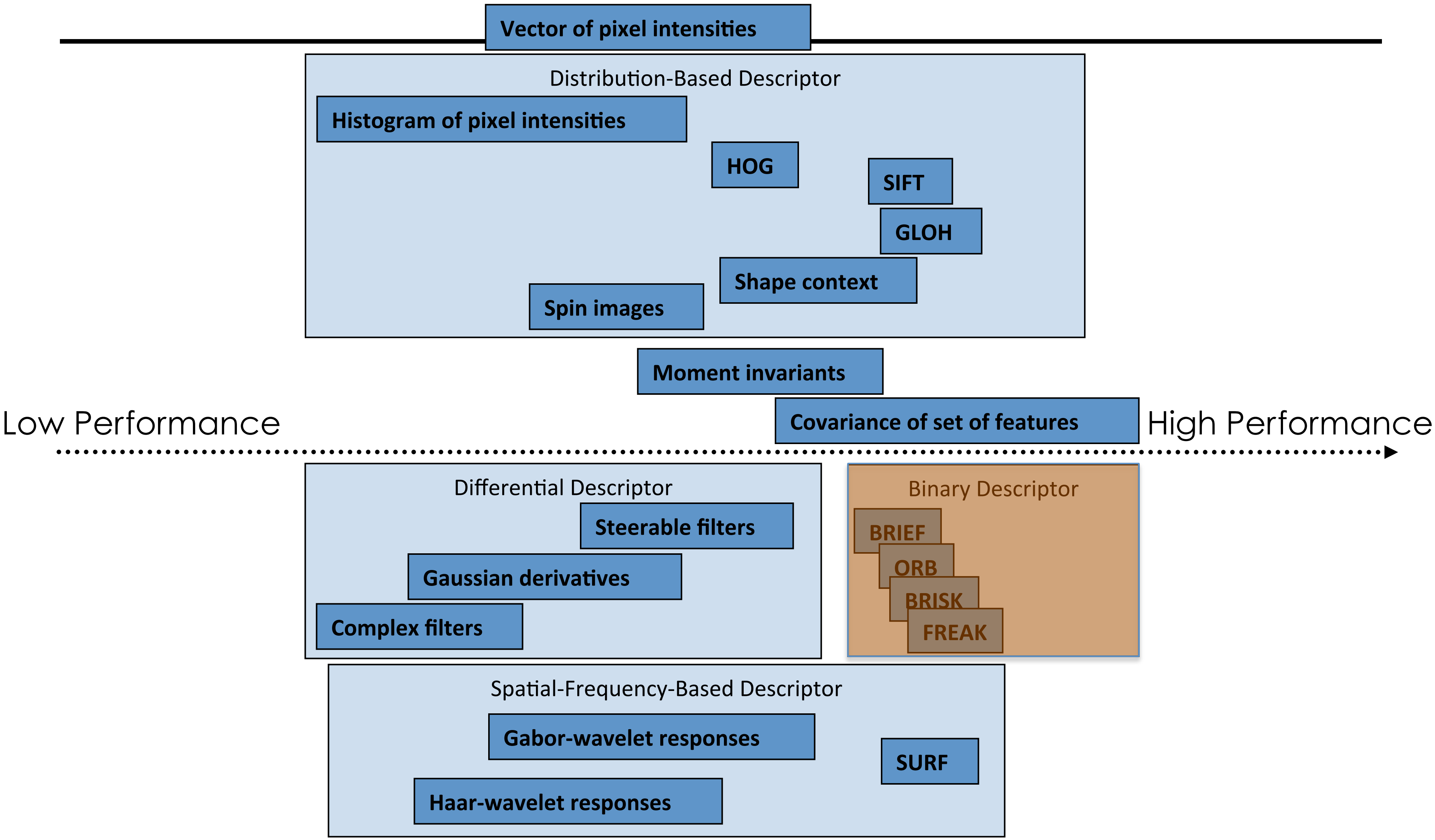
Source: Steve Seitz

II. Tracklet Generation: An arm-race of image descriptors



[1] Gabriel, P., Hayet, J., Piater, J., Verly, J.: Object tracking using color interest points
[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection,",
[3] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification,".

II. Tracklet Generation: An arm-race of image descriptors



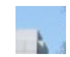
[1] Gabriel, P., Hayet, J., Plater, J., Verly, J.: Object tracking using color interest points
[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection,",
[3] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification,".

II. Tracklet Generation: Binary descriptors

II. Tracklet Generation: BRIEF[1] / ORB[2] / BRISK[3]

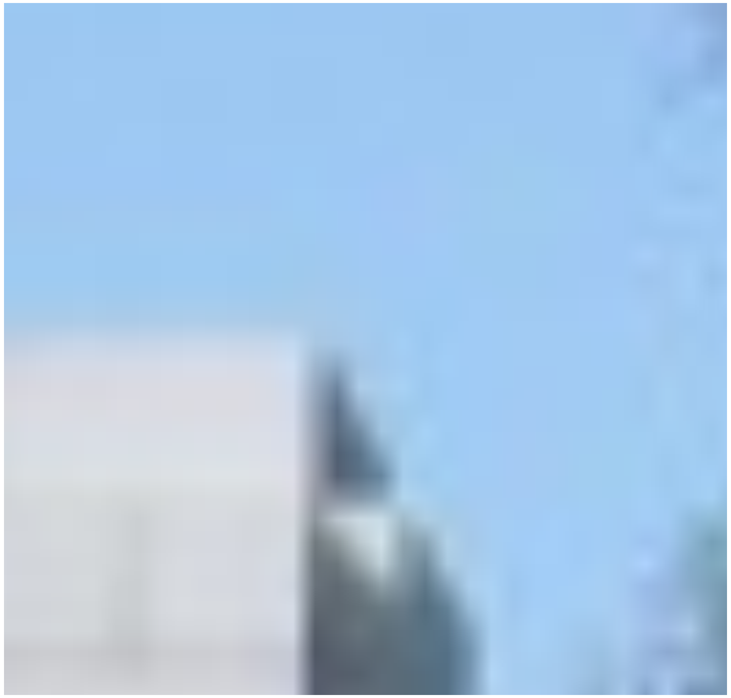
[1] Calonder, Michael, et al. "Brief: Binary robust independent elementary features." *ECCV 2010*.
[2] Rublee, Ethan, et al. "ORB: an efficient alternative to SIFT or SURF. " ICCV 2011
[3] Leutenegger,. , et al. "BRISK: Binary robust invariant scalable keypoints." ICCV 2011

II. Tracklet Generation: BRIEF[1] / ORB[2] / BRISK[3]



[1] Calonder, Michael, et al. "Brief: Binary robust independent elementary features." *ECCV 2010*.
[2] Rublee, Ethan, et al. "ORB: an efficient alternative to SIFT or SURF. " ICCV 2011
[3] Leutenegger,. , et al. "BRISK: Binary robust invariant scalable keypoints." ICCV 2011

II. Tracklet Generation: BRIEF[1] / ORB[2] / BRISK[3]

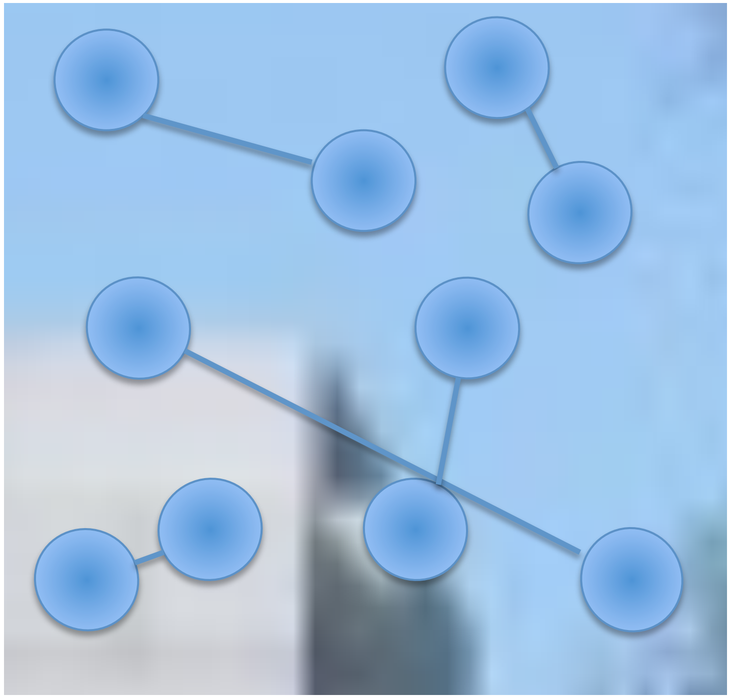


[1] Calonder, Michael, et al. "Brief: Binary robust independent elementary features." *ECCV 2010*.
[2] Rublee, Ethan, et al. "ORB: an efficient alternative to SIFT or SURF. " ICCV 2011
[3] Leutenegger,. , et al. "BRISK: Binary robust invariant scalable keypoints." ICCV 2011

II. Tracklet Generation: **BRIEF[1] / ORB[2]** / BRISK[3]

[1] Calonder, Michael, et al. "Brief: Binary robust independent elementary features." *ECCV 2010*.
[2] Rublee, Ethan, et al. "ORB: an efficient alternative to SIFT or SURF. " ICCV 2011
[3] Leutenegger,. , et al. "BRISK: Binary robust invariant scalable keypoints." ICCV 2011

II. Tracklet Generation: **BRIEF[1]** / ORB[2] / BRISK[3]



A sequence of 1-bit DoG

[1] Calonder, Michael, et al. "Brief: Binary robust independent elementary features." *ECCV 2010*.
[2] Rublee, Ethan, et al. "ORB: an efficient alternative to SIFT or SURF. " ICCV 2011
[3] Leutenegger,. , et al. "BRISK: Binary robust invariant scalable keypoints." ICCV 2011

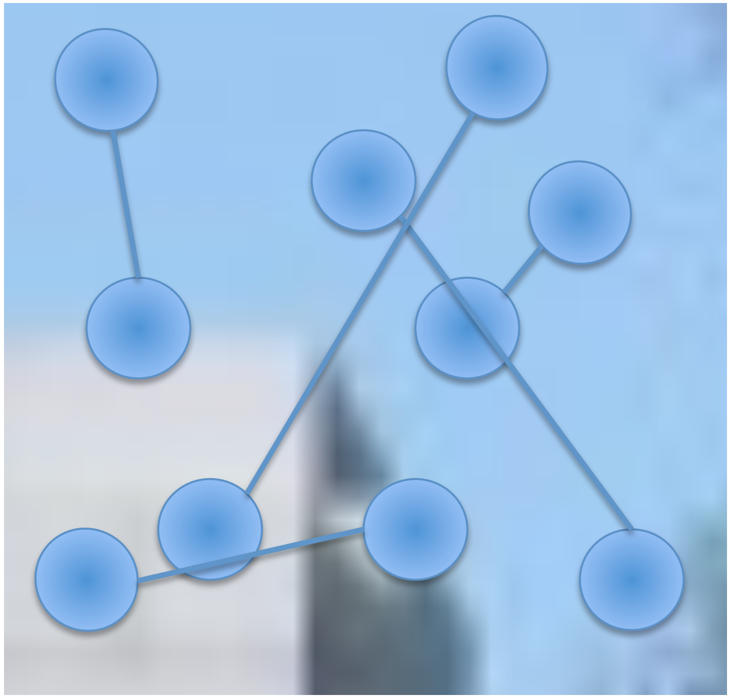
II. Tracklet Generation: BRIEF[1] / **ORB[2]** / BRISK[3]

- Select

1) Most discriminant

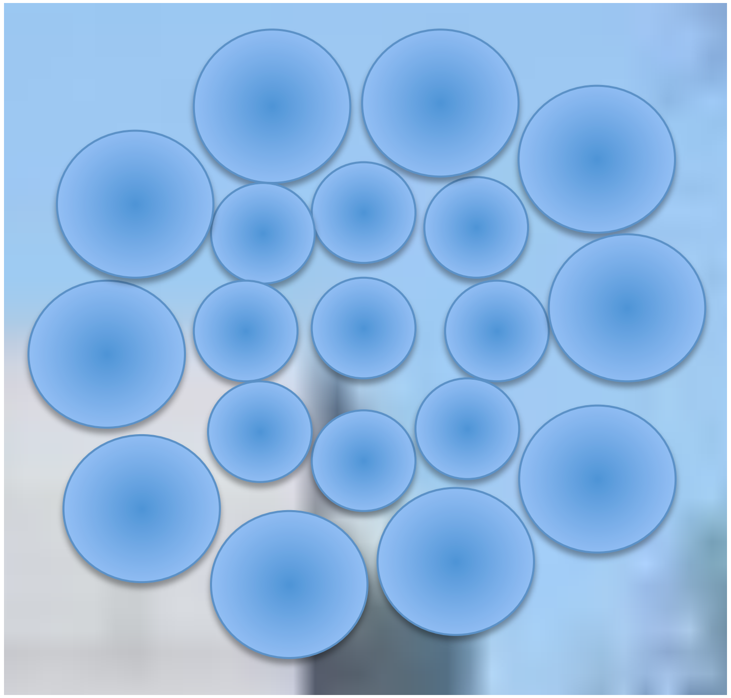
AND

2) Less correlated

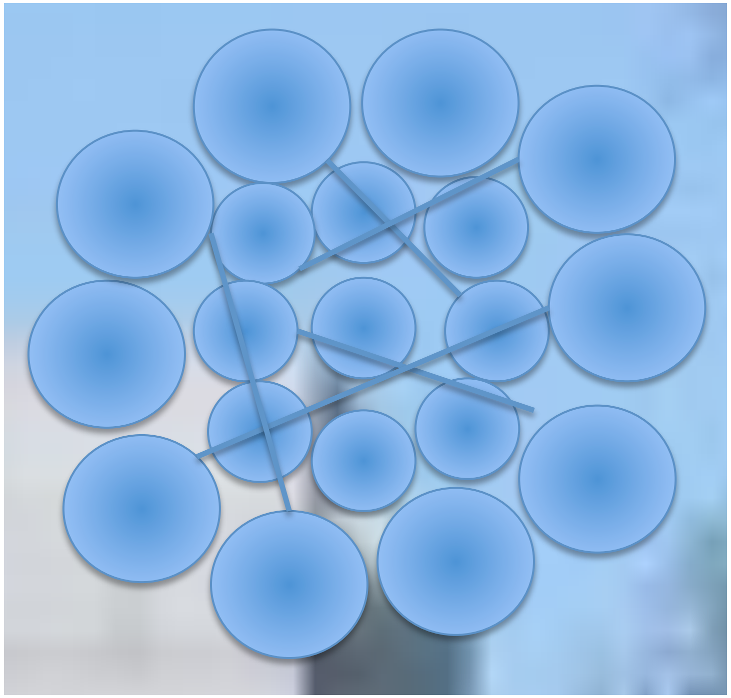


[1] Calonder, Michael, et al. "Brief: Binary robust independent elementary features." *ECCV 2010*.
[2] Rublee, Ethan, et al. "ORB: an efficient alternative to SIFT or SURF. " ICCV 2011
[3] Leutenegger,. , et al. "BRISK: Binary robust invariant scalable keypoints." ICCV 2011

II. Tracklet Generation: BRIEF[1] / ORB[2] / **BRISK[3]**

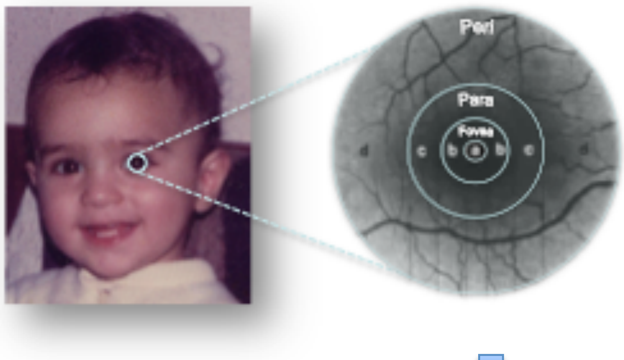


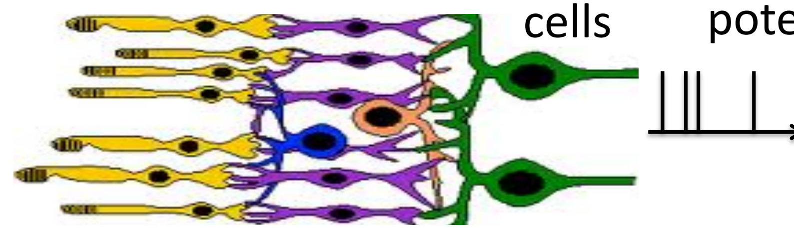
II. Tracklet Generation: BRIEF[1] / ORB[2] / **BRISK[3]**

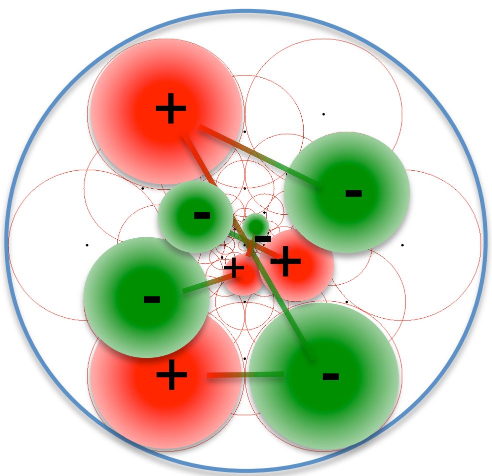


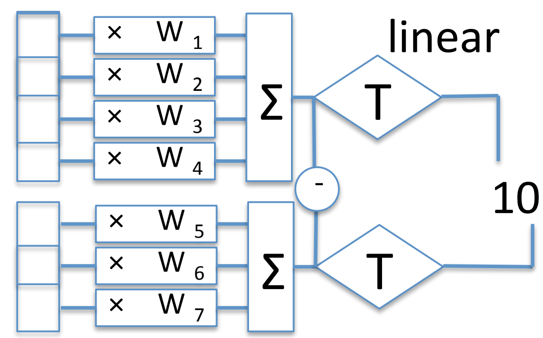
[1] Calonder, Michael, et al. "Brief: Binary robust independent elementary features." *ECCV 2010*.
[2] Rublee, Ethan, et al. "ORB: an efficient alternative to SIFT or SURF. " ICCV 2011
[3] Leutenegger,. , et al. "BRISK: Binary robust invariant scalable keypoints." ICCV 2011

II. Tracklet Generation: Retina-inspired [1]



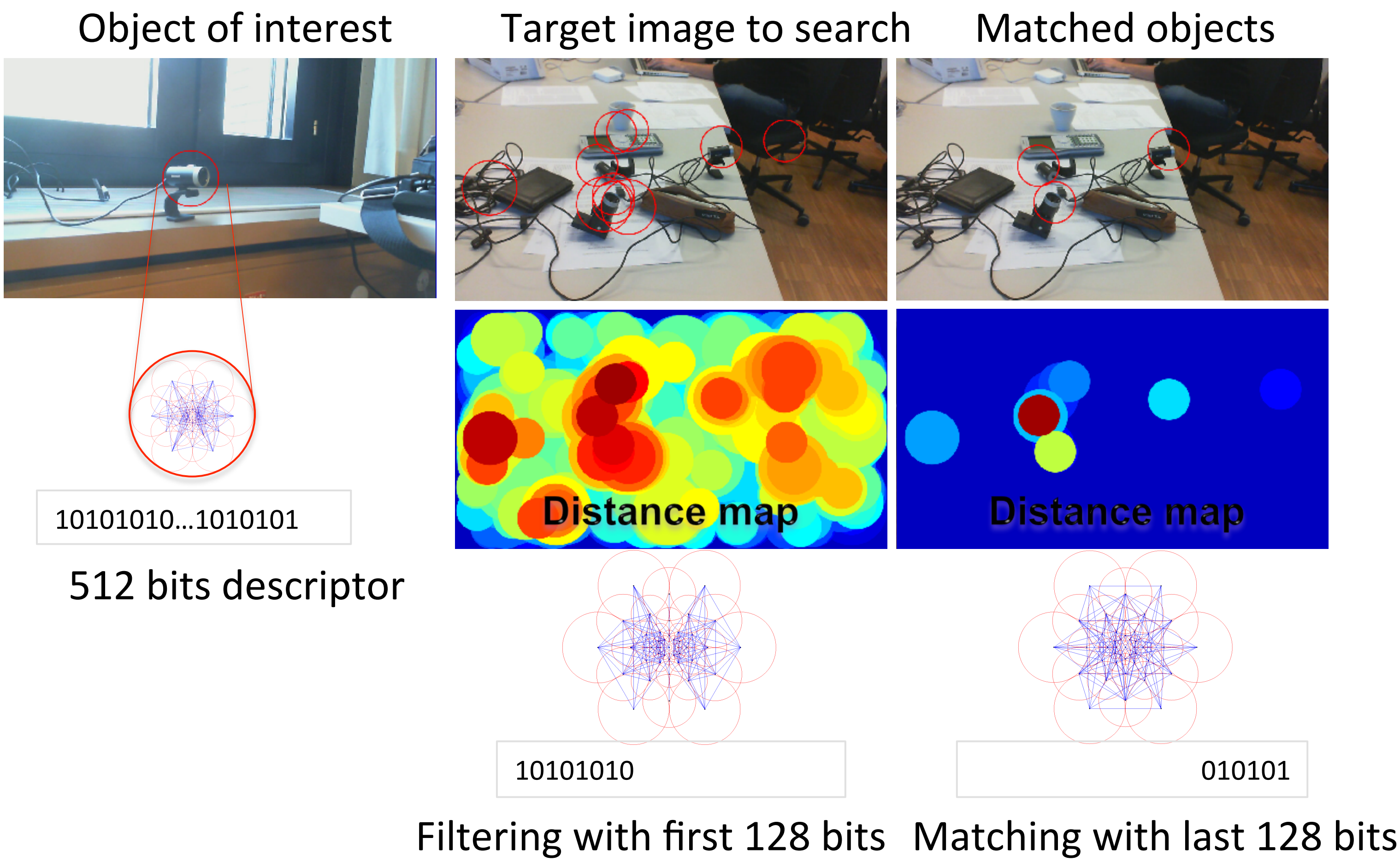






[1] Alahi, et al. "Freak: Fast retina keypoint." *CVPR 2012*

II. Tracklet Generation: Saccadic search



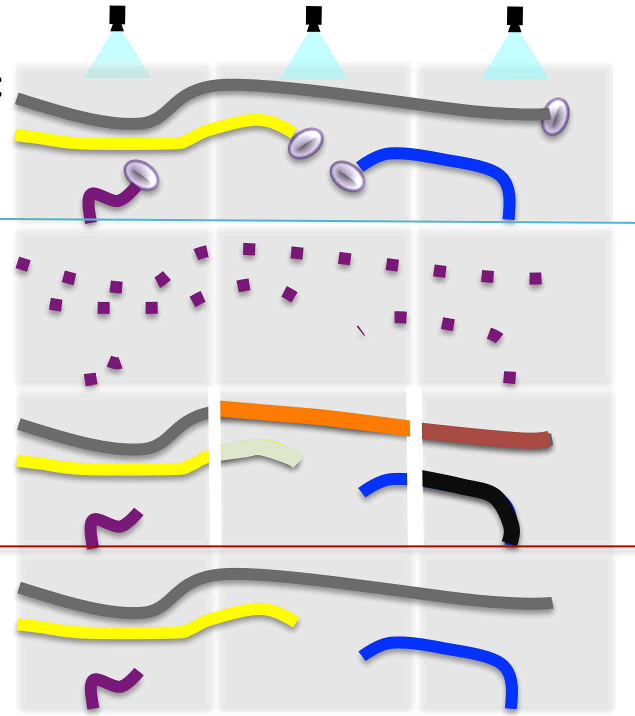
[1] Alahi, et al. "Freak: Fast retina keypoint." *CVPR 2012*

Tracking Example

Outline:

From Foreground Extraction **To** Tracking 42 million Pedestrians

Network of 3 sensors:
4 people walking
(top view)



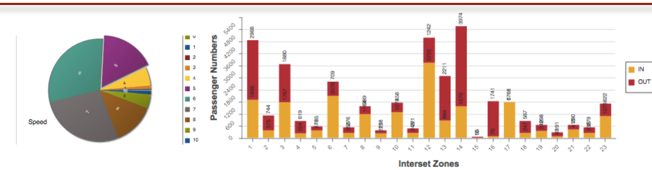
>100 GB/day

1 Detection intra-sensor — 5GB/day

2 Tracklet Generation — 4 GB/day

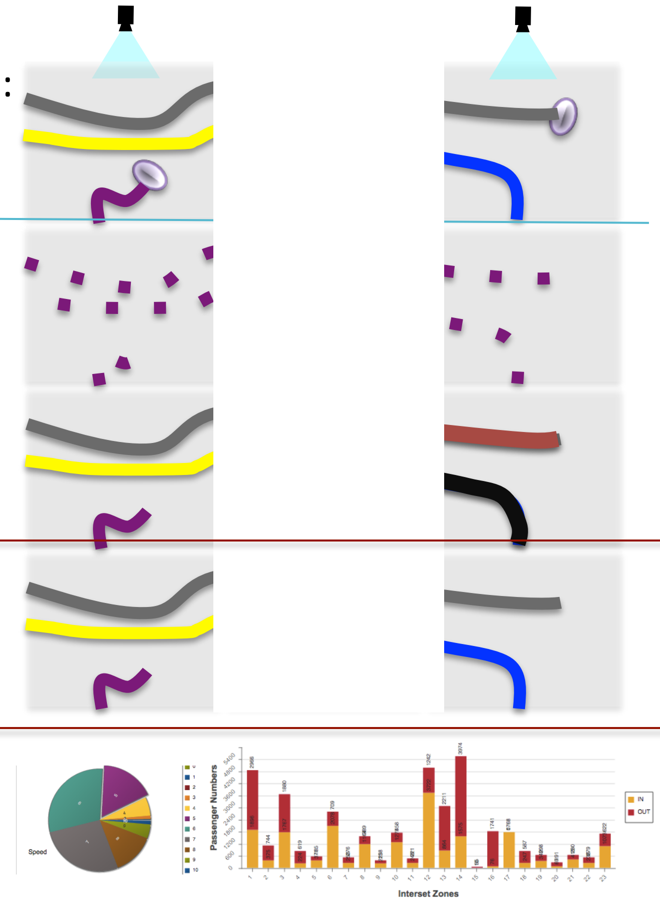
3 Tracklet Association — 5 GB/day

4 Statistic Generation — 0.05 GB/day



Tracklet association in scattered network

Network of 3 sensors:
4 people walking
(top view)



Network of 3 sensors: 4 people walking (top view)	>100 GB/day
Detection intra-sensor	5GB/day
Tracklet Generation	4 GB/day
Tracklet Association	5 GB/day
Statistic Generation	0.05 GB/day

STANFORD UNIVERSITY

III- Tracklet association: Problem formulation

Let:

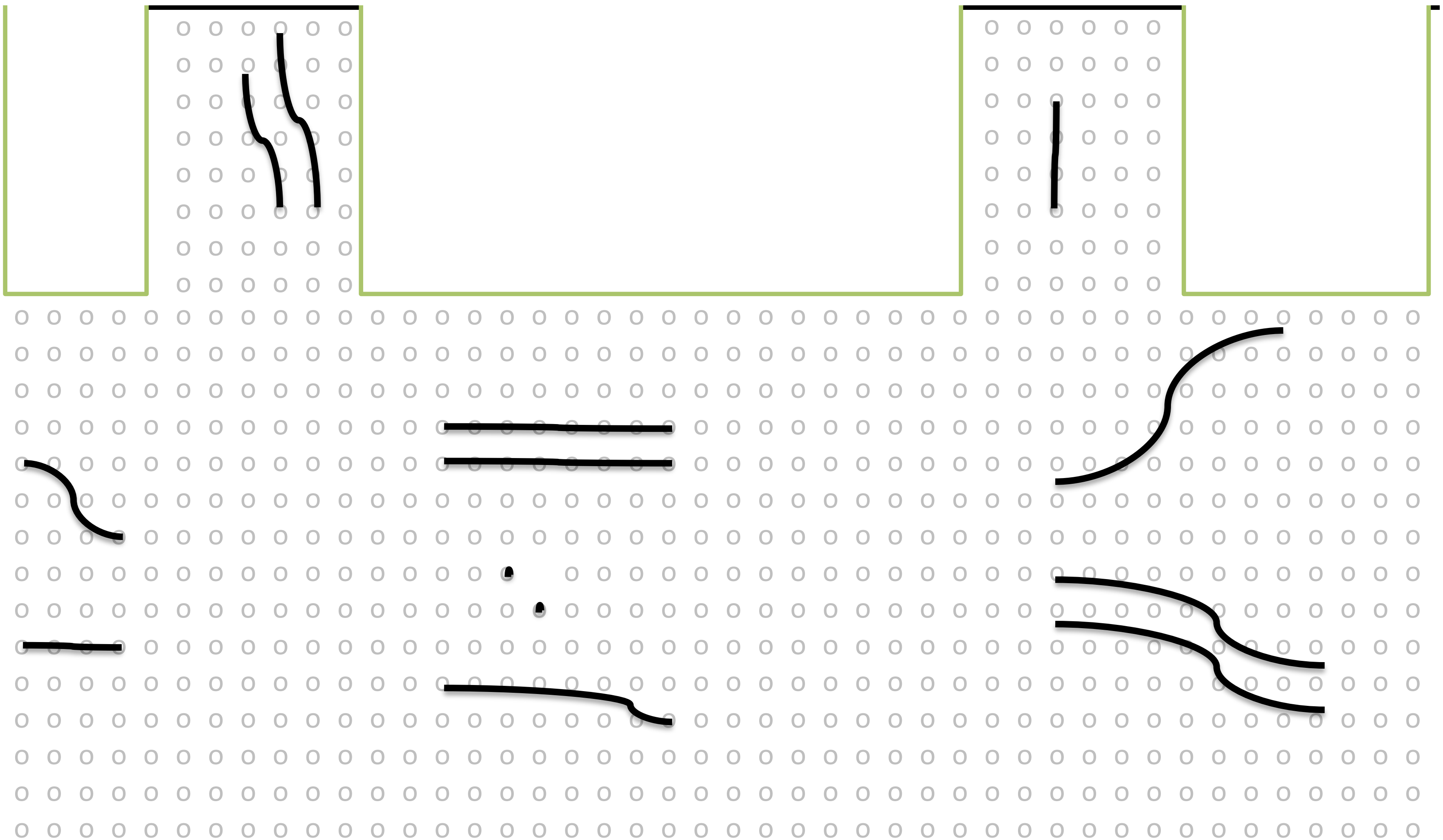
- **T**: all long term trajectories
- **t**: tracklets (tracklets capture within each camera)
- Problem: Maximizing the a posterior probability (MAP) of **T** :

$$\mathbf{T}^* = \arg\max_{T} P(\mathbf{T} \mid \mathbf{t}) \quad (1)$$
$$= \arg\max \prod_i P(\mathbf{t}_i \mid \mathbf{T}) P(\mathbf{T}) \quad (2),$$

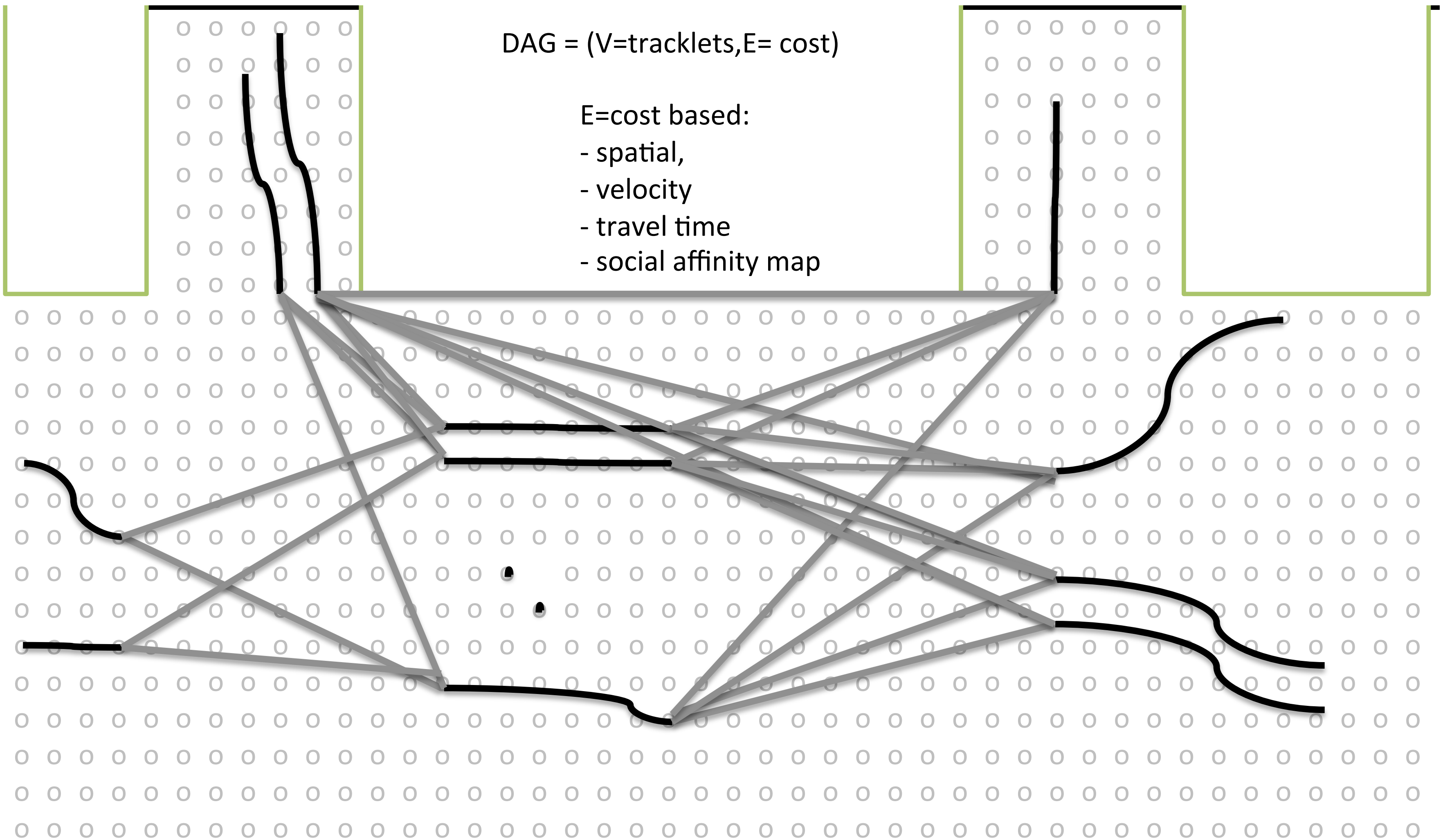
where $P(\mathbf{T}) = \prod_k P(\mathbf{T}_k)$ (since trajectories do not overlap)

$P(\mathbf{T}_k) = P(\mathbf{t}^s{}_k) \ldots P(\mathbf{t}^t{}_k \mid \mathbf{t}^{t-1}{}_k)\, P(\mathbf{t}^e{}_k)$ (markov chain)

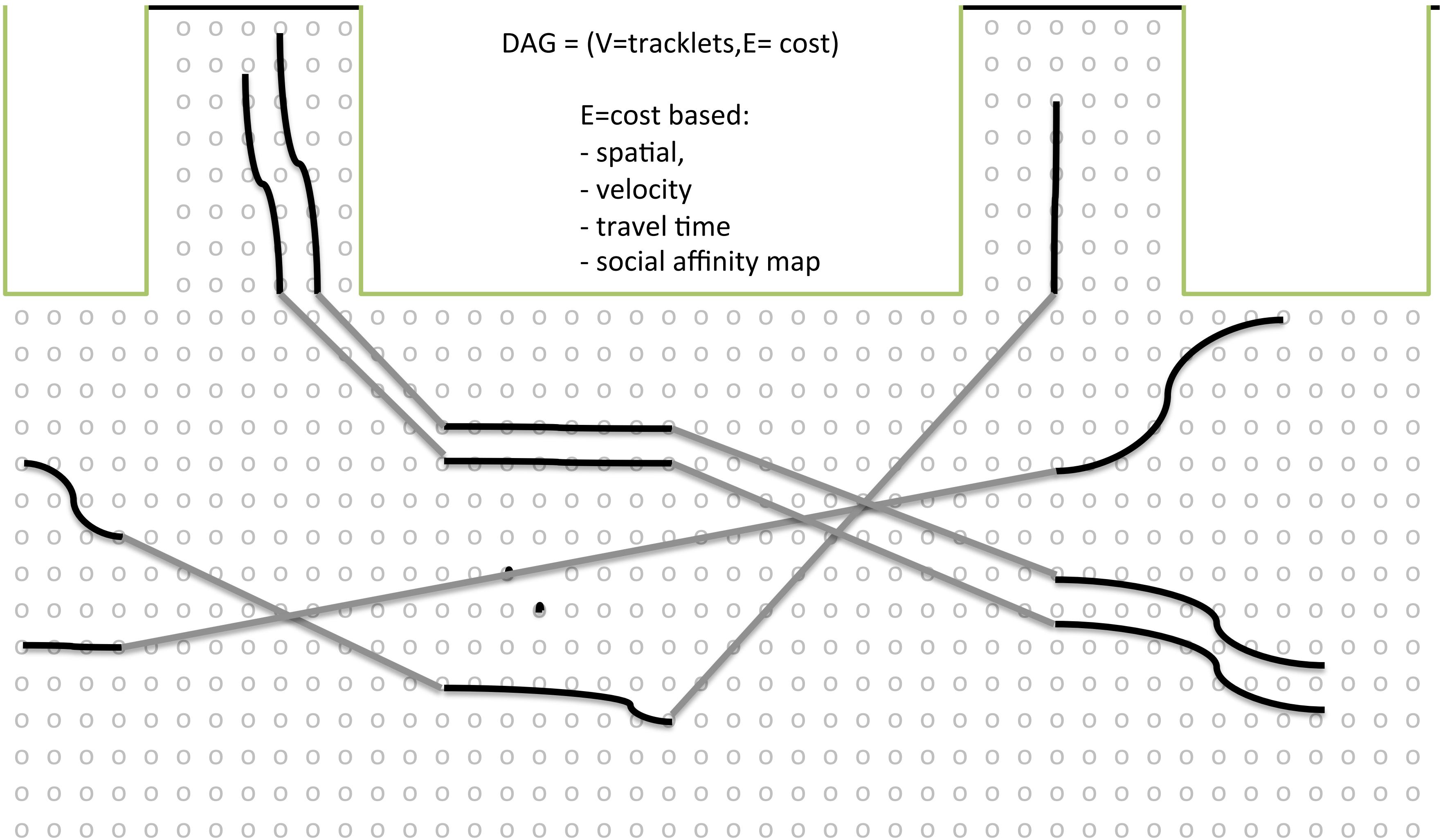
III. Tracklet association (Top view)



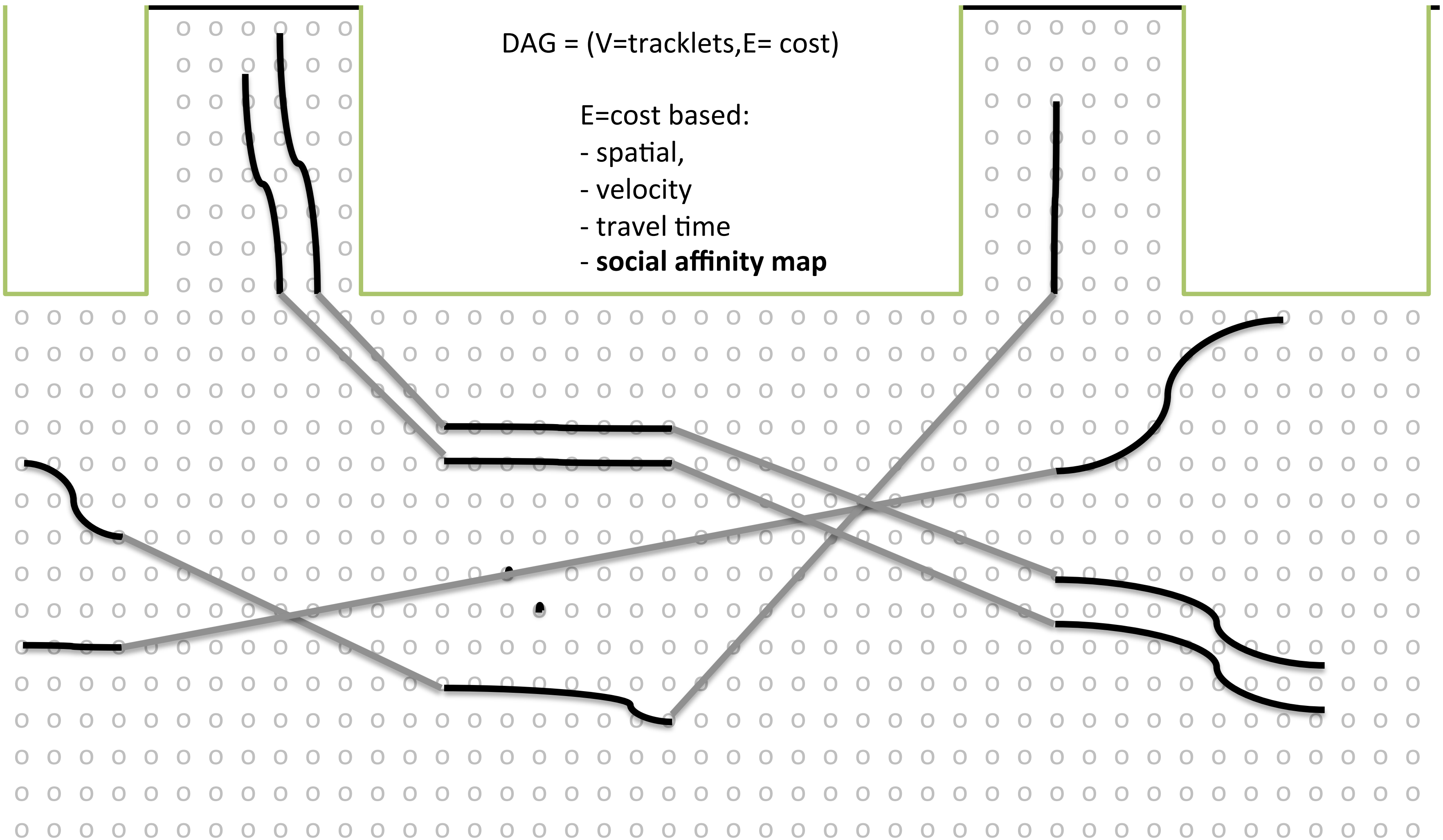
III. Tracklet association (Top view)



III. Tracklet association (Top view)



III. Tracklet association (Top view)

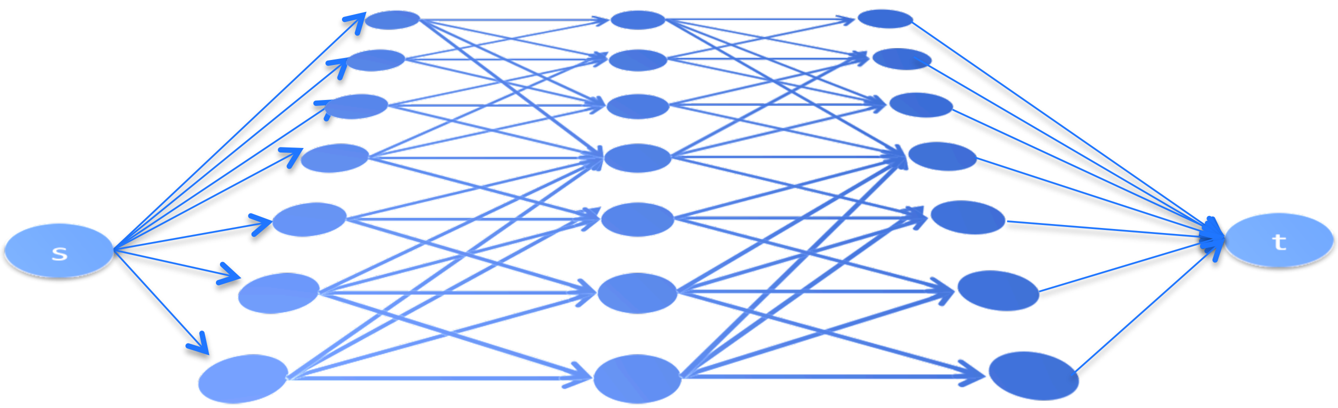


Network flow optimization

Objective: minimum cost maximum flow

$$\arg\min_{f} c(f)$$

$$c(f) = \sum \alpha_i f_i + \sum \beta_{ij} f_{ij}$$



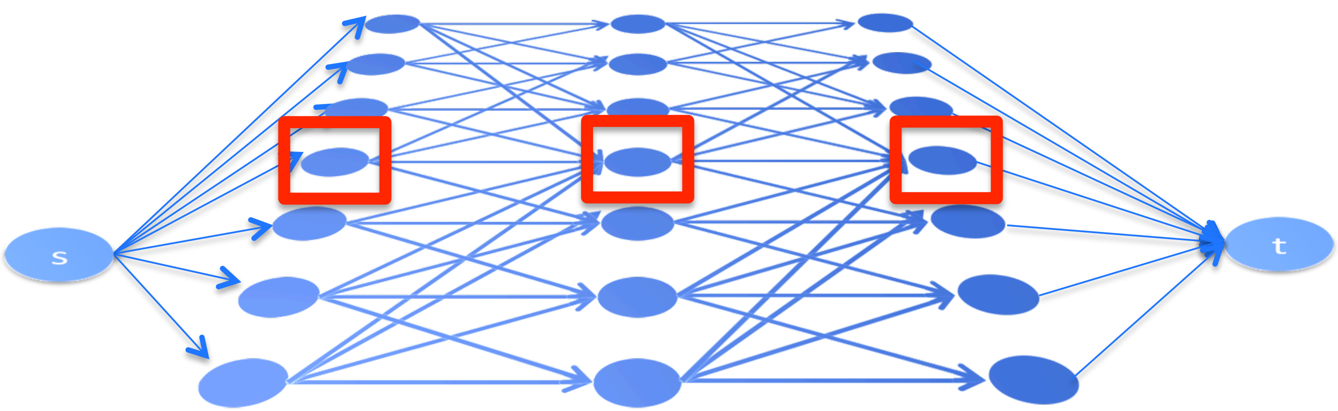
Where α_i, β_{ij}, γ_{OD} are the costs, and f_i the flows

Network flow optimization

Objective: minimum cost maximum flow

$$\arg\min_{f} c(f)$$

$$c(f) = \boxed{\sum \alpha_i f_i} + \sum \beta_{ij} f_{ij}$$



Cost α_i based:

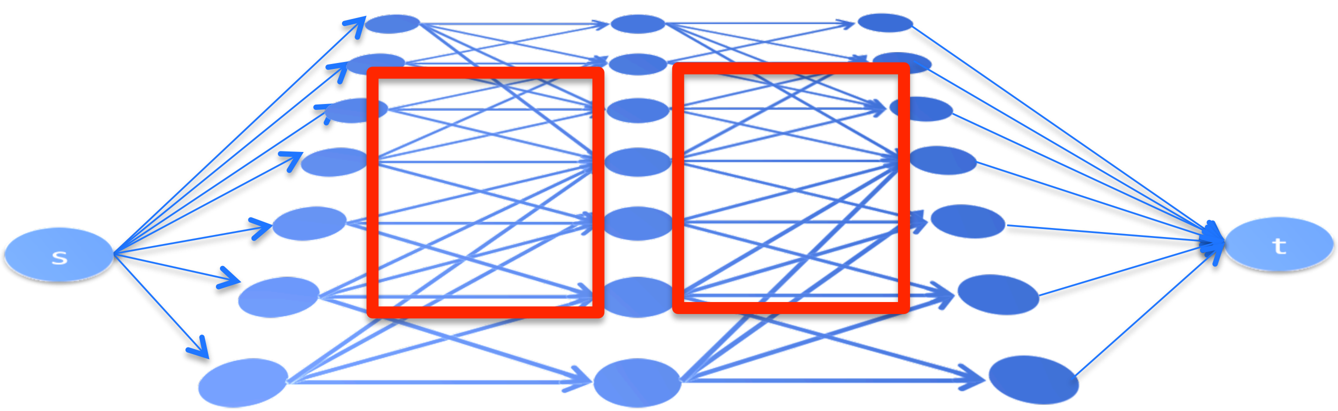
- Detection likelihood

Network flow optimization

Objective: minimum cost maximum flow

$$\arg\min_{f} c(f)$$

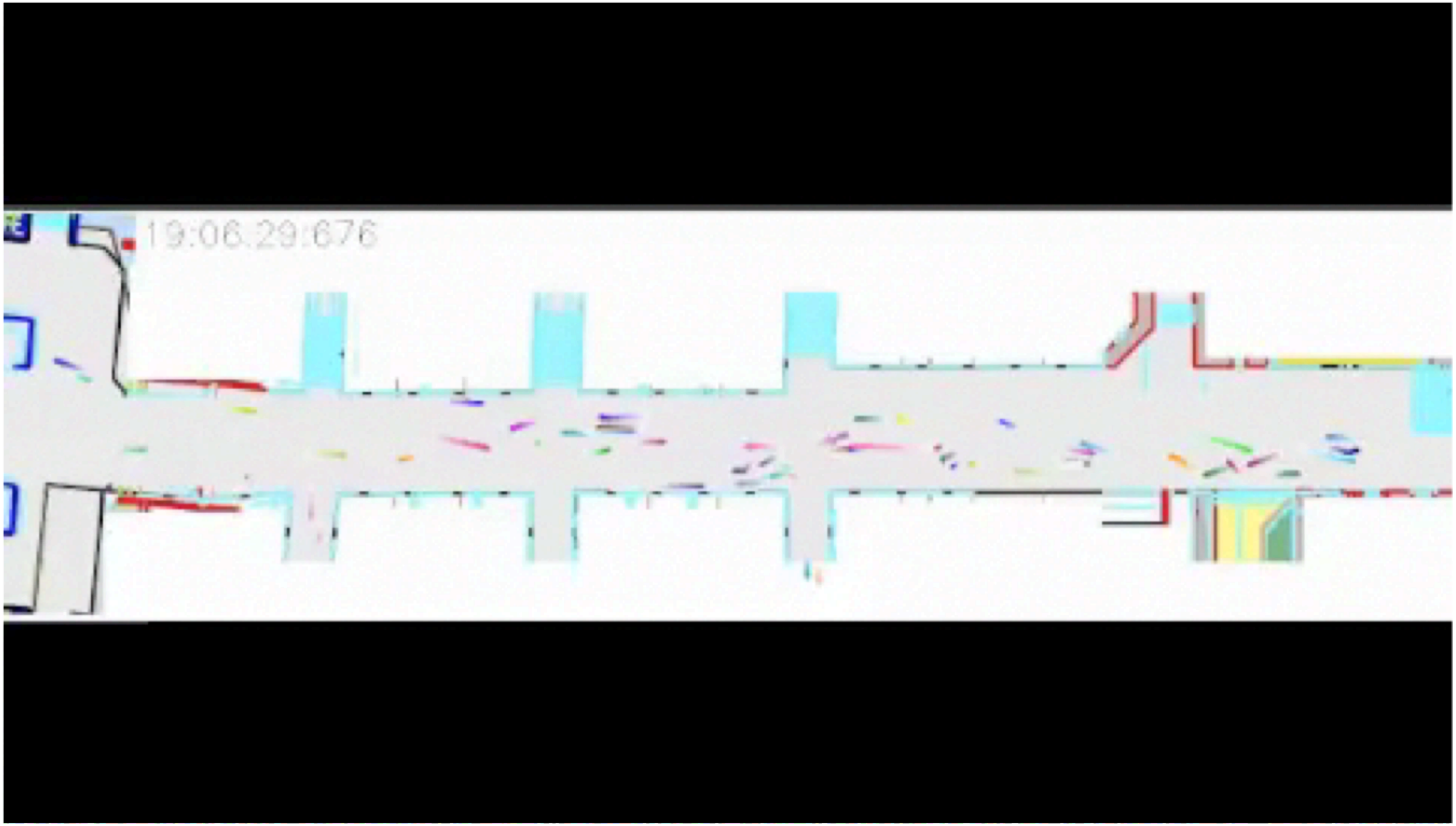
$$c(f) = \sum \alpha_i f_i + \boxed{\sum \beta_{ij} f_{ij}}$$



Cost β_{ij} based:

- spatial
- velocity
- **S**ocial **A**ffinity **M**ap

Tracklet association With Social Affinity Map



Conclusion

A new dimension to "Google Analytics":

Analyzing people outside of website