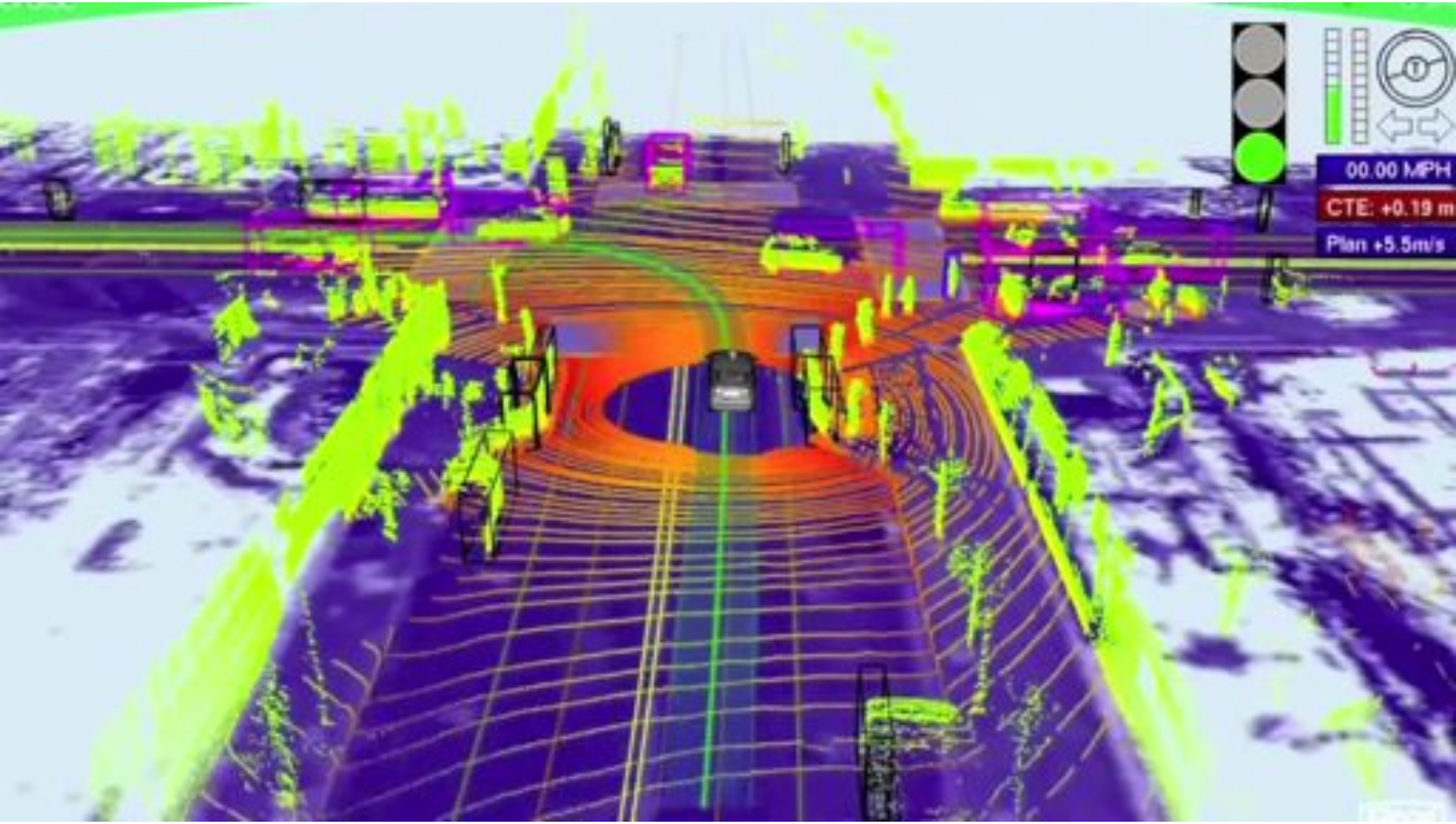
Guest lecture:
Visual Tracking

Alexandre Alahi

Stanford Vision Lab / CVGL

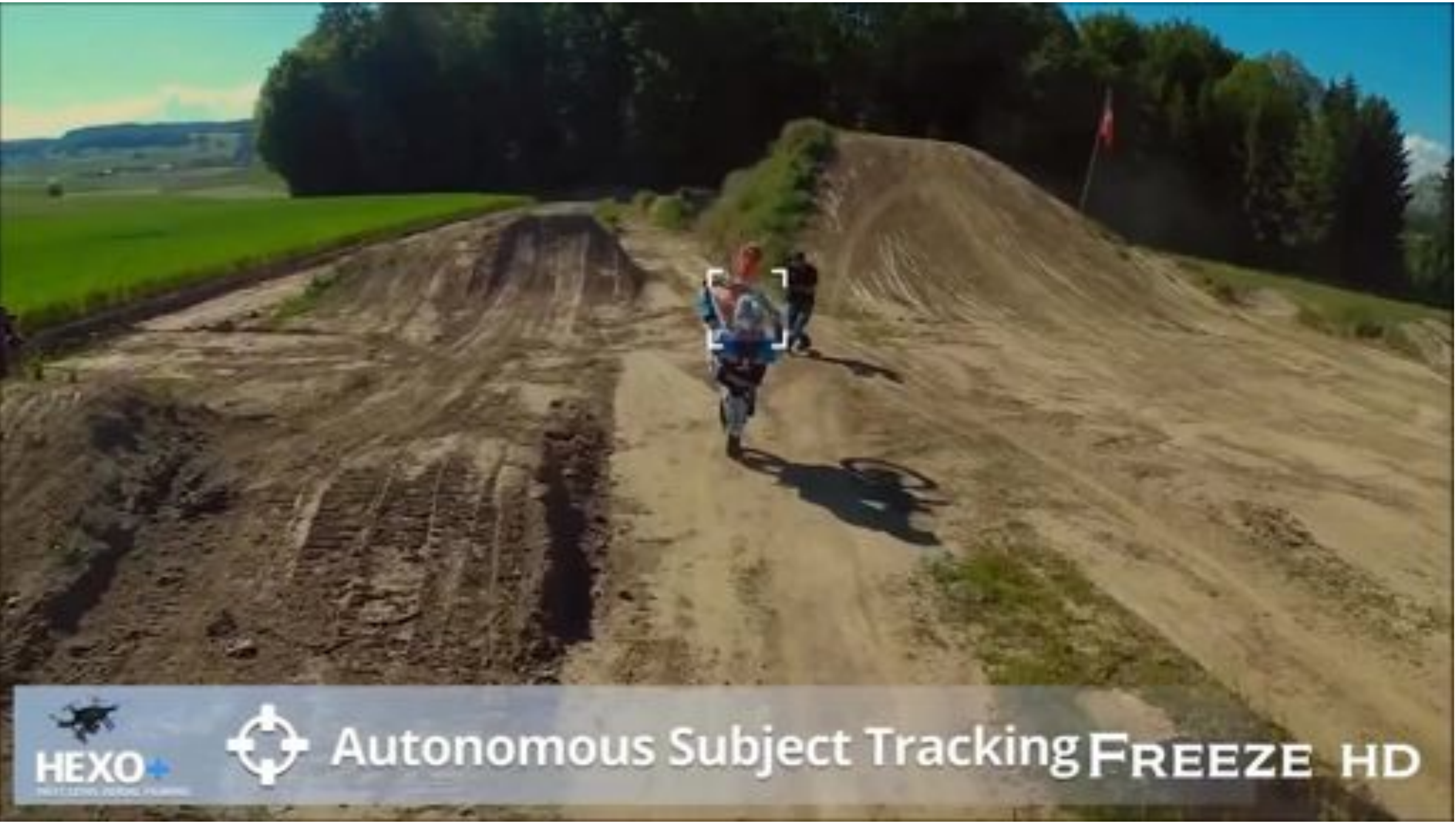

Tracklets:
complete
only sink


00.00 MPH
CTE: +0.19 m
Plan +5.5m/s

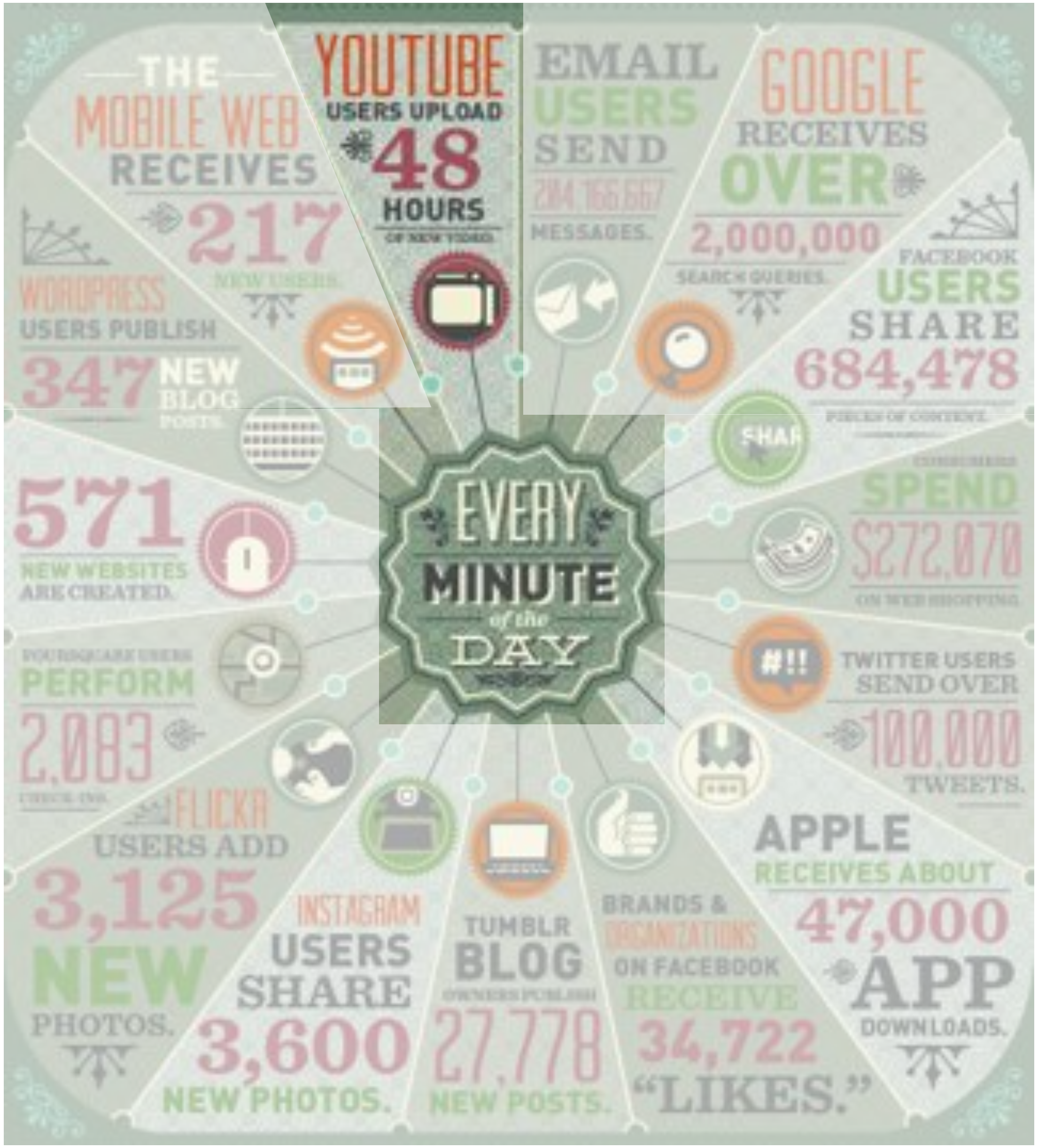

STANFORD


Visit store.ovi.com
ovi
Visit store.ovi.com
5125


#nextlevelaerialfilming
Autonomous
Subject Tracking


HEXO+
Autonomous Subject Tracking
FREEZE HD

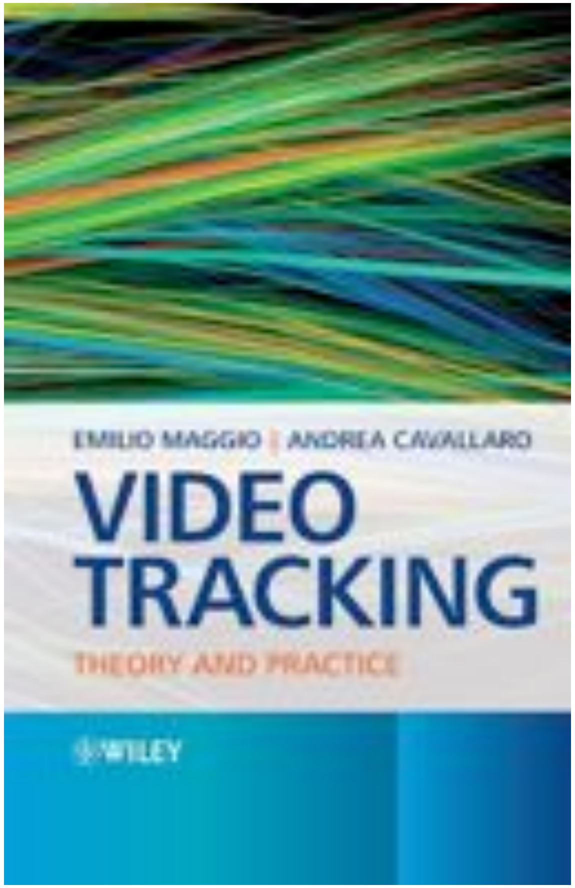

Pinter Wollman et al 2011 J Roy Soc Interface



EVERY MINUTE of the DAY
THE MOBILE WEB RECEIVES 217 NEW USERS
YOUTUBE USERS UPLOAD 48 HOURS
EMAIL USERS SEND 204,166,667 MESSAGES.
GOOGLE RECEIVES OVER 2,000,000 SEARCH QUERIES.
WORDPRESS USERS PUBLISH 347 NEW BLOG POSTS.
FACEBOOK USERS SHARE 684,478
571 NEW WEBSITES ARE CREATED.
SPEND $272,070
PERFORM 2,083
TWITTER USERS SEND OVER 100,000 TWEETS.
FLICKR USERS ADD 3,125 NEW PHOTOS.
INSTAGRAM USERS SHARE 3,600 NEW PHOTOS.
TUMBLR BLOG 27,778 NEW POSTS.
BRANDS & ORGANIZATIONS ON FACEBOOK RECEIVE 34,722 "LIKES."
APPLE RECEIVES ABOUT 47,000 APP DOWNLOADS.

All started in the early 60s

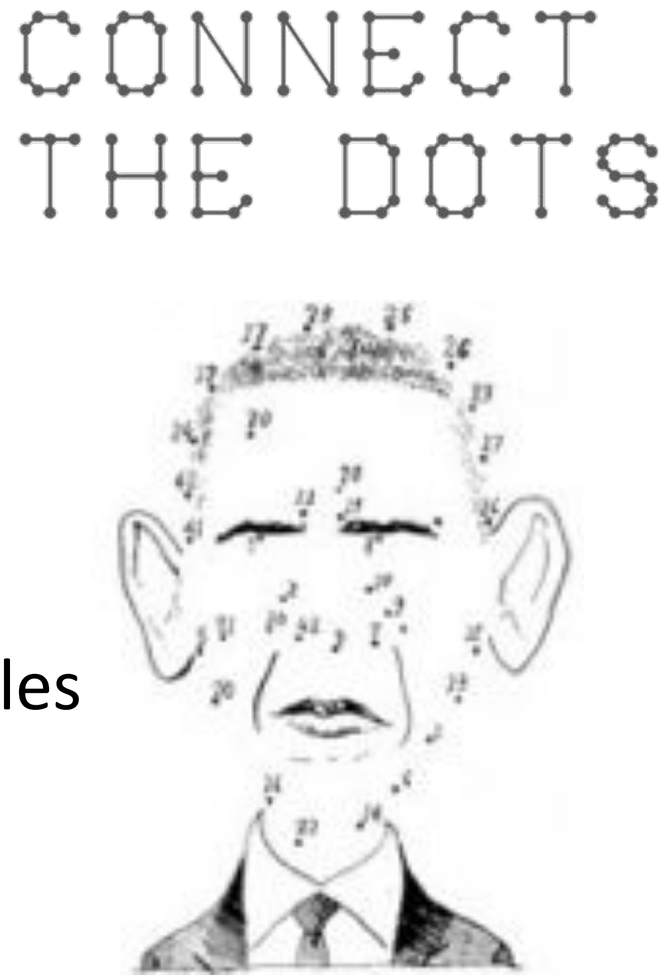
- With Kalman filter for military
- A book on Video Tracking: Theory and Practice



What is tracking about?

- Data association
- Similarity measurement
- Correlation
- Matching/Retrieval

- Reasoning with “strong” priors
- Detection with very similar examples



Outline

1. Problem statement
2. Challenges
3. Object representation
4. Single target tracking
5. Multi-target tracking
6. Tips & references

Problem statement

- Input: target
- Objective: Estimate target <u>state</u> over <u>time</u> (space)
- State:
  - Position
  - Appearance
  - Shape
  - Velocity
  - Affine transformation w.r.t. previous patch

Problem statement

- Input: target
- Objective: Estimate target state over time
- State: e.g. position

- Design/pipeline elements: (O.S.S.)
  - **O**bject representation
  - **S**imilarity measure
  - **S**earching process

Outline

1. Problem statement
2. **Challenges**
3. Object representation
4. Single target tracking
5. Multi-target tracking
6. Tips & references

What are the challenges?

- Variations due to geometric changes
  (pose, articulation, scale)
- Variations due to photometric factors
  (illumination, appearance)
- Occlusions
- Non-linear motion
- Very limited resolution, blurry
  (standard recognition might fail)
- Similar objects in the scene

See live demo

Algorithms common issues

- Track initiation & termination
- Occlusion handling
- Merging/switching
- Drifting due to wrong update of the target model

See live demo

Outline

1. Problem statement
2. Challenges
3. **Object representation**
    1. Low/mid/high level features
    2. Grid/Pyramid/Cascade
    3. Patch/keypoints
4. Single target tracking
5. Multi-target tracking
6. Tips & references

Object representation

- Goal:

we want a representation that is:

  – Descriptive enough to disambiguate target VS background
  – Flexible enough to cope with:
    - Scale
    - Pose
    - Illumination
    - Partial occlusions

Object representation

- Object approximation:
  - Segmentation / Polygonal approximation
  - Bounding ellipse/box
  - Position only

- Goal: Measure affinity

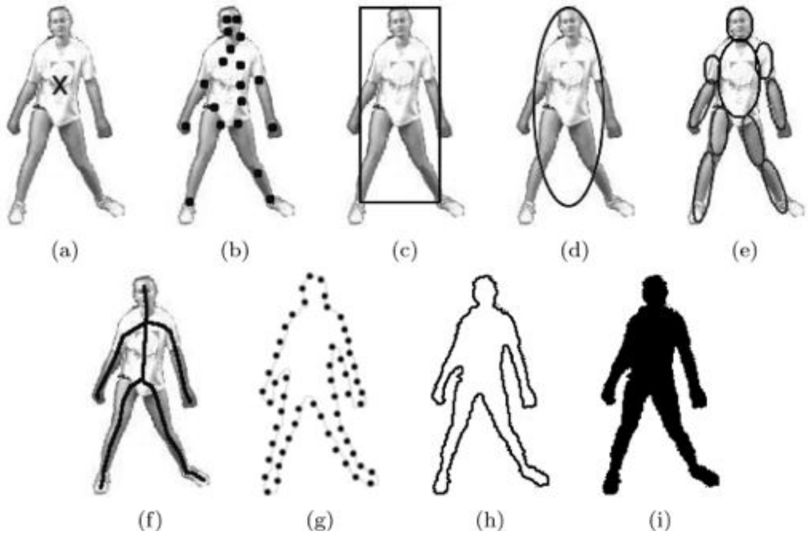


Image from A. Yilmaz et. Al : Object tracking: A survey. ACM Computing Surveys, 2006

Measuring Affinity

From Lecture 2

- In general: $$aff(x, y) = \exp\left(-\frac{1}{2\sigma_d^2}\|f(x) - f(y)\|^2\right)$$

- Examples:
  - Distance: $f(x) = location(x)$
  - Intensity: $f(x) = intensity(x)$
  - Color: $f(x) = color(x)$
  - Texture: $f(x) = filterbank(x)$

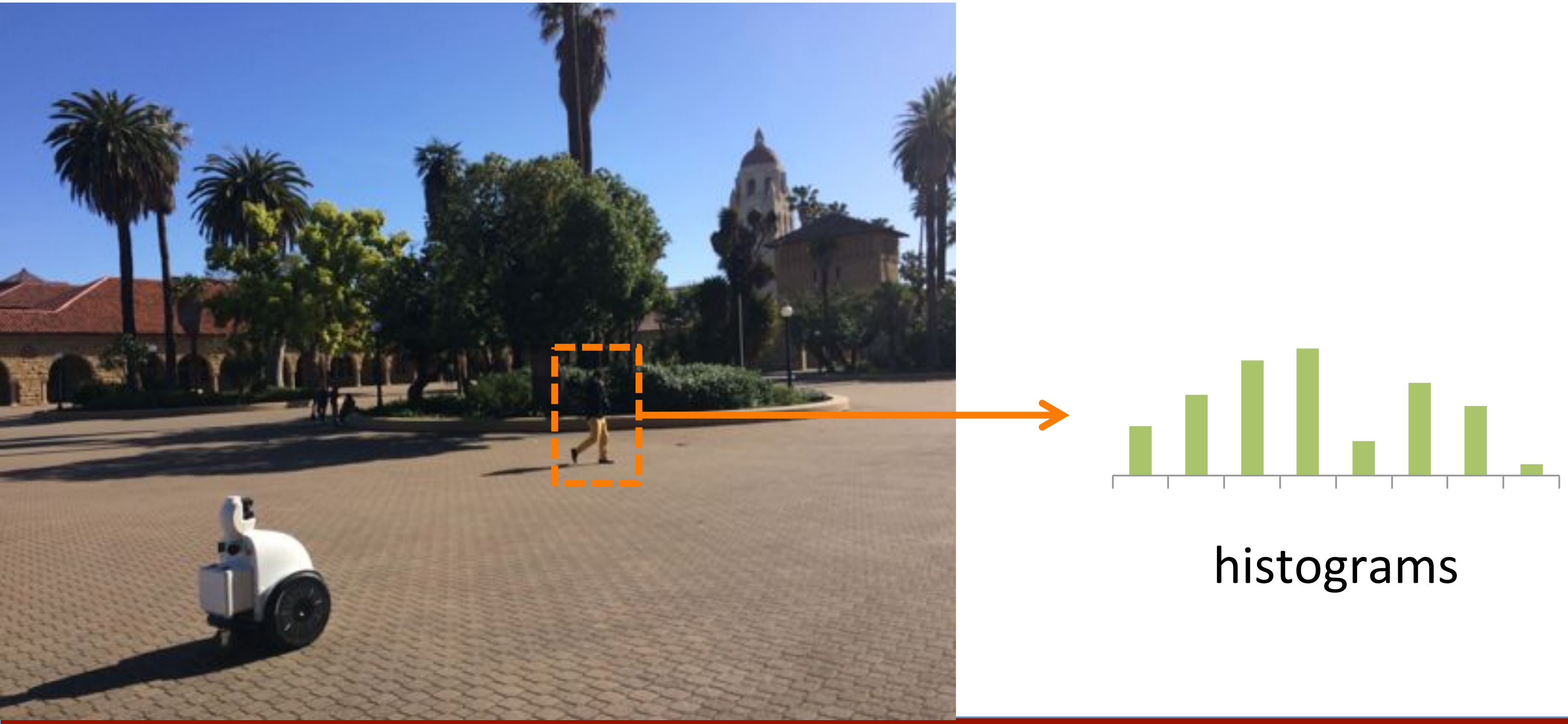
Pixels => Regions

- Note: Can also modify distance metric

slide credit: Forsyth & Ponce

Object representation: From light to useful information

- Low/mid/high level features



From cs231

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 $\theta = \tan^{-1}\left(\frac{\partial f}{\partial y} / \frac{\partial f}{\partial x}\right)$

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

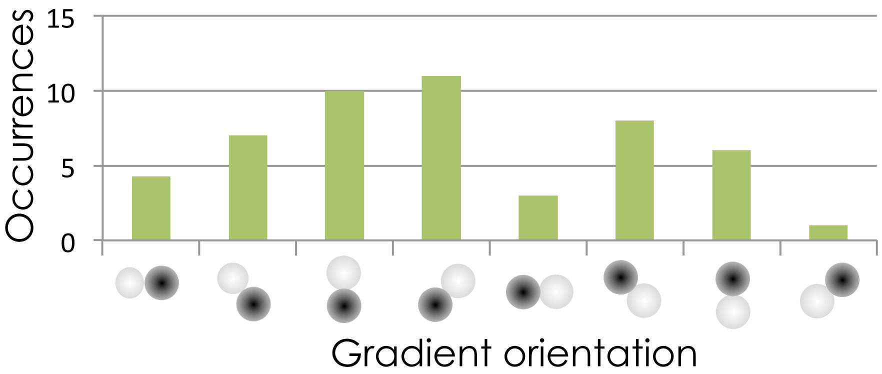
The *edge strength* is given by the gradient magnitude

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

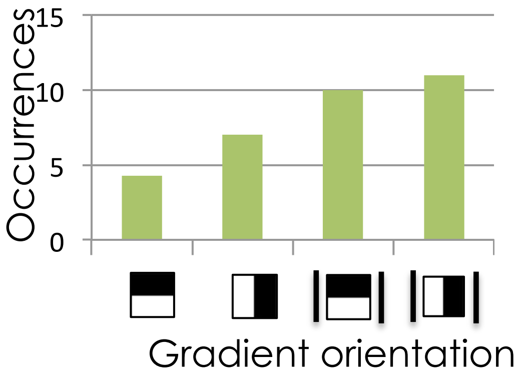
Source: Steve Seitz

Low-level features

- Integer responses



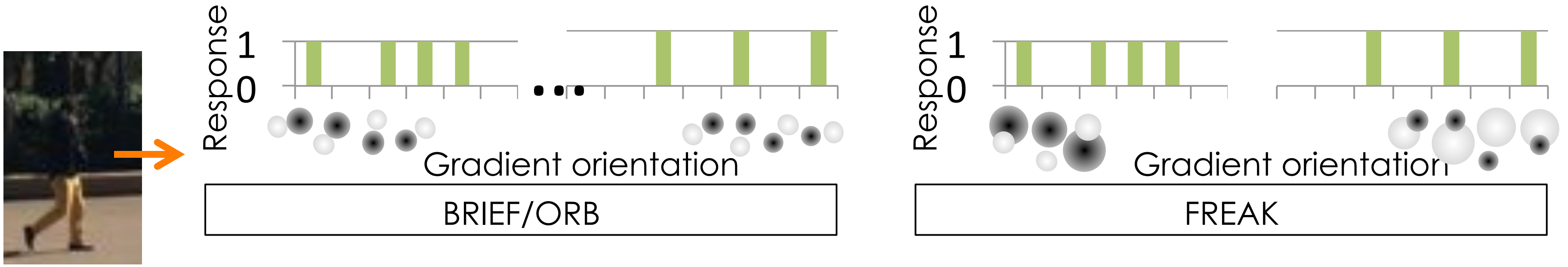
HoG feature used in SIFT-like descriptor



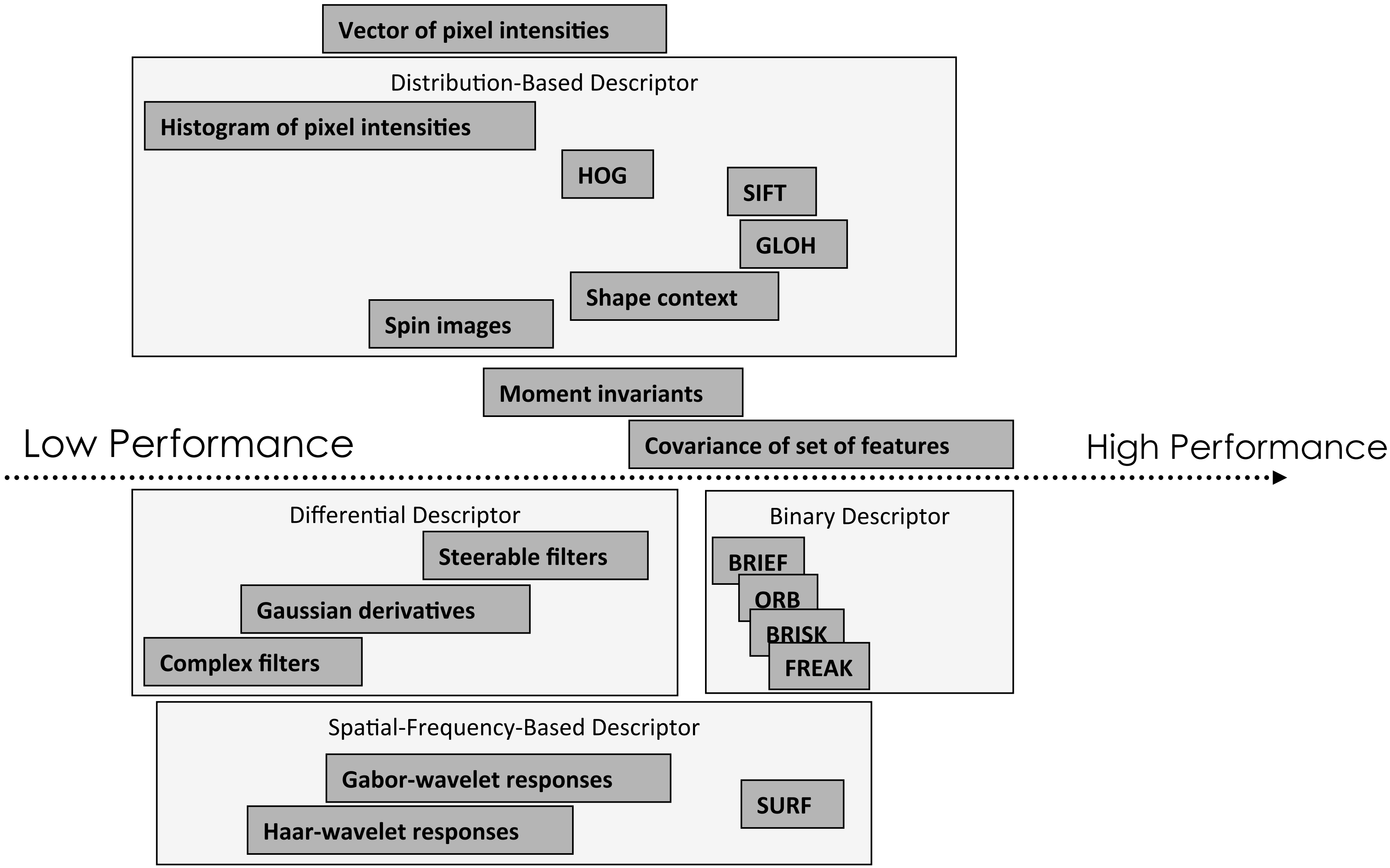
Haar feature used in SURF

Low-level features

- Binary responses

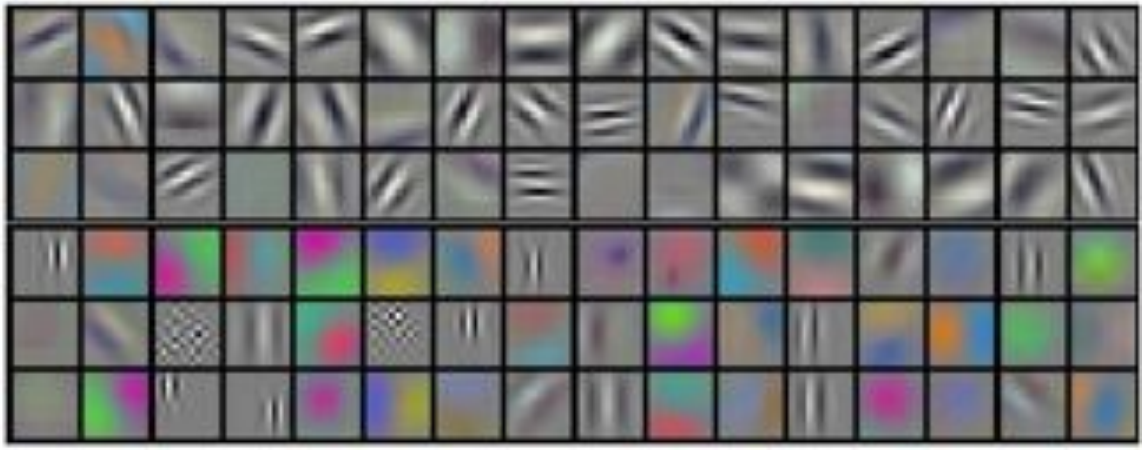


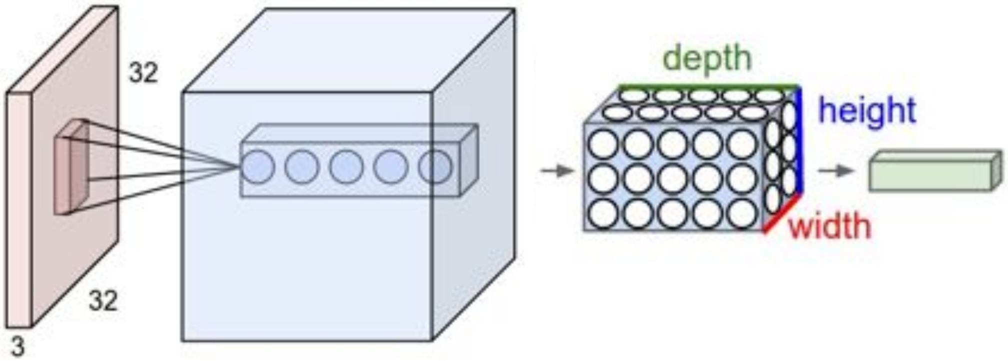
A bulk of Low-level features



Mikolajczyk et. al. "A performance evaluation of local descriptors." PAMI 2005

Recent trend: CNN features





Object representation: Sampling strategies

- Grid/pyramid/cascade of coarse-to-fine

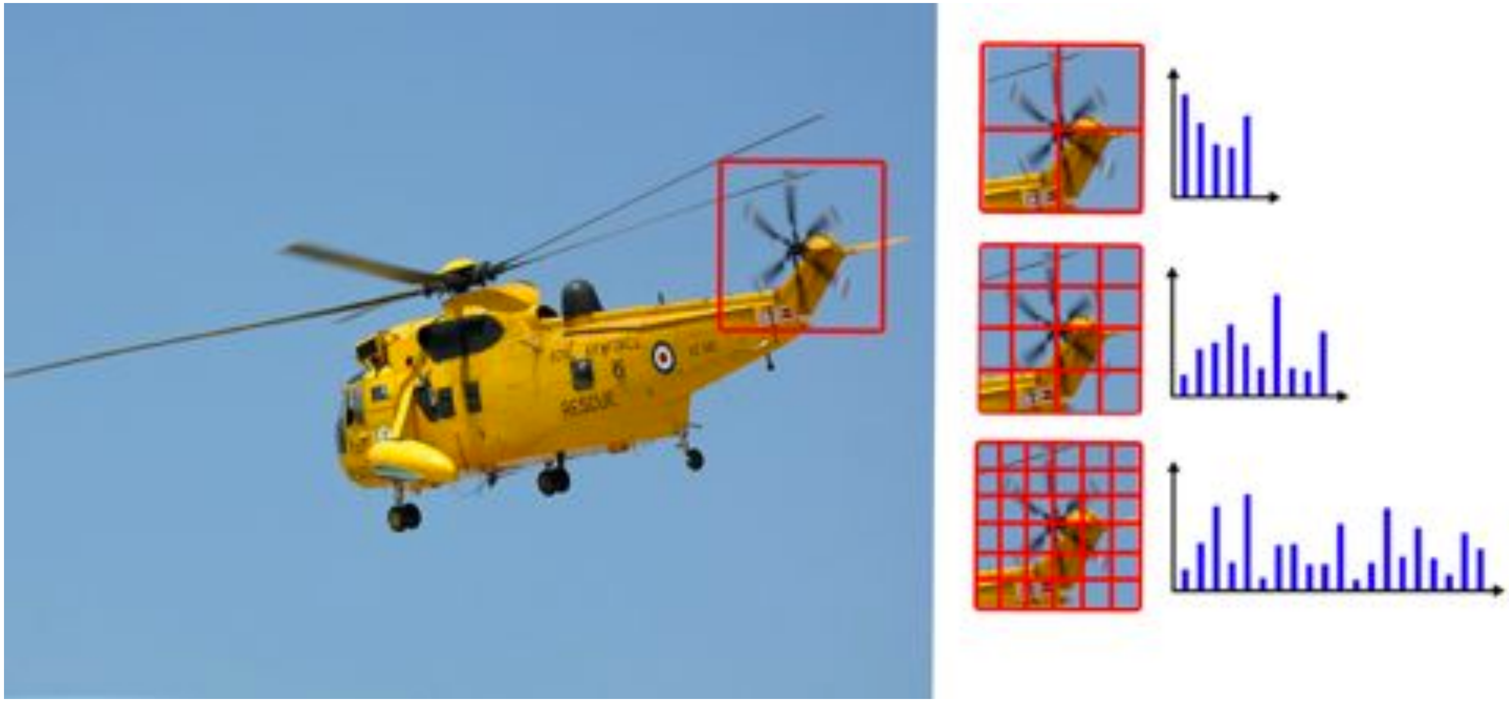
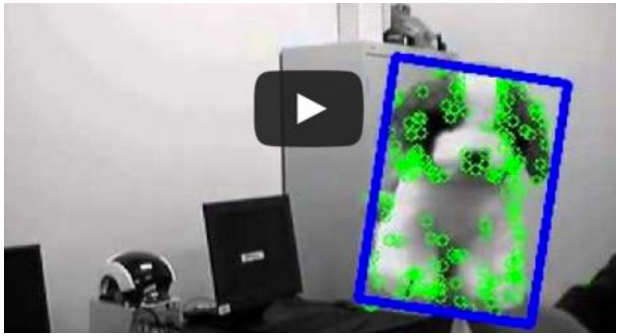


Image from L. Seidenari

Object representation: Sampling strategy

- Local patches/ Keypoints [1]



[1] A. Alahi et. al., Biologically-inspired keypoint, to be published by Wiley

Outline

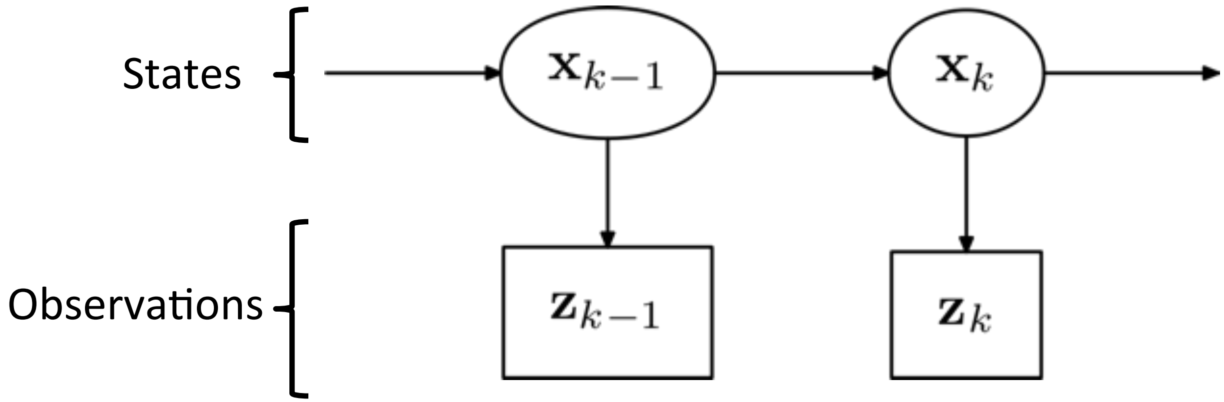
1. Problem statement
2. Challenges
3. Object representation
4. **Single target tracking**
    1. **Bayesian estimation**
    2. **On-line learning**
5. Multi-target tracking
6. Tips & references

Single target tracking

- Formulation
  - Input: bounding box at starting frame
  - Output: next bounding boxes across the next frames

Single target tracking
- Probabilistic tracking-

- Tracking as a Bayesian network
- Hidden Markov Model



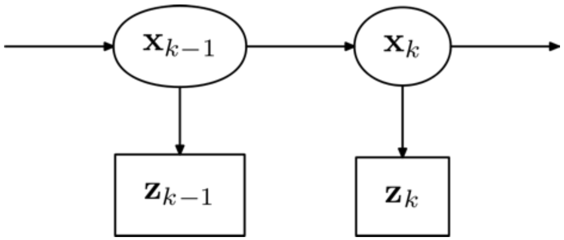
- Markov assumptions

$$p(x_k \mid x_{1:k-1}) = p(x_k \mid x_{k-1})$$

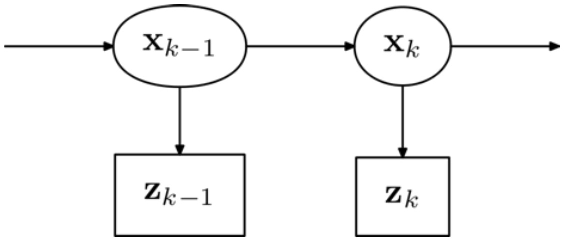
$$p(z_k \mid x_{1:k}) = p(z_k \mid x_k)$$

Single target tracking
- Probabilistic tracking-

- Recursive Bayes filters
- Find posterior $p(x_k \mid z_{1:k})$
- State eq. (motion dynamics) $f(x_k \mid x_{k-1})$
- Observation eq. (image) $g(z_k \mid x_k)$



Single target tracking
- Probabilistic tracking-



- Recursive Bayes filters
- Find posterior $p(x_k \mid z_{1:k})$
- State eq. (motion dynamics) $f(x_k \mid x_{k-1})$
- Observation eq. (image) $g(z_k \mid x_k)$

- Prediction

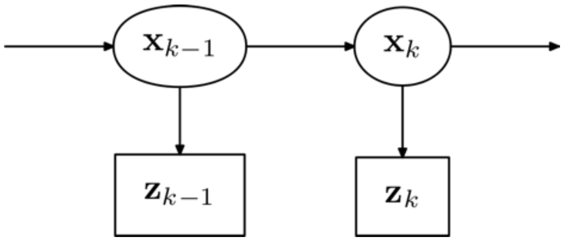
Previous posterior

$$p(x_k \mid z_{1:k-1}) = \int f(x_k \mid x_{k-1}) \boxed{p(x_{k-1} \mid z_{1:k-1})} dx_{k-1}$$

- Update

$$p(x_k \mid z_{1:k}) = \frac{g(z_k \mid x_k) p(x_k \mid z_{1:k-1})}{\int g(z_k \mid x_k) p(x_k \mid z_{1:k-1}) dx_k}$$

Single target tracking
- Probabilistic tracking-



- Solving Bayes Equations
  - Gaussian & Linear
    - Kalman filter [1]
  - Gaussian non-linear
    - Extended Kalman filter
  - Non-Gaussian non-linear
    - Monte Carlo methods (Condensation [2])
  - Hill-climbing on posterior
    - Mean-shift

[1] Kalman, Rudolph Emil. "A new approach to linear filtering and prediction problems." Journal of Fluids Engineering , 1960
[2]Isard, Michael, and Andrew Blake. "Condensation—conditional density propagation for visual tracking." IJCV 1998

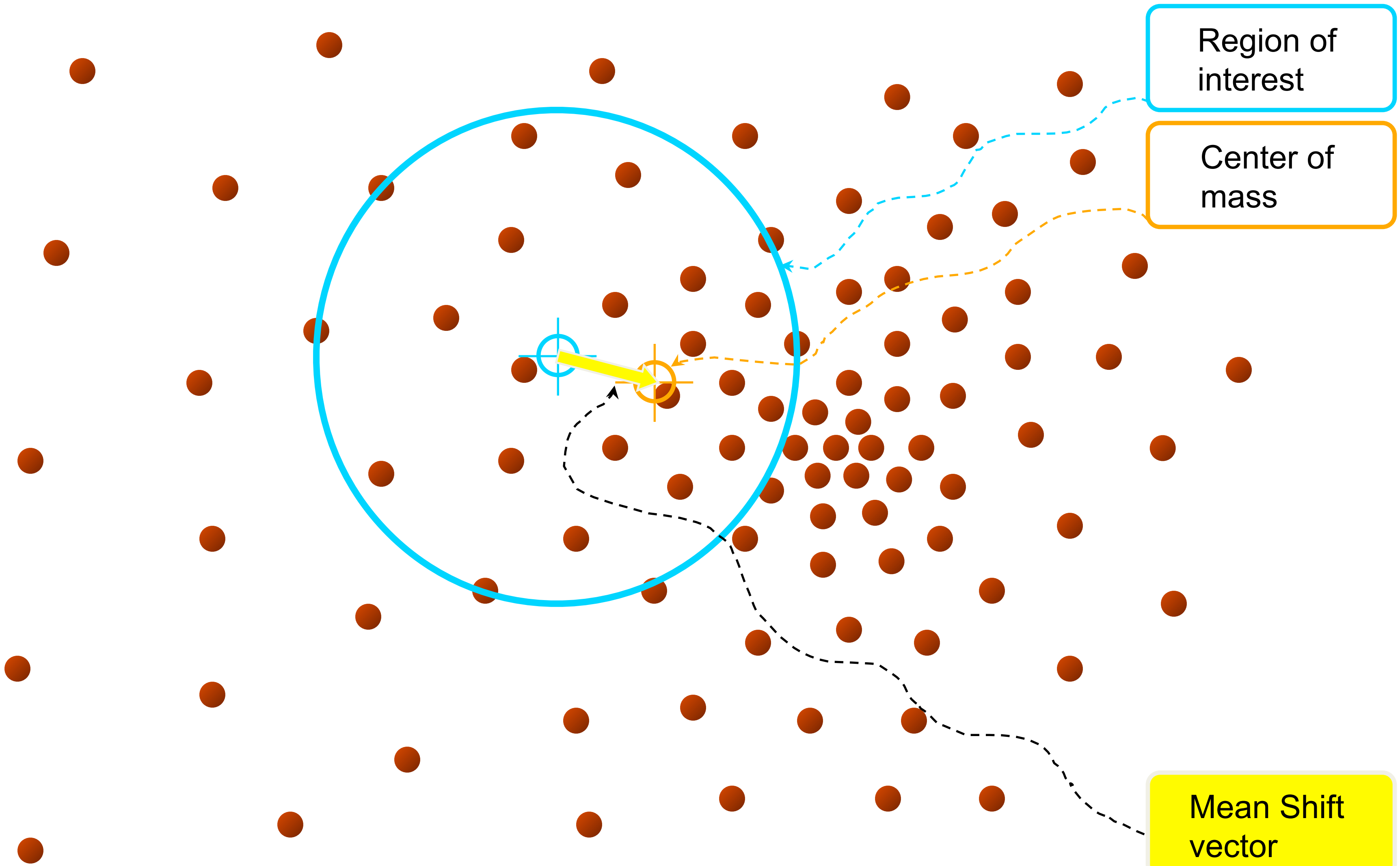
Single target tracking
- Probabilistic tracking-

- Kernel-based tracking [1]

- Mean-shift
  - Non-parametric feature space
  - Locate the maxima of a density function
  - Color histogram / Bhattacharyya

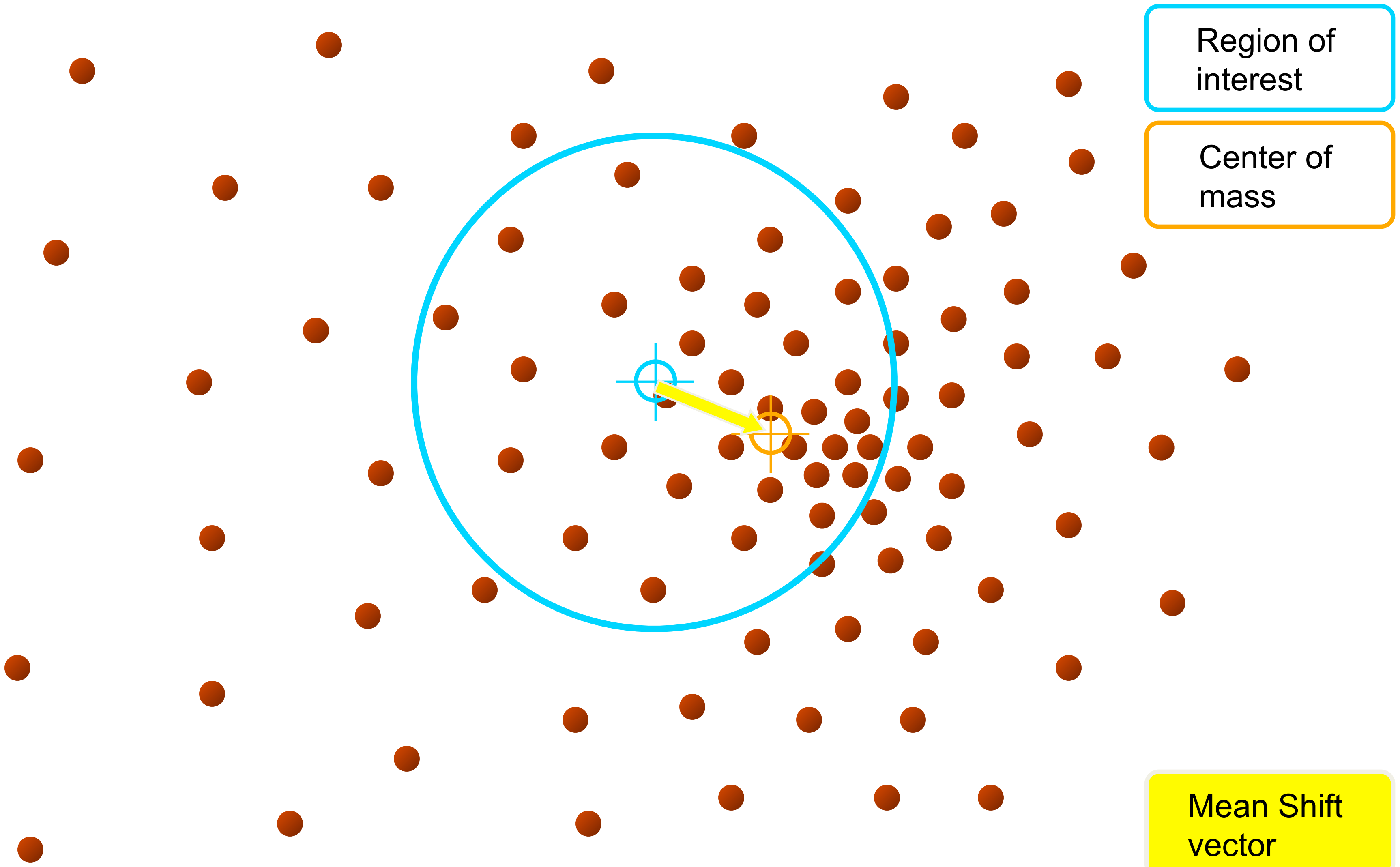
[1] Comaniciu, Dorin, Visvanathan Ramesh, and Peter Meer. "Kernel-based object tracking." PAMI (2003)

Mean-Shift



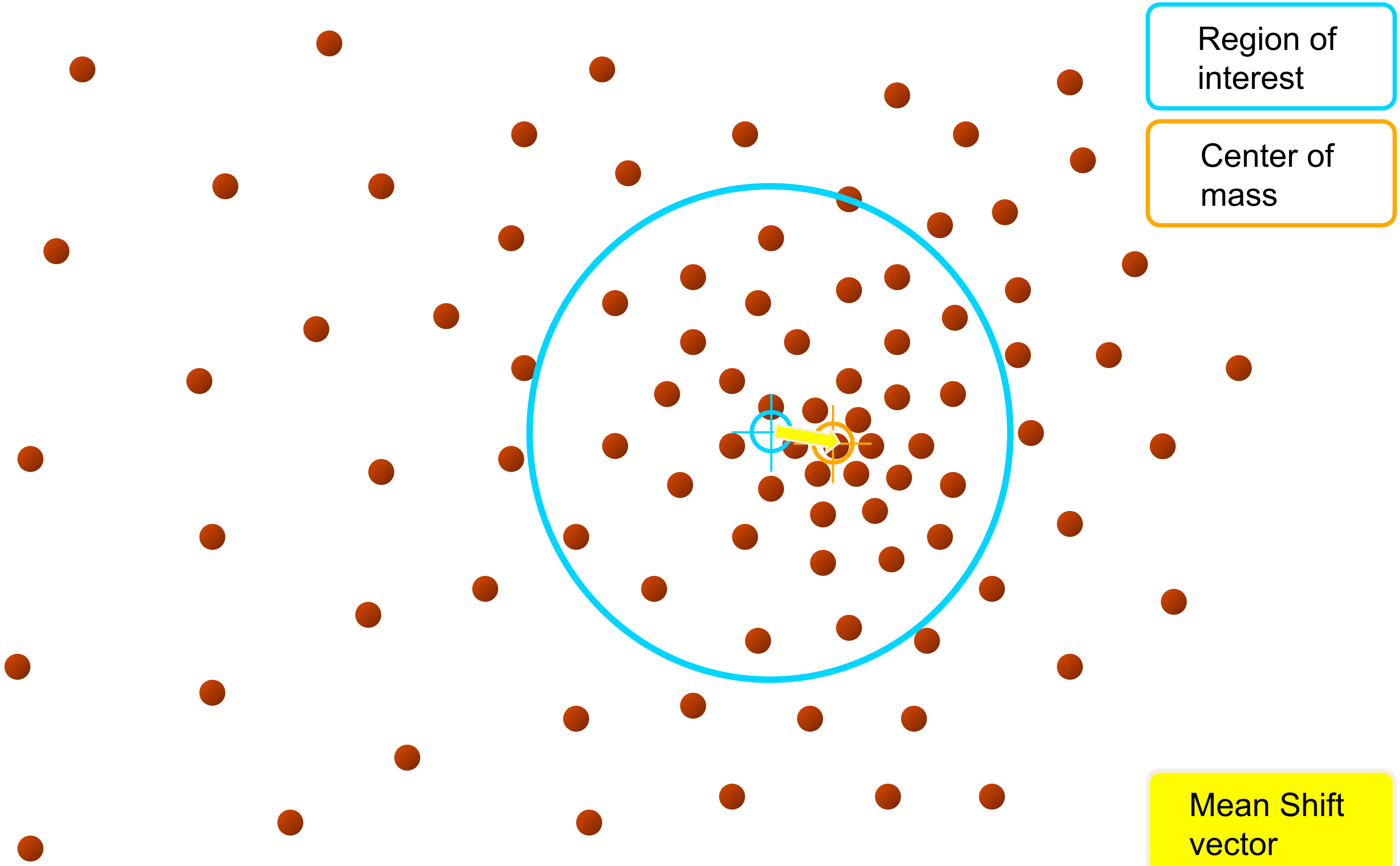
Slide by Y. Ukrainitz & B. Sarel

Mean-Shift



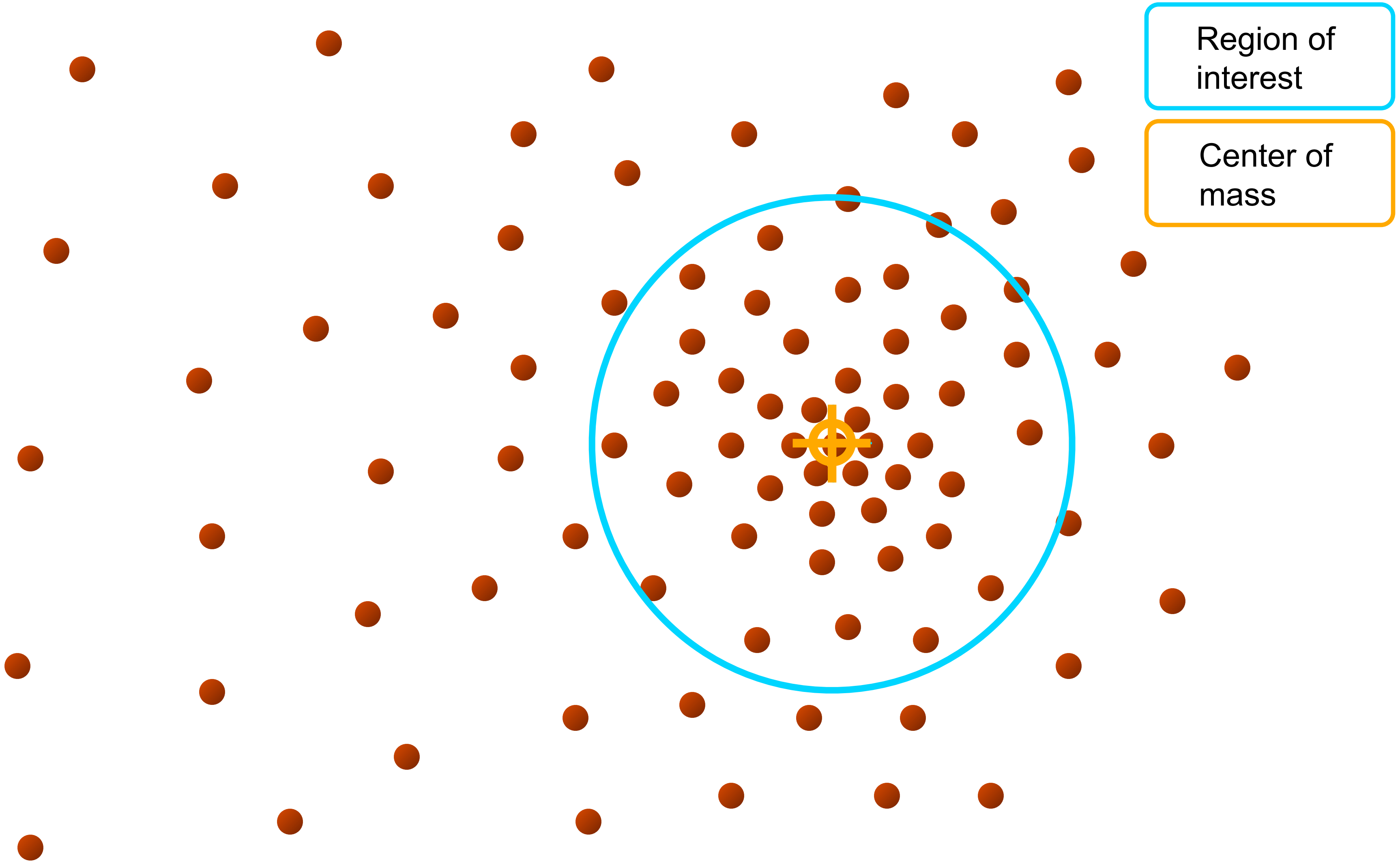
Slide by Y. Ukrainitz & B. Sarel

Mean-Shift



Slide by Y. Ukrainitz & B. Sarel

Mean-Shift



Slide by Y. Ukrainitz & B. Sarel

Single target tracking
- Probabilistic tracking-

- Mean-shift

Pros:

- Fast
- No need for texture
- Tolerate for minor change of appearance

Cons:

- Only one hypothesis, no fallback if tracker is lost
- A single histogram does not capture variation of appearance
- Limited discriminative power with background

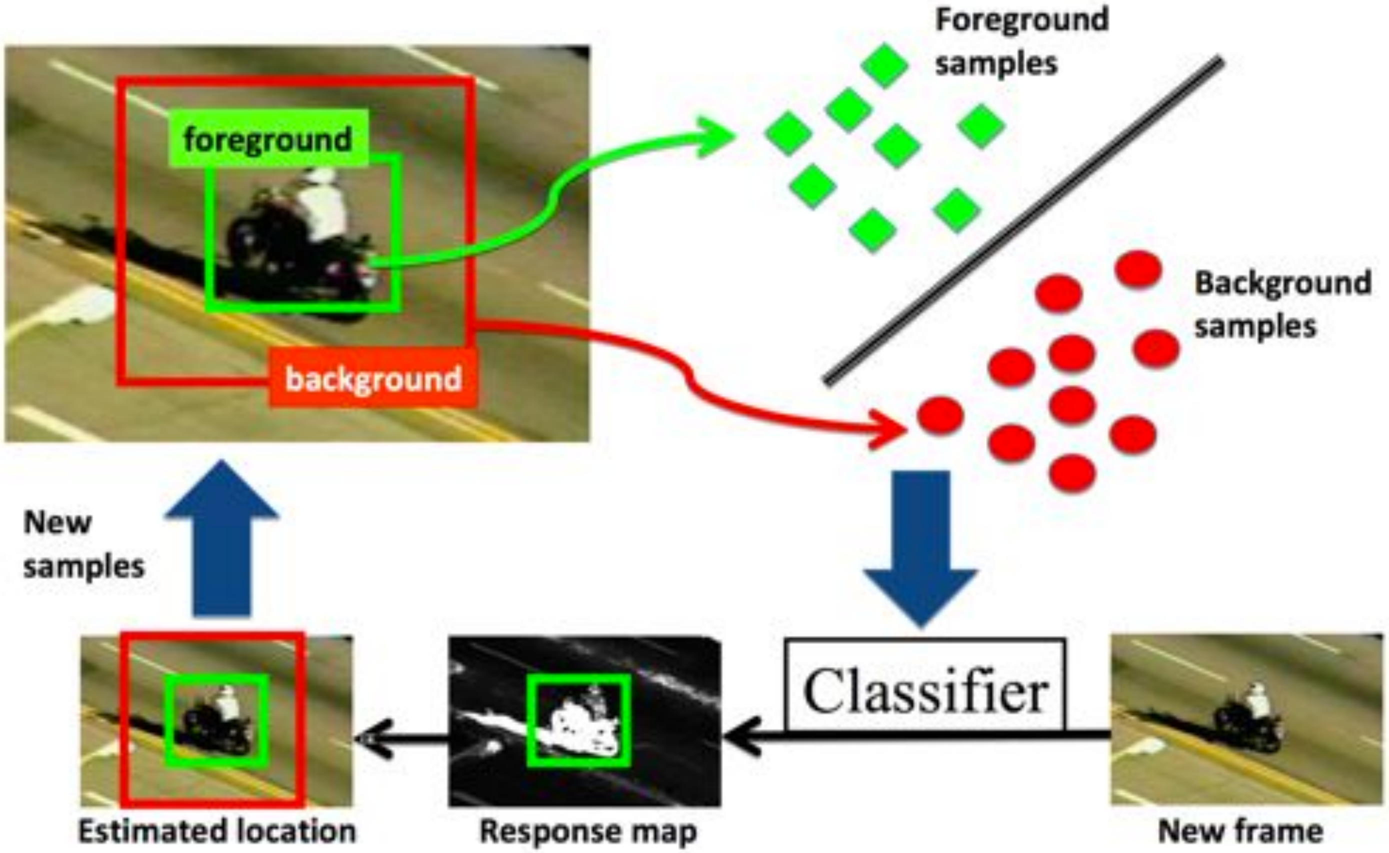
Single target tracking
- On-line learning -

- Discriminative modeling (tracking-by-detection)

- Learn and apply a detector or predictor

- Challenges:
  - What are training data? Labeled?
  - How to avoid drift? Handle occlusion?
  - How to control complexity?

Single target tracking
- On-line learning -



Slide from Collins, PSU

Single target tracking
- On-line learning -

- On-line discriminative learning
- One shot learning
- On-line update of the classifier

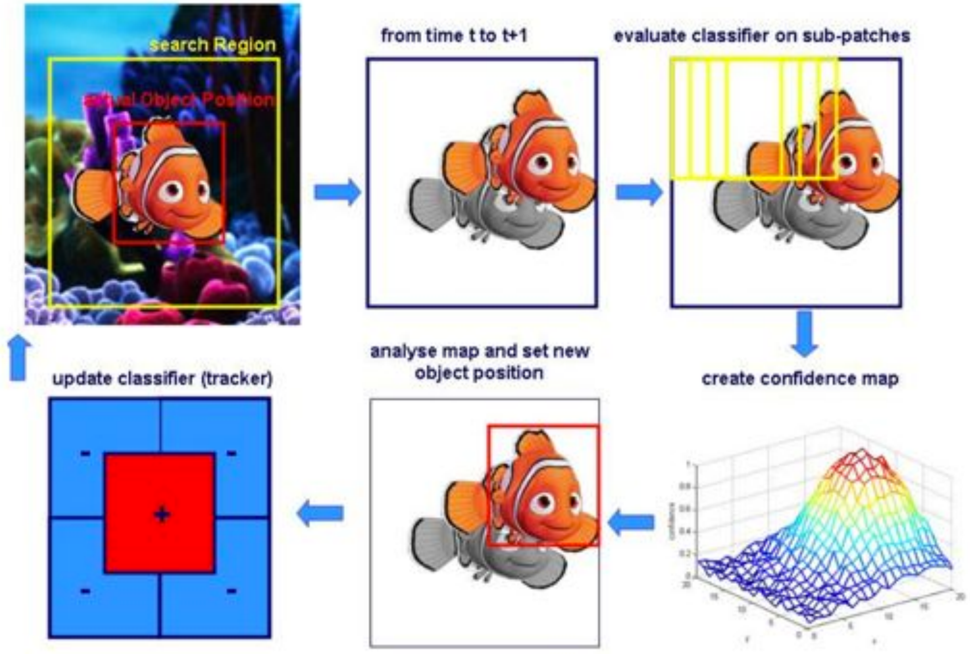


Figure from Grabner and Bischof CVPR 06

Single target tracking
- On-line learning -

- Examples of on-line discriminative learning
  - Multiple Instance Learning [1]
  - Kernelized Structured SVM [2]
  - Combine short track + detector [3]

[1] Babenko, Boris, Ming-Hsuan Yang, and Serge Belongie. "Visual tracking with online multiple instance learning." CVPR 2009
[2] Hare, Sam, Amir Saffari, and Philip HS Torr. "Struck: Structured output tracking with kernels." ICCV 2011
[3] Kalal, Zdenek, Krystian Mikolajczyk, and Jiri Matas. "Tracking-learning-detection." PAMI 2012

Single target tracking
- On-line learning -

- On-line discriminative learning

Pros:

- Can handle several appearance changes
- Can detect after full occlusion

Cons:

- Can drift
- Learning is not trivial

Outline

1. Problem statement
2. Challenges
3. Object representation
4. Single target tracking
5. **Multi-target tracking**
    1. **Formulation**
    2. **Graph-based**
6. Tips & references

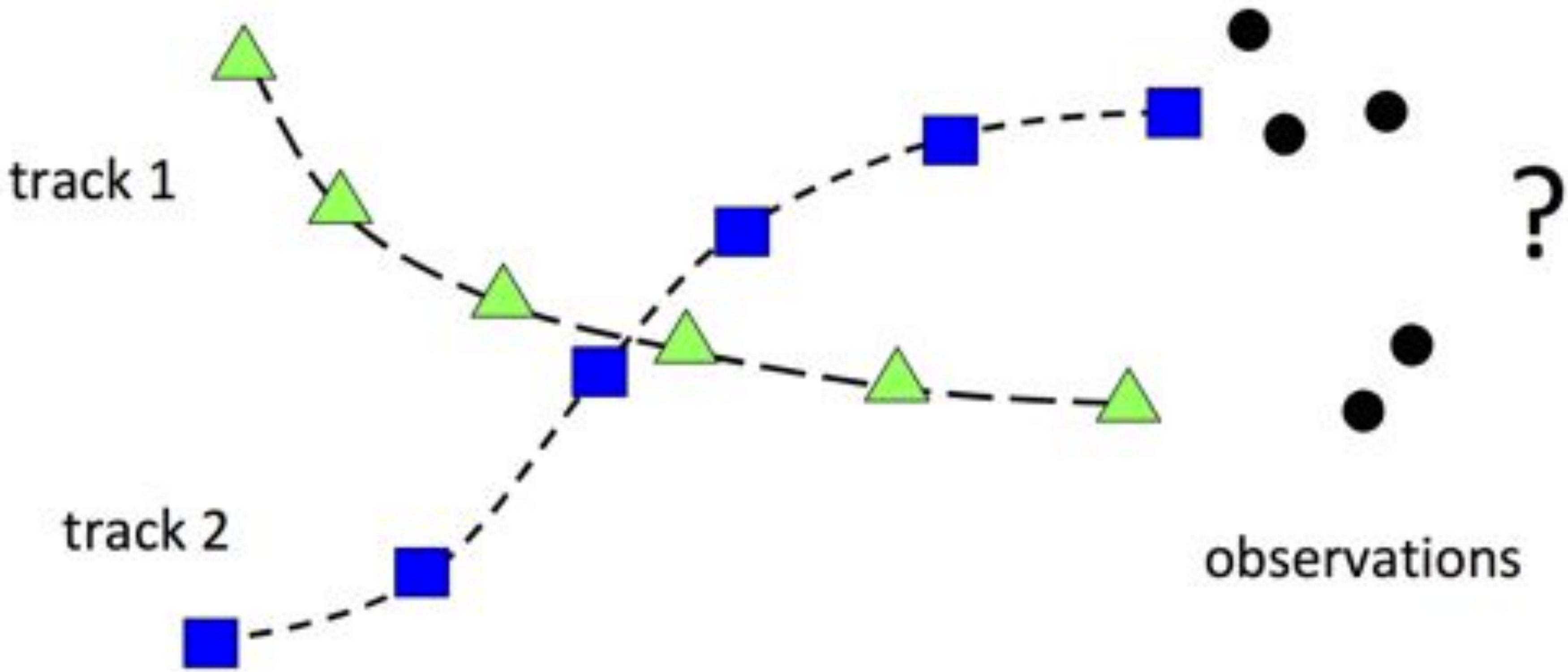
Multi-target tracking

- Formulation
  - Input: a set of detections (from next module R-CNN)
  - Output: state (id) for each detections

What is Multi-target tracking about?

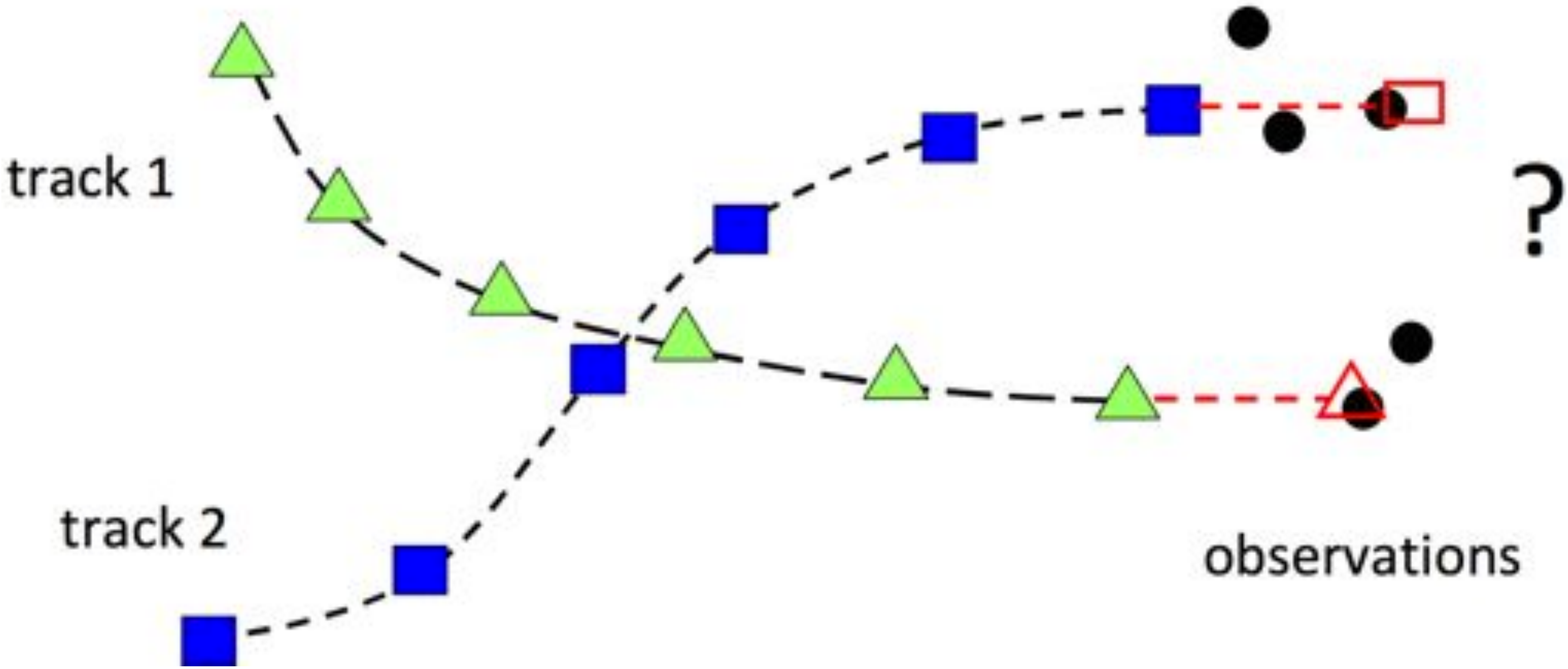
- Data association
- Assignment problems
- Discrete combinatorial optimization

Multi-target tracking



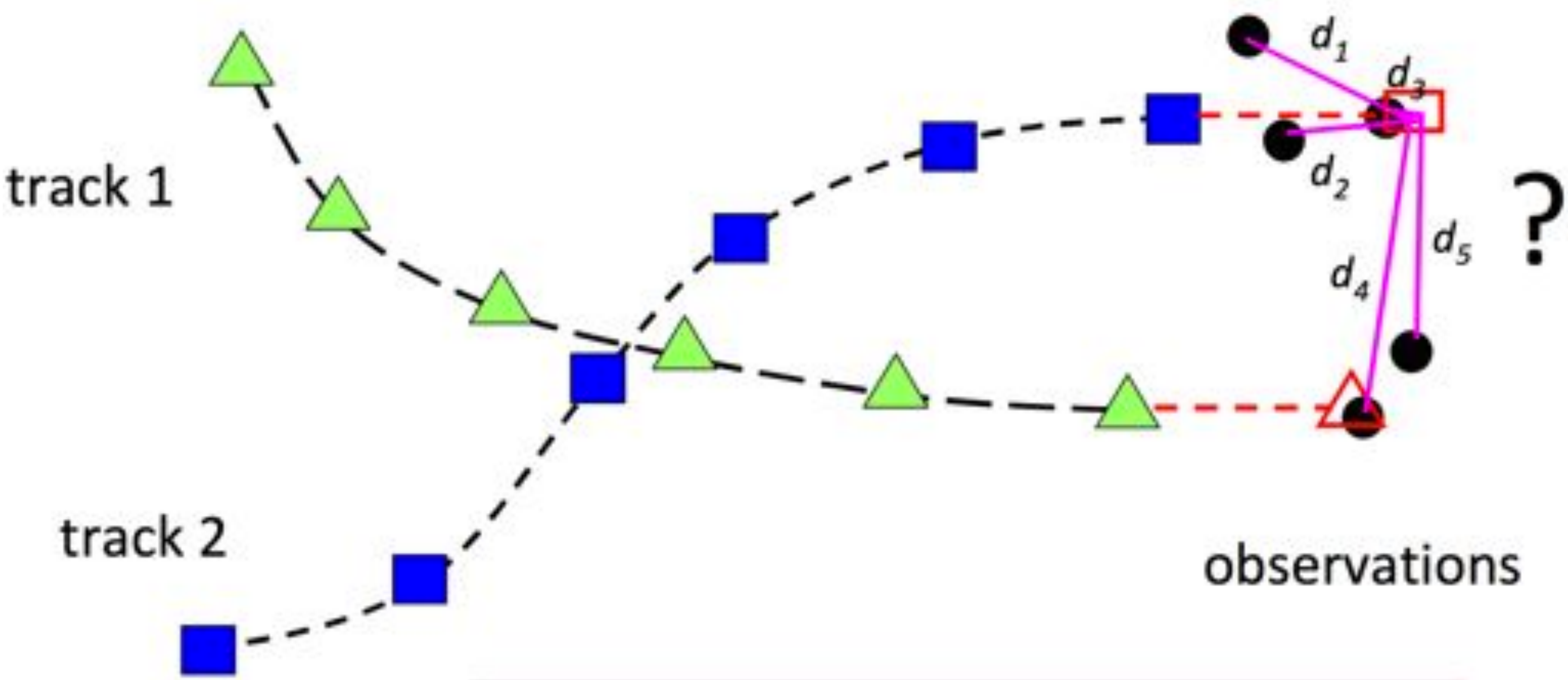
Slide from Collins, PSU

Multi-target tracking



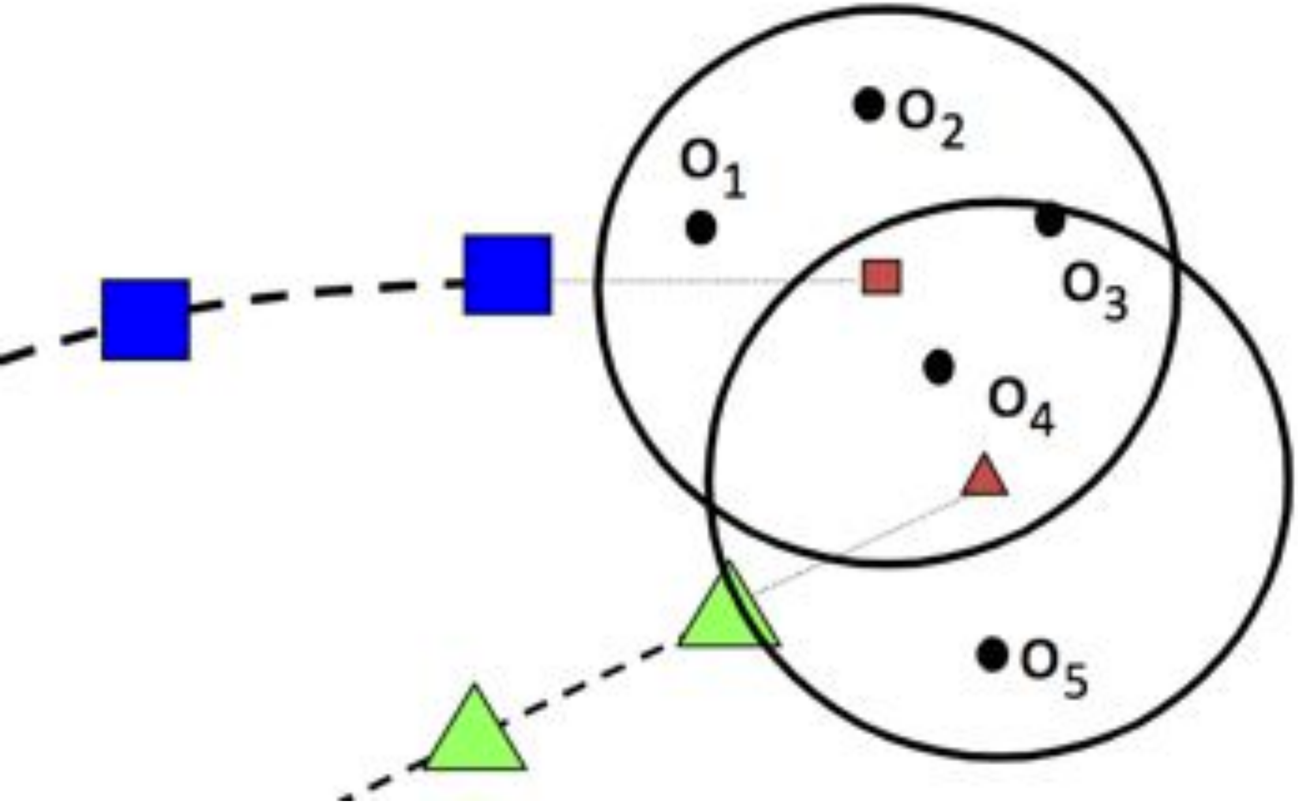
Slide from Collins, PSU

Multi-target tracking



Slide from Collins, PSU

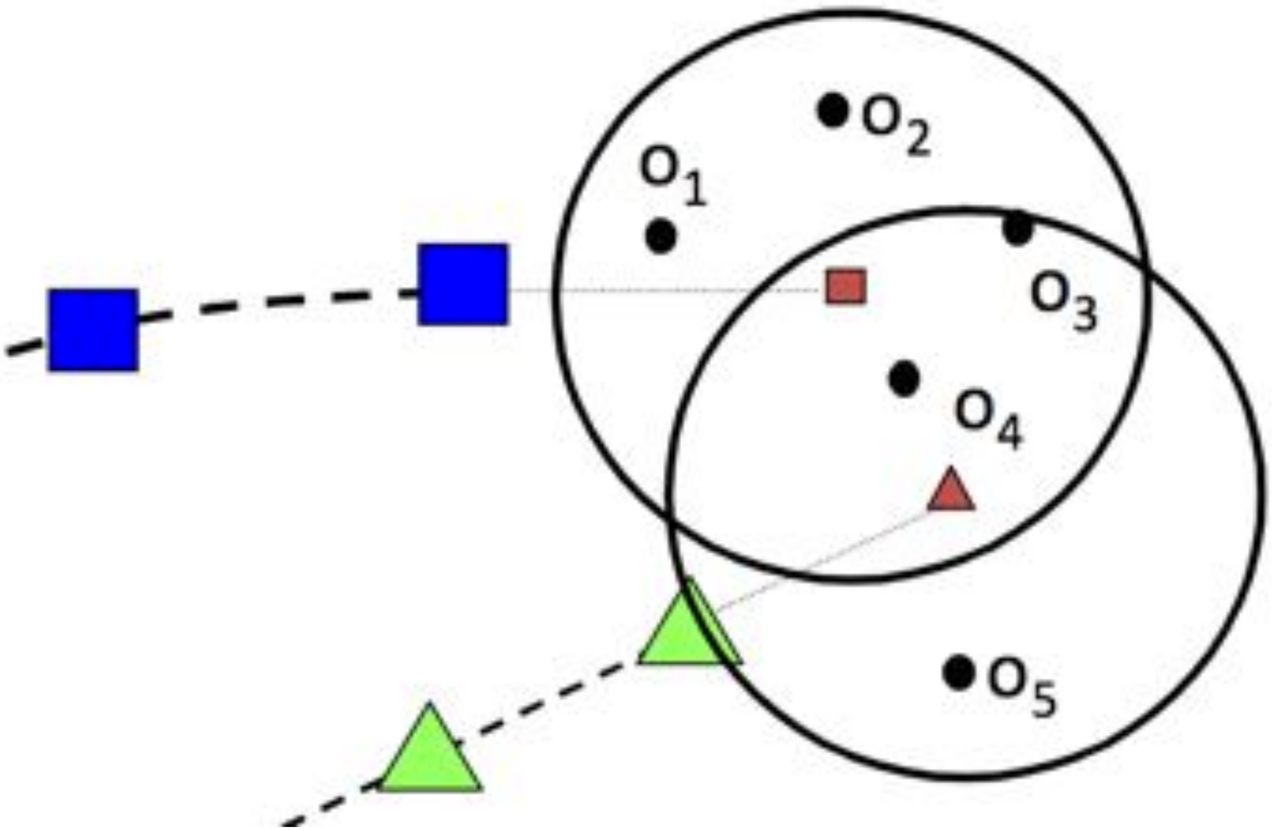
Multi-target tracking



	ai1	ai2
1	3.0	
2	5.0	
3	6.0	1.0
4	$\boxed{9.0}$	$\boxed{8.0}$
5		3.0

Slide from Collins, PSU

Multi-target tracking



	ai1	ai2
1	3.0	
2	5.0	
3	6.0	1.0
4	9.0	8.0
5		3.0

Non-optimal!

Slide from Collins, PSU

Multi-target tracking

- Mathematical definition

maximize: $\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}x_{ij}$

subject to:

$$\left.\begin{array}{ll} \sum_j x_{ij} = 1; & i = 1,2,\ldots,n \\ \sum_i x_{ij} = 1; & j = 1,2,\ldots,n \\ x_{ij} \in \{0,1\} & \end{array}\right\}$$

constraints that say
X is a permutation matrix

Where w is the affinity matrix and x is the assignments

Hungarian algorithm finds the optimal assignment

Slide from Collins, PSU

Multi-target tracking

	1	2	3	4	5
1	$\boxed{0.95}$	0.76	0.62	0.41	0.06
2	0.23	0.46	0.79	$\boxed{0.94}$	0.35
3	0.61	0.02	$\boxed{0.92}$	0.92	0.81
4	0.49	$\boxed{0.82}$	0.74	0.41	0.01
5	0.89	0.44	0.18	0.89	$\boxed{0.14}$

Greedy Solution

Score=3.77

	1	2	3	4	5
1	$\boxed{0.95}$	0.76	0.62	0.41	0.06
2	0.23	0.46	$\boxed{0.79}$	0.94	0.35
3	0.61	0.02	0.92	0.92	$\boxed{0.81}$
4	0.49	$\boxed{0.82}$	0.74	0.41	0.01
5	0.89	0.44	0.18	$\boxed{0.89}$	0.14

Optimal Solution

Score=4.26

Slide from Collins, PSU

Multi-target tracking

- Hungarian algorithm

- Pro
  - Optimal single frame assignment
- Con
  - Not optimal for multiple frames

Multi-target tracking

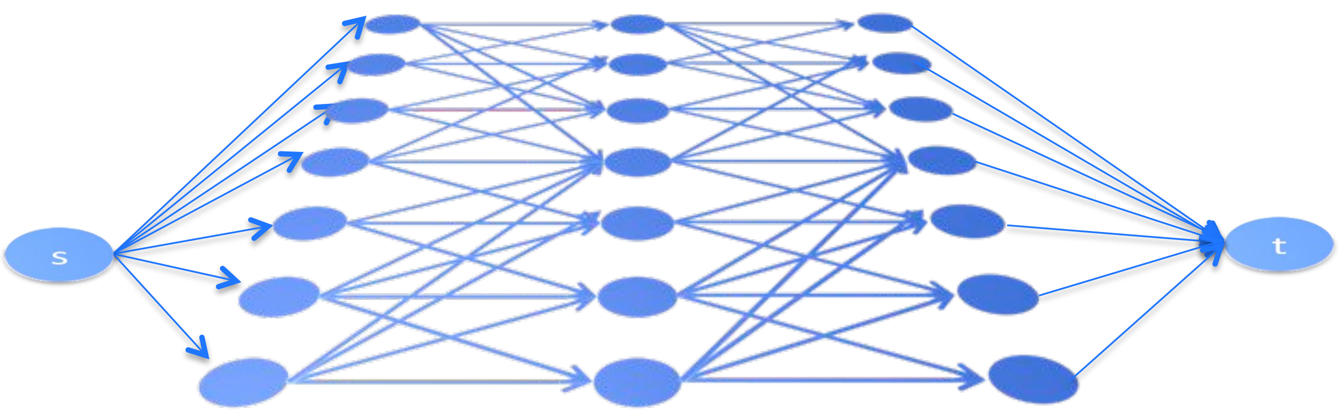
- Goal: seek a globally optimal solution across several frames

Multi-target tracking

Objective: minimum cut maximum flow

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

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



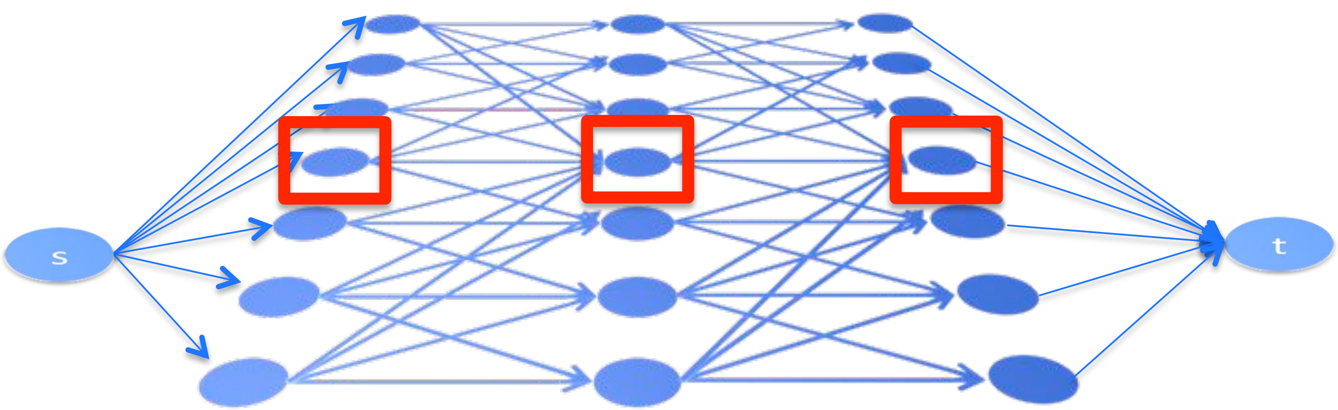
Where α_{i}, β_{ij}, γ_{OD} are the costs,
and f_i the flows

Multi-target tracking

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:

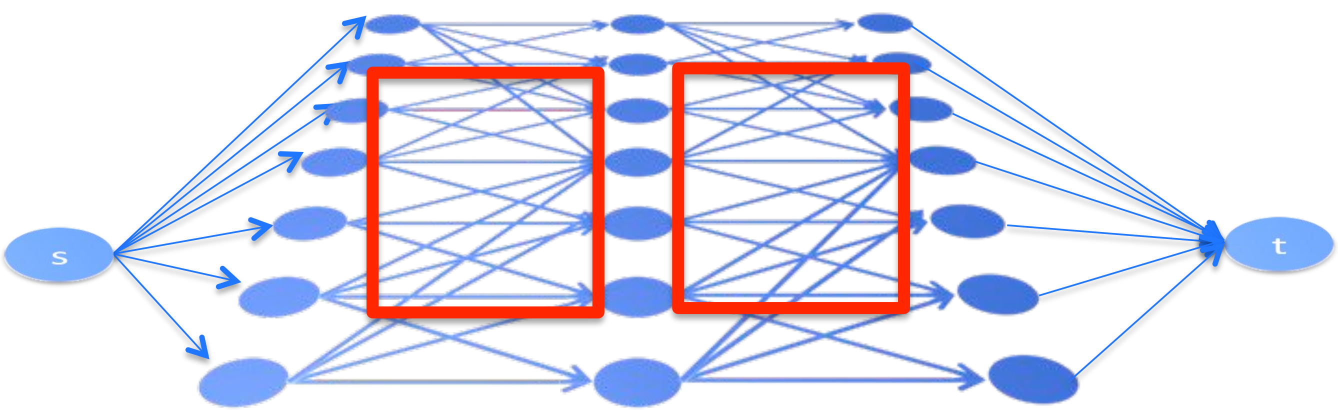
- Detection likelihood

Multi-target tracking

Objective: minimum cut maximum flow

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

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

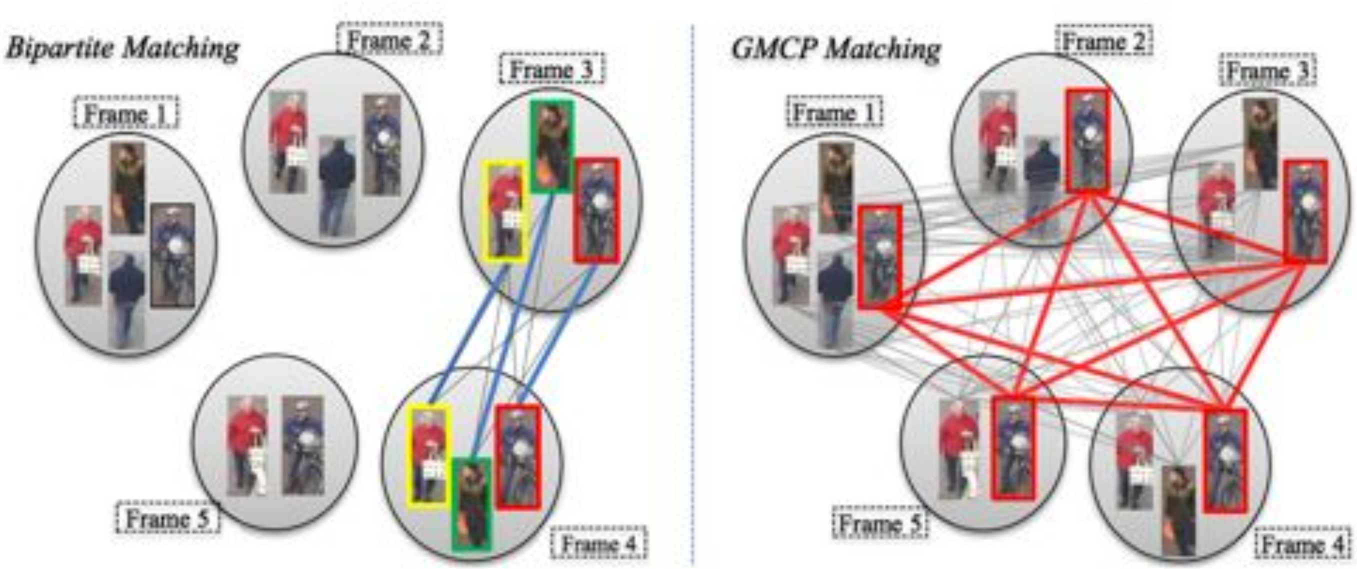


Cost β_{ij} based:

- spatial
- velocity

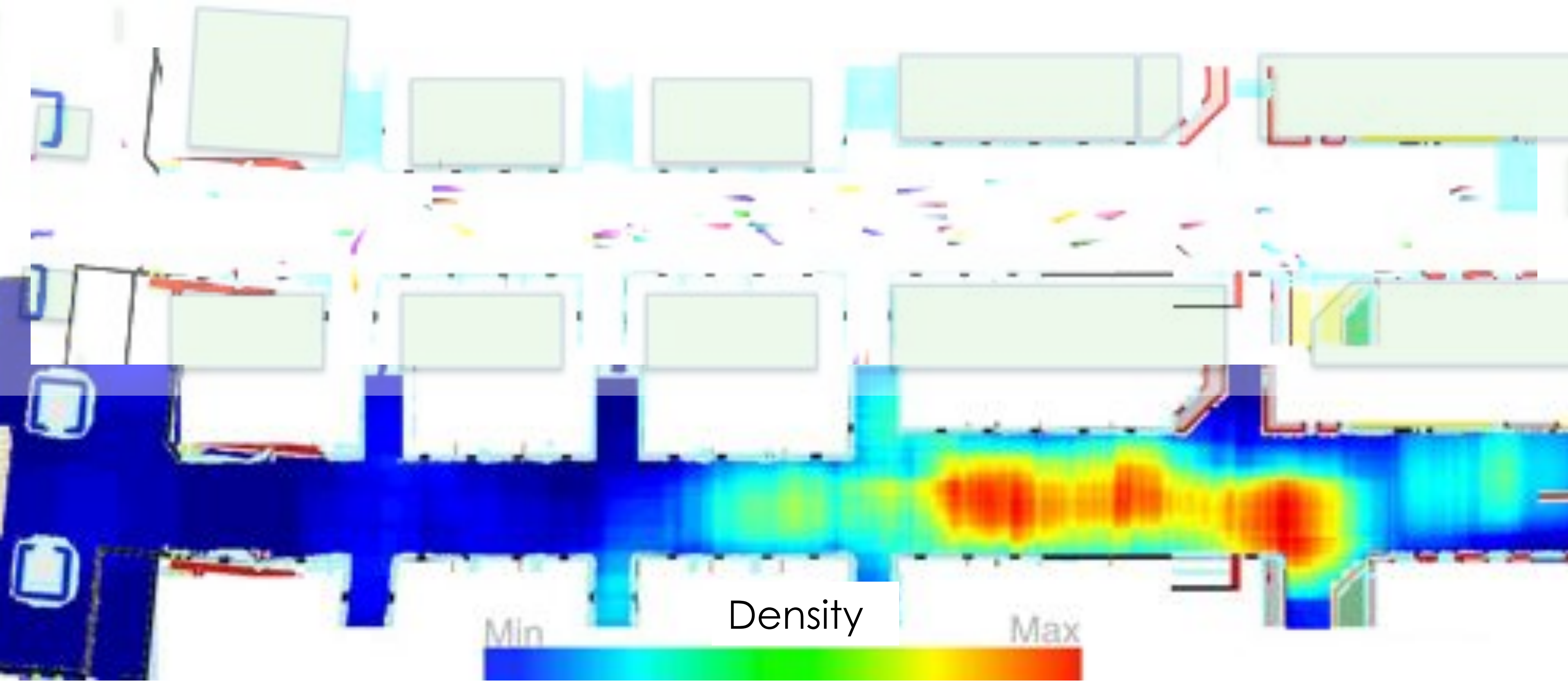
Multi-target tracking

- Optimal assignment for fully connected graph [1]



[1]Zamir, Amir Roshan et. al. "Gmcp-tracker: Global multi-object tracking using generalized minimum clique graphs." ECCV 2012

42 million of collected trajectories



Outline

1. Problem statement
2. Challenges
3. Object representation
4. Single target tracking
5. Multi-target tracking
6. **Tips** & references

Tips

- Model context
(a popular strategy since early 90s in CV community)
- Discriminative learning
- Sparsity driven




Some readings

- Tracking by matching
  - Isard, Michael, and Andrew Blake. "Condensation— conditional density propagation for visual tracking." International journal of computer vision 29.1 (1998): 5-28.
  - S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan. Locally Orderless Tracking. In CVPR, 2012
- Tracking by matching with an extended appearance model
  - D. Ross, J. Lim, R.-S. Lin, and M.-H. Yang. Incremental Learning for Robust Visual Tracking. IJCV, 77(1):125–141, 2008.
- Tracking with sparsity constraint
  - W. Zhong, H. Lu, and M.-H. Yang. Robust Object Tracking via Sparsity-based Collaborative Model. In CVPR, 2012.
  - Kwon, Junseok, and Kyoung Mu Lee. "Visual tracking decomposition."Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010.
  - Li, Hanxi, Chunhua Shen, and Qinfeng Shi. "Real-time visual tracking using compressive sensing." Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011.

Some readings

- Tracking by detections (ML approach, using a discriminative classification)
  - Babenko, Boris, Ming-Hsuan Yang, and Serge Belongie. "Visual tracking with online multiple instance learning." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.
  - Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-Learning-Detection," Pattern Analysis and Machine Intelligence 2011.
  - S. Hare, A. Saffari, and P. H. S. Torr. Struck: Structured Output Tracking with Kernels. In ICCV, 2011.
  - F. Henriques, R. Caseiro, P. Martins, and J. Batista. Exploiting the Circulant Structure of Tracking-by-Detection with Kernels. In ECCV, 2012
  - Nebehay, Georg, and Roman Pflugfelder. "Consensus-based matching and tracking of keypoints for object tracking." Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on. IEEE, 2014.

Some readings

- Multi-target tracking (data association)
  - Berclaz, Jerome, et al. "Multiple object tracking using k-shortest paths optimization." Pattern Analysis and Machine Intelligence, IEEE Transactions on 33.9 (2011): 1806-1819.
  - Pirsiavash, Hamed, Deva Ramanan, and Charless C. Fowlkes. "Globally-optimal greedy algorithms for tracking a variable number of objects." (CVPR), 2011
  - Zamir, Amir Roshan, Afshin Dehghan, and Mubarak Shah. "Gmcp-tracker: Global multi-object tracking using generalized minimum clique graphs." Computer Vision–ECCV 2012. Springer Berlin Heidelberg, 2012. 343-356.
  - Liu, Jingchen, et al. "Tracking sports players with context-conditioned motion models." (CVPR), 2013.