Probabilistic Machine Learning Models for Intelligent Sensing: Computer Vision and Beyond

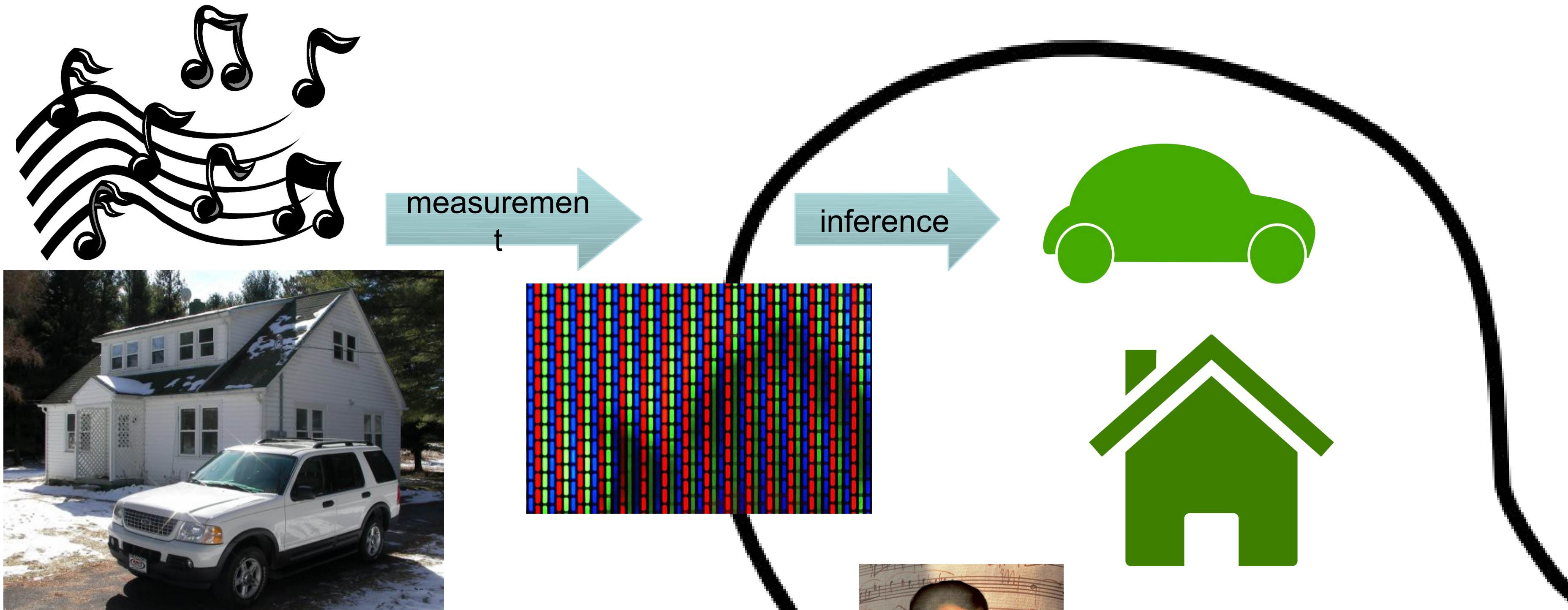
Dr. Timothy Hospedales

Centre for Intelligent Sensing
Queen Mary University of London

CIS centre for intelligent sensing

Queen Mary University of London

Intelligent Sensing?



Challenges:

- Ill posed
- How to constrain the parameters?
- Resolve conflicts

Intelligent Sensing?

Intelligent Sensing

- How to resolve different sources of data?
- How to solve ill-posed problems?
- How to not start from scratch every time?
- How to make the right measurements?

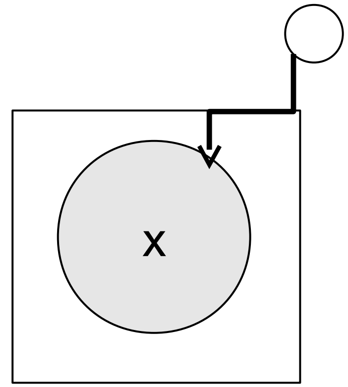
Outline

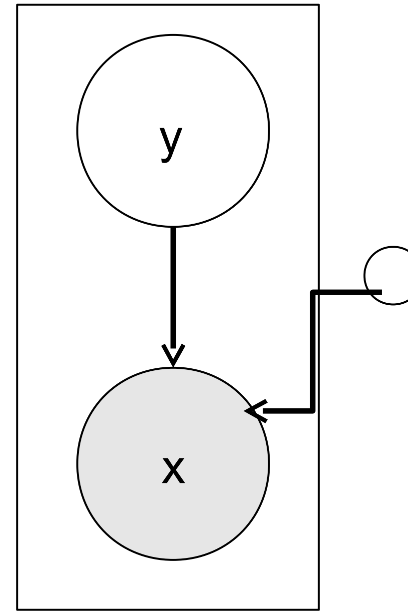
Intelligent Sensing

- **Data Fusion**
- Data and annotation efficiency
  - Unsupervised
  - Weakly-supervised
  - Semi-supervised
  - Multi-label/multi-instance
  - Zero-shot learning
- Observation efficiency
  - Active learning

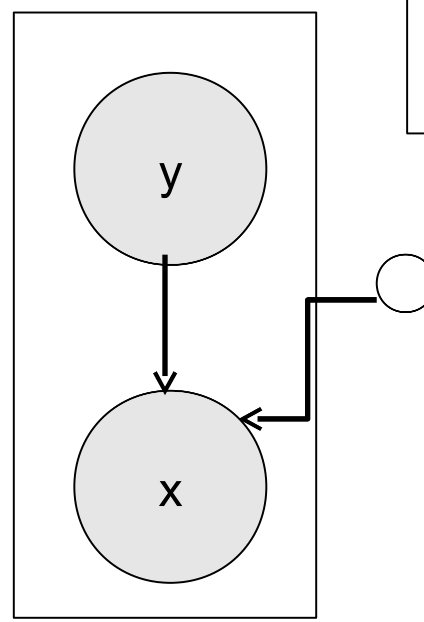
Notation

- Unsupervised Learning
  - Observe {x}, model p(x)
  - Observe {x}, model p(x,y)





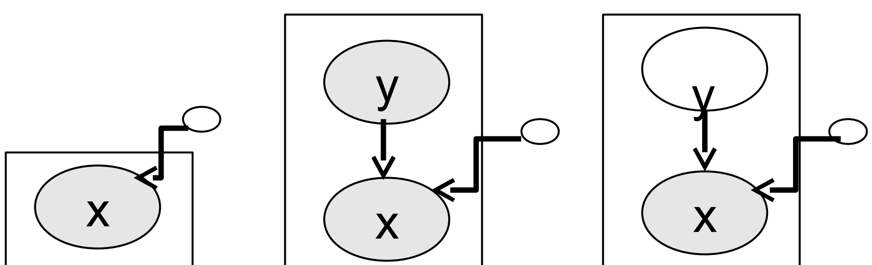
- Supervised Learning
  - Observe {x,y}, model p(x,y)



Machine Learning

Classic Problems:

- Inference $$p(y \mid x) = p(x \mid y)p(y) / p(x)$$
- Marginal Likelihood $$p(x) = \int p(x, y)\, dy$$
- ML Learning, $$\hat{\theta} = \operatorname{argmax} p(X \mid \theta)$$
  - Density Estimation
- EM Learning $$\hat{\theta} = \operatorname{argmax} \int p(X, Y \mid \theta)\, dY$$
- Model Selection $$M = \operatorname{argmax} \int p(X, Y, \theta \mid M) p(M)\, dY\, d\theta$$



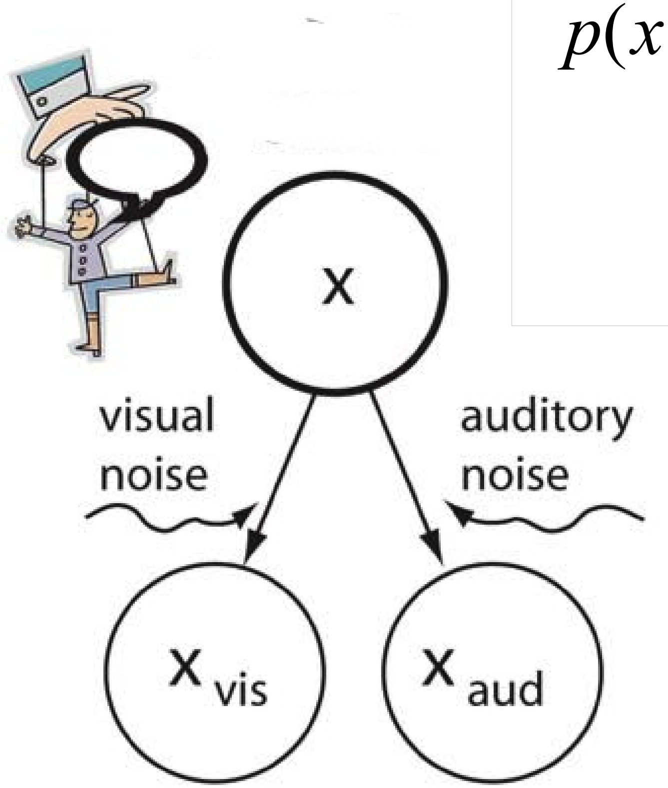
CIS centre for intelligent sensing

Queen Mary University of London

- Fusion and Data Association

Multisensory perception

Fusing Multiple Data Sources



$$p(x \mid x_a, x_v) \propto p(x_a, x_v \mid x)p(x)$$
$$= p(x_a \mid x)p(x_v \mid x)p(x)$$

Optimal fusion depends on reliability of each modality

But how do you know the reliability of your senses?

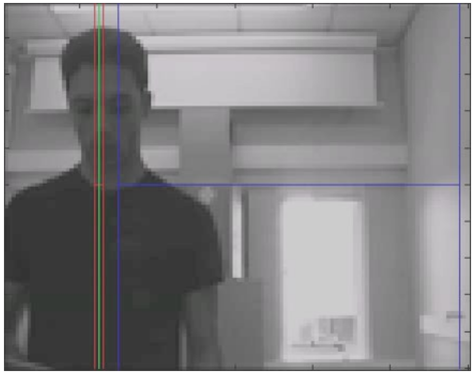
Fusing Multiple Data Sources

Aim: Given an audio-visual stream

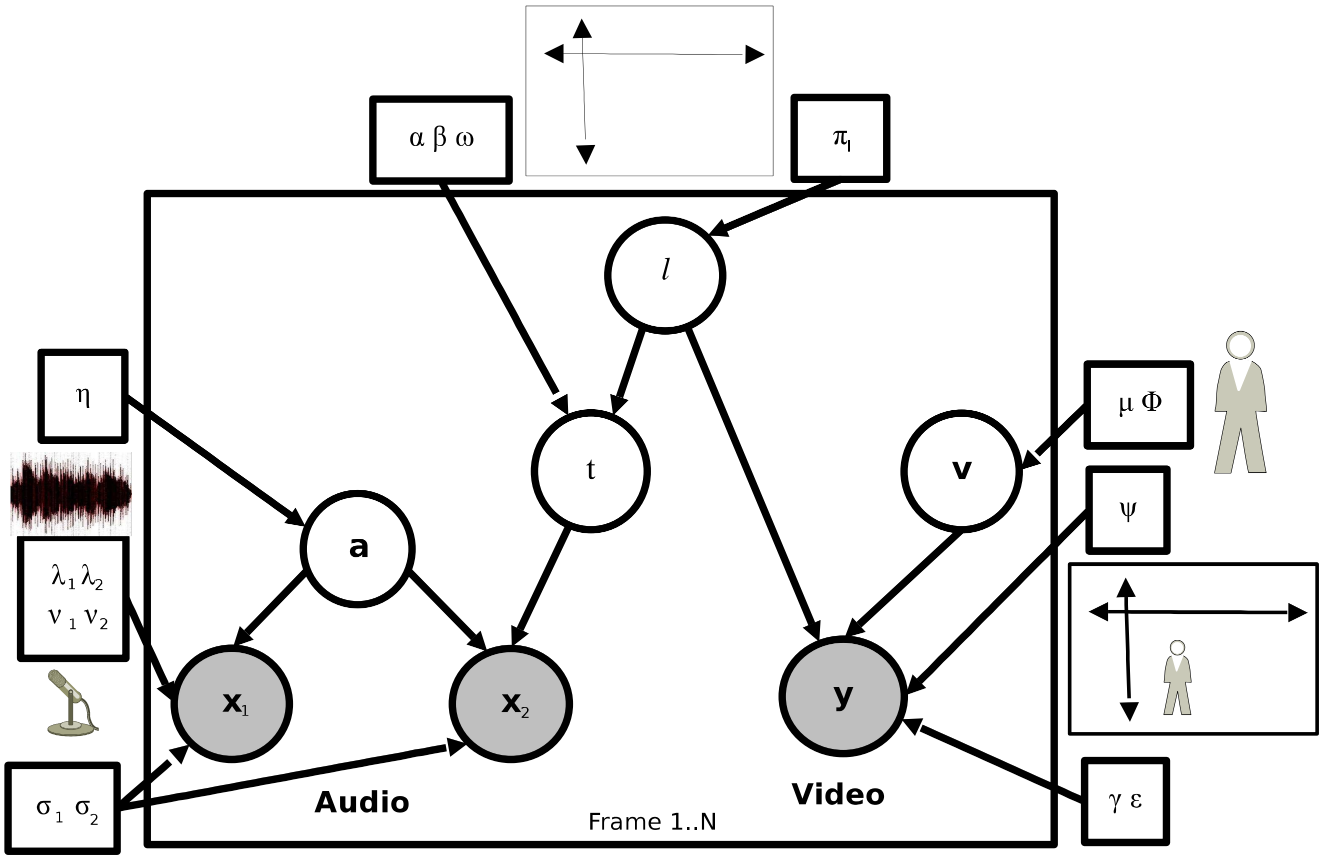
- Learn the user's appearance & sound
- Learn the microphone and camera characteristics
- Fuse appearance and sound info for optimal localization

Challenges:

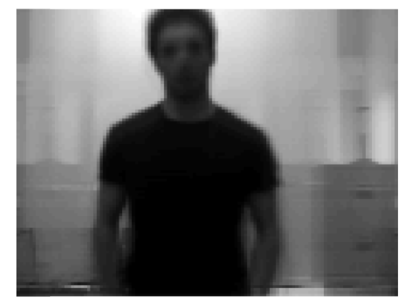
- No supervision
- (No background subtraction)
- Real-time inference



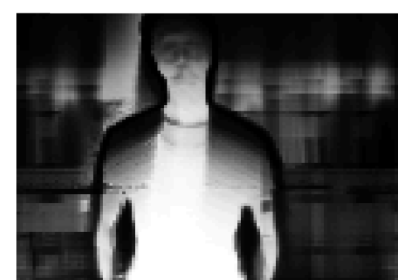
Fusing Multiple Data Sources



(f) Learned Template Mean



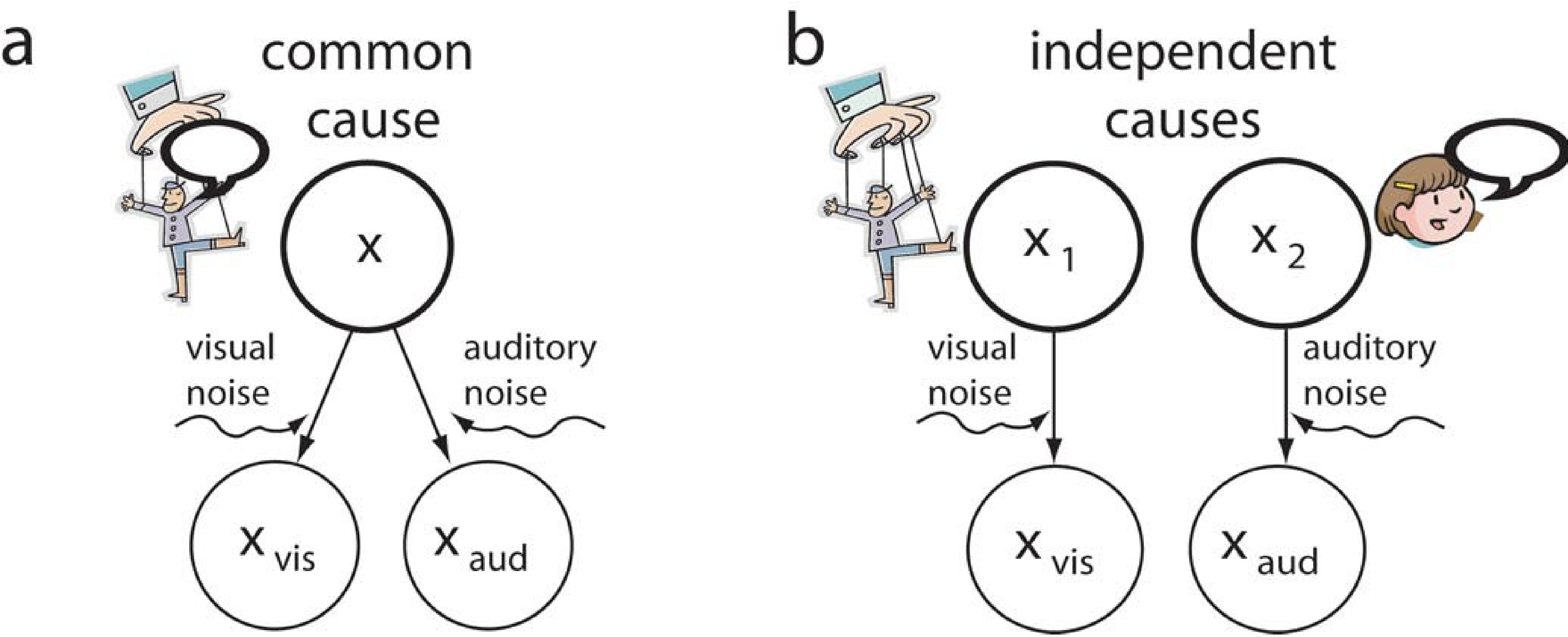
(g) Learned Template Precision



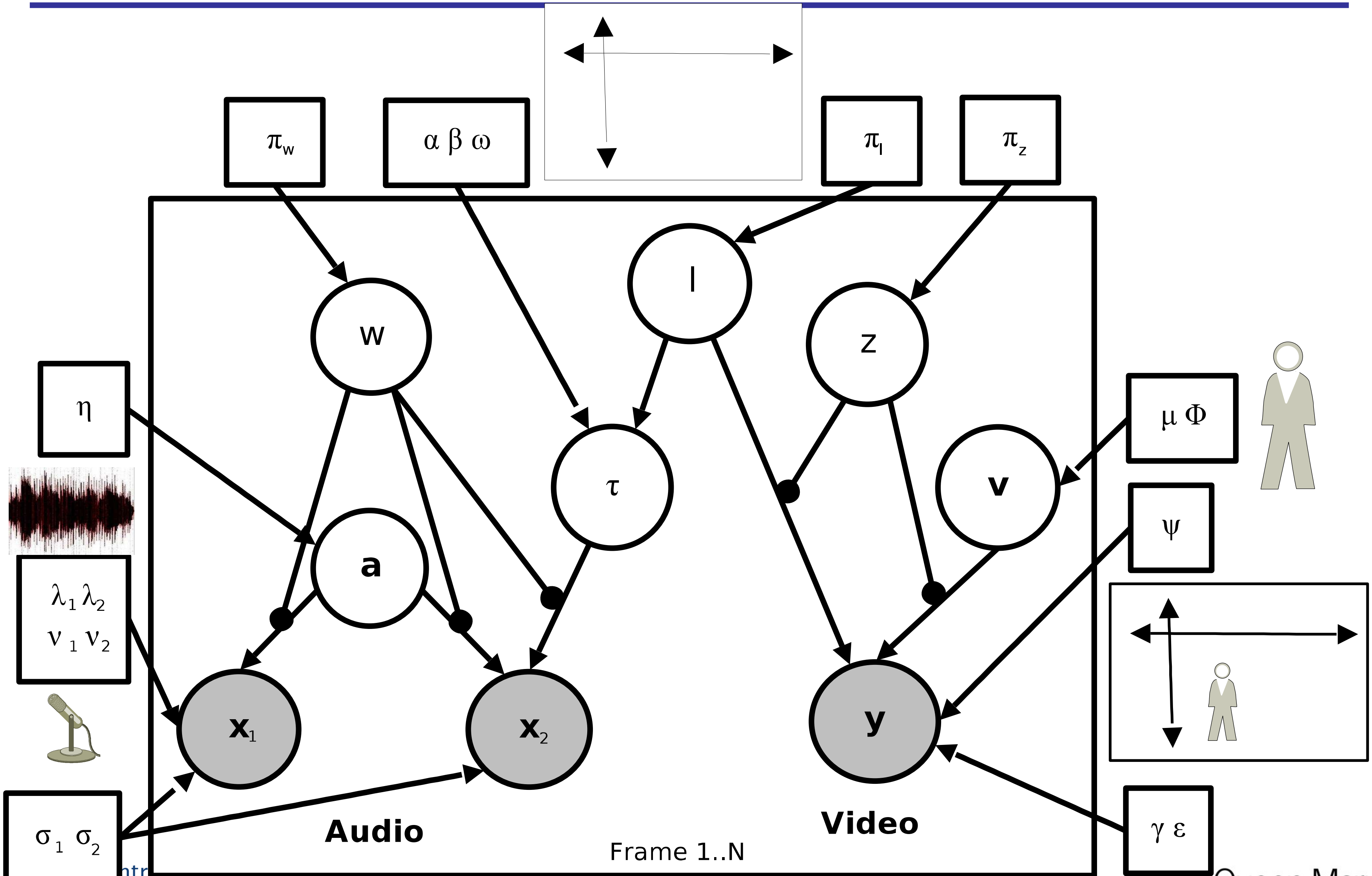
[IJCAI'07, PAMI'08]

CIS centre for intelligent sensing

Fusion without correspondence?



Multiple Data Sources?



[IJCAI'07, PAMI'08]

Fusion without correspondence?

EM Learning

Model Selection

CIS centre for intelligent sensing

[IJCAI'07, PAMI'08]

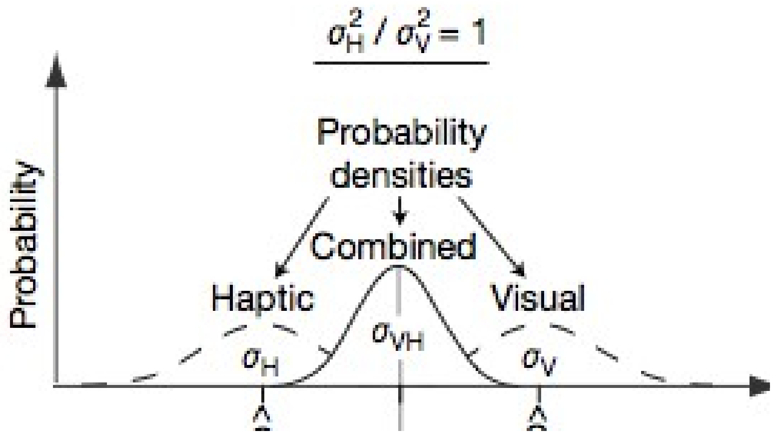
Queen Mary University of London

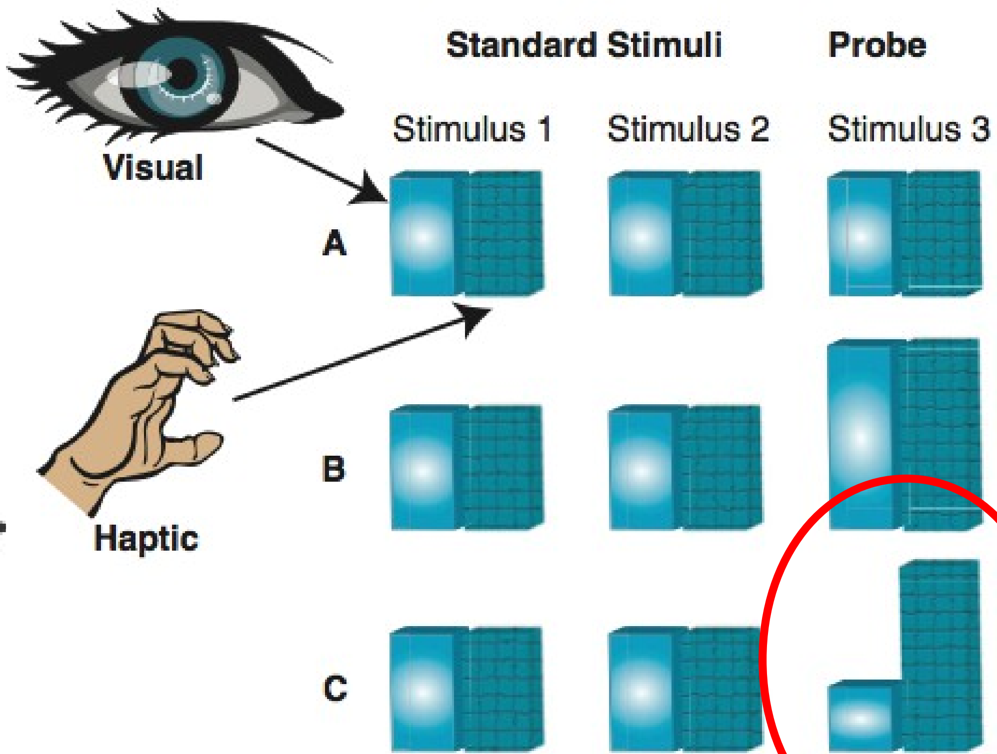
- Abnormality Detection

Multisensory perception

Human Multisensory Oddity detection

Optimal Fusion





$$\hat{S}_c = w_1 \hat{S}_1 + w_2 \hat{S}_2$$

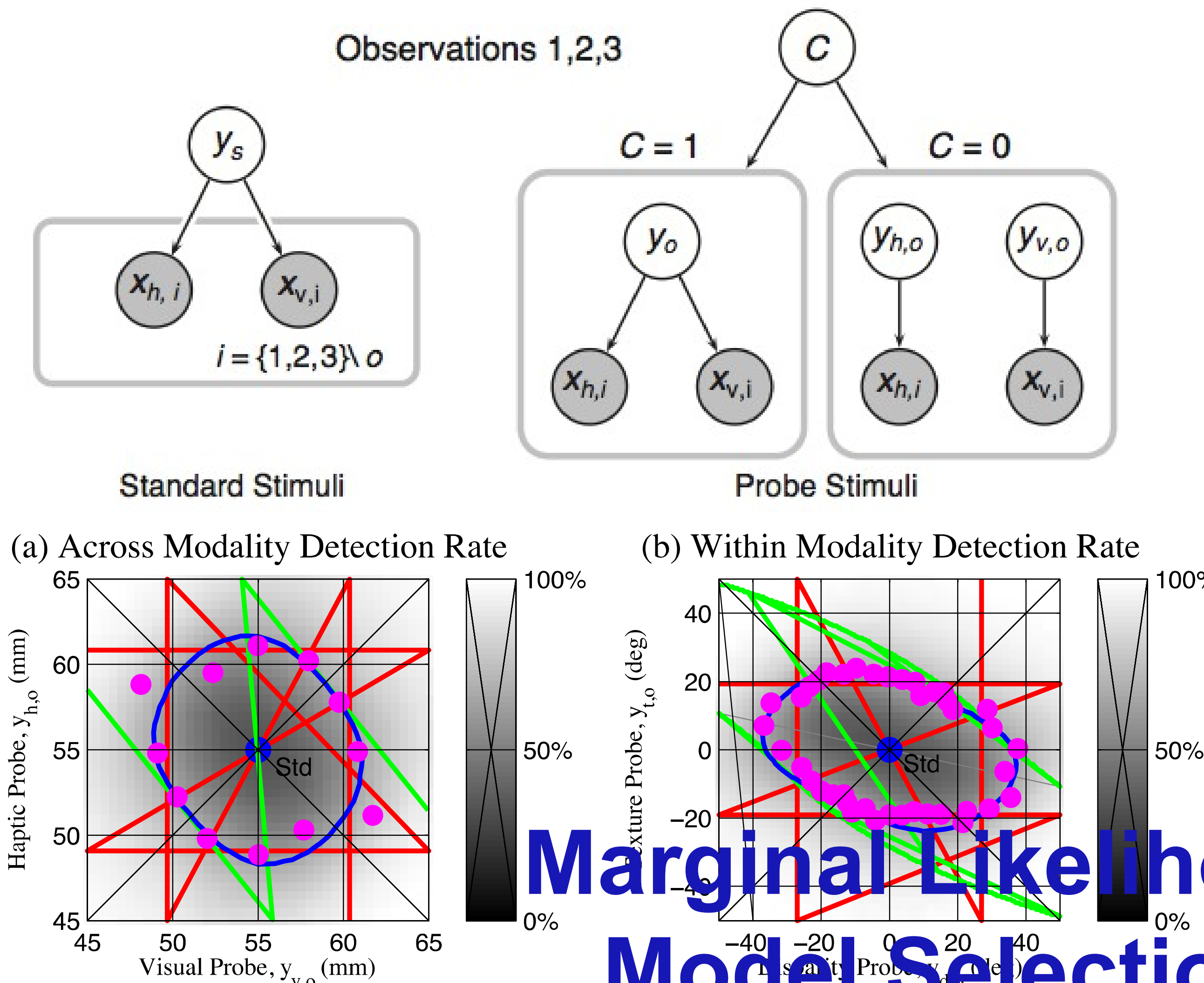
People (eventually) notice this

Hillis et al, Science, 2002
Banks et al, Nature, 2002

CIS centre for intelligent sensing

Queen Mary University of London

Human Multisensory Oddity detection



Marginal Likelihood Model Selection

[PLOS'09, OUP'11]

Data Fusion

- Optimally integrating multiple sensors
- Resolving multi-sensor data association

Outline

Intelligent Sensing

- Data Fusion
- **Data and annotation efficiency**
  - Unsupervised
  - Weakly-supervised
  - Semi-supervised
  - Multi-label/multi-instance
  - Zero-shot learning
- Observation efficiency
  - Active learning

- Video Surveillance: Anomaly Detection & Clustering

Unsupervised learning

Unsupervised Learning / Surveillance

- Aim: Given a video stream
  - Get domain knowledge by learning about activities
  - Detect abnormal behaviors as outliers against the model
- Challenges:
  - No tracking
  - No camera calibration
  - No supervision
  - Complex behaviors
  - Real-time

CIS centre for intelligent sensing

[ICCV’09, IJCV’11]

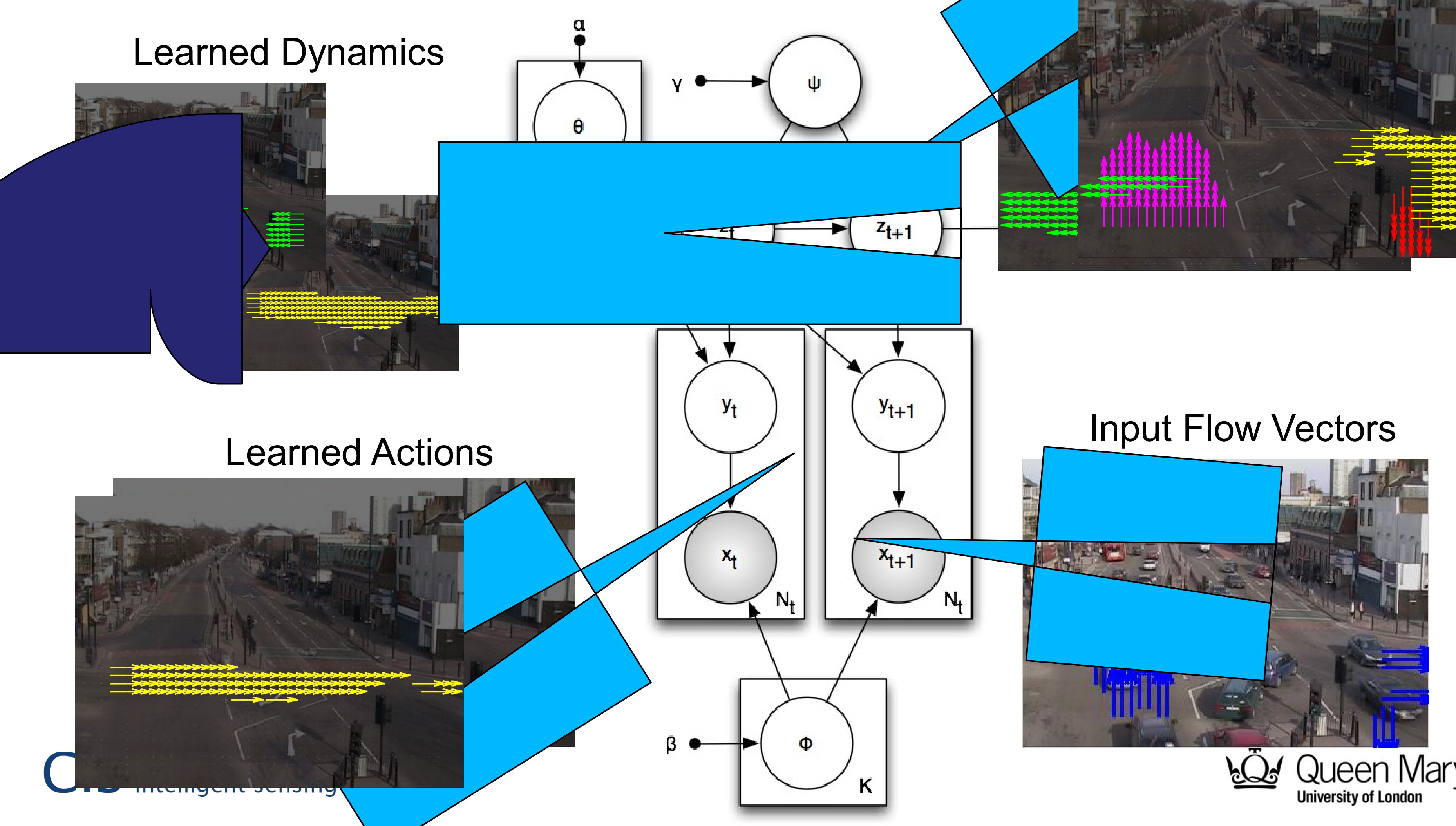
Queen Mary University of London

Unsupervised / Surveillance: MCTM

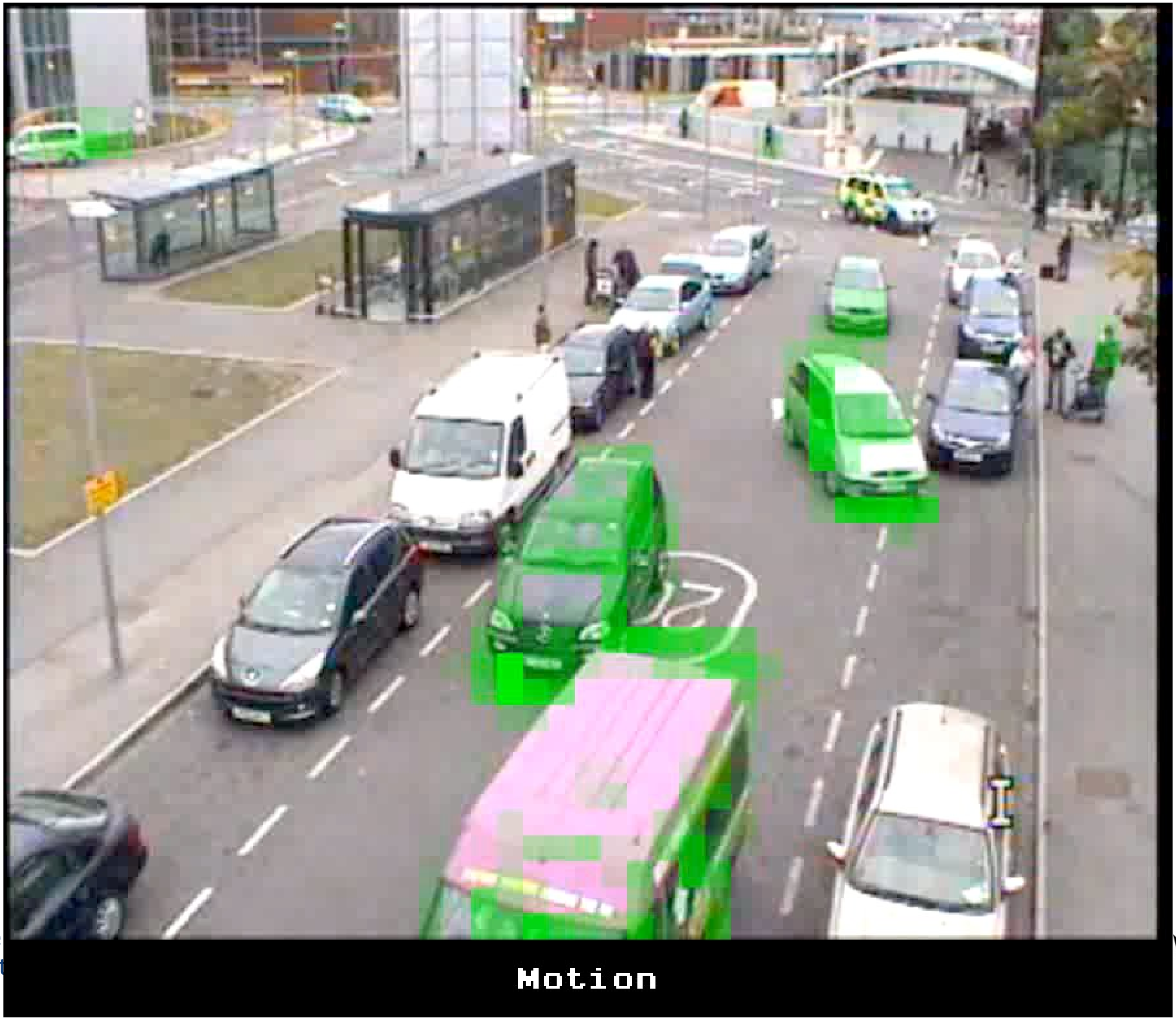
New generative model…

“Markov Clustering Topic Model”

ed Behaviors



Unsupervised / Surveillance: MCTM Learning



CIS

Queen Mary
University of London

Unsupervised / Surveillance: MCTM: Abnormality

$$p(x) = \int p(x, y)\, dy$$

$$p(y \mid x)$$

EM Learning

Inference

Marginal Likelihood

CIS centre for intelligent sensing

[ICCV'09, IJCV'11]

Queen Mary University of London

Unsupervised Learning

- What if you have seen a few known (but rare) behaviors you want to detect?
  - E.g., surveillance
- These may also be subtle

Weakly Supervised Learning / Surveillance

- Aim: Given a video stream
  - Learn a detector for a rare and subtle behavior
- Challenge
  - Use only weak annotation
  - Sparse training data
  - No tracking
  - Real-time

- View as hard multi-instance learning (MIL) problem

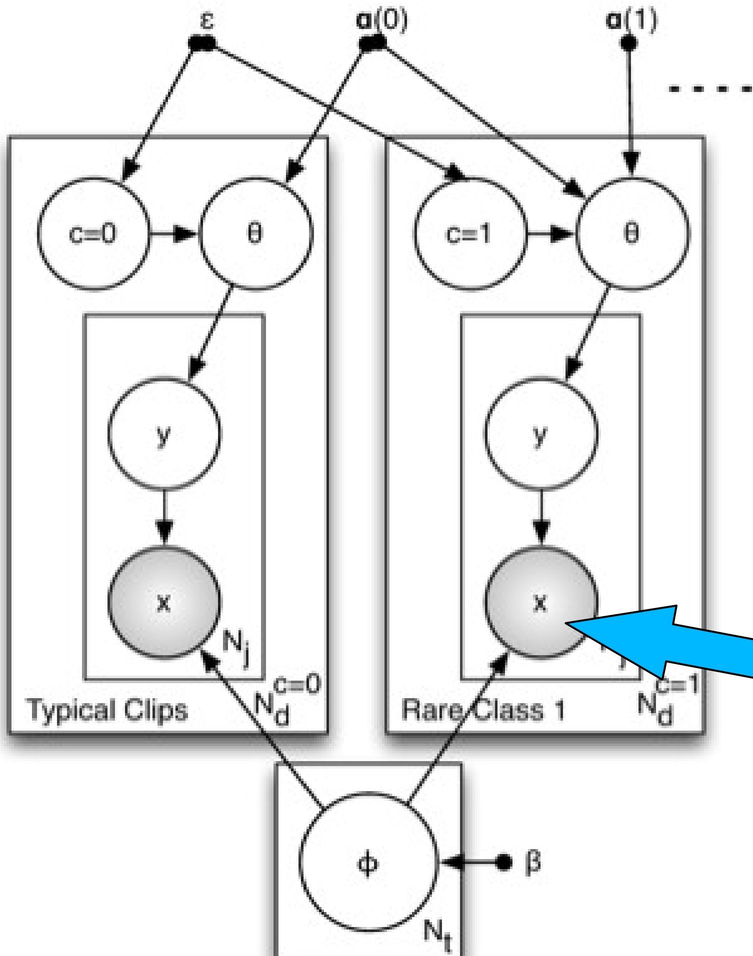
WSL / Surveillance: Rare Events

Example Challenge

1 Example Each

100 Examples

WSL / Surveillance: WSJTM



Classify: Compute p(C|X)

- Bayesian Model Selection
  - Variational Importance Sampler

Locate: Infer p(Y|X,C)

- Gibbs

WSL / Surveillance: WSJTM: Results

WSJTM Classifier: Trained with Weak and Sparse Labels

EM Learning

Model Selection

CIS centre for intelligent sensing

Queen Mary University of London

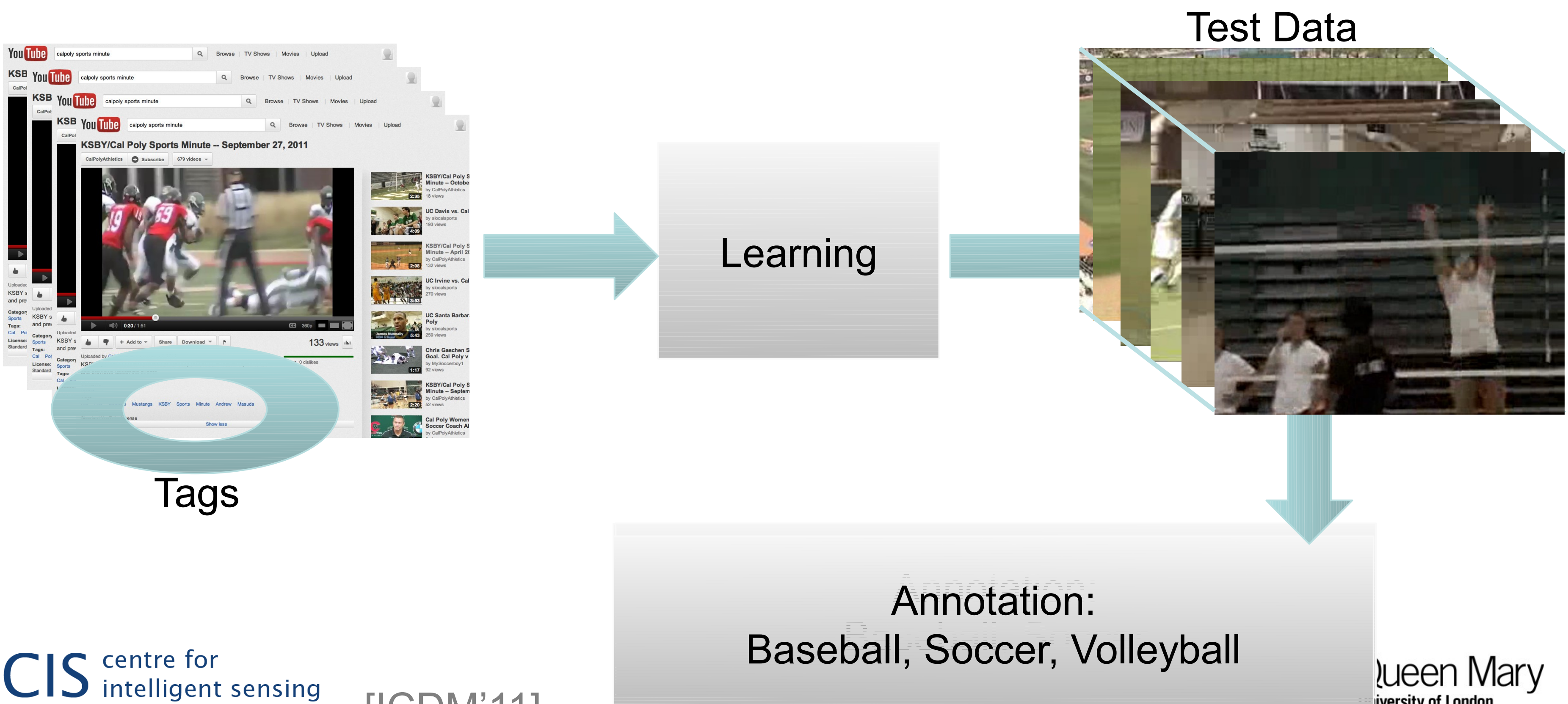
[ACCV'10, PAMI'11]

Weakly Supervised Learning

- What if you need more than one label per instance?
  - E.g., multi-media indexing.
- The weak supervision problem gets harder…

Weakly Supervised Multi Label: Tagging

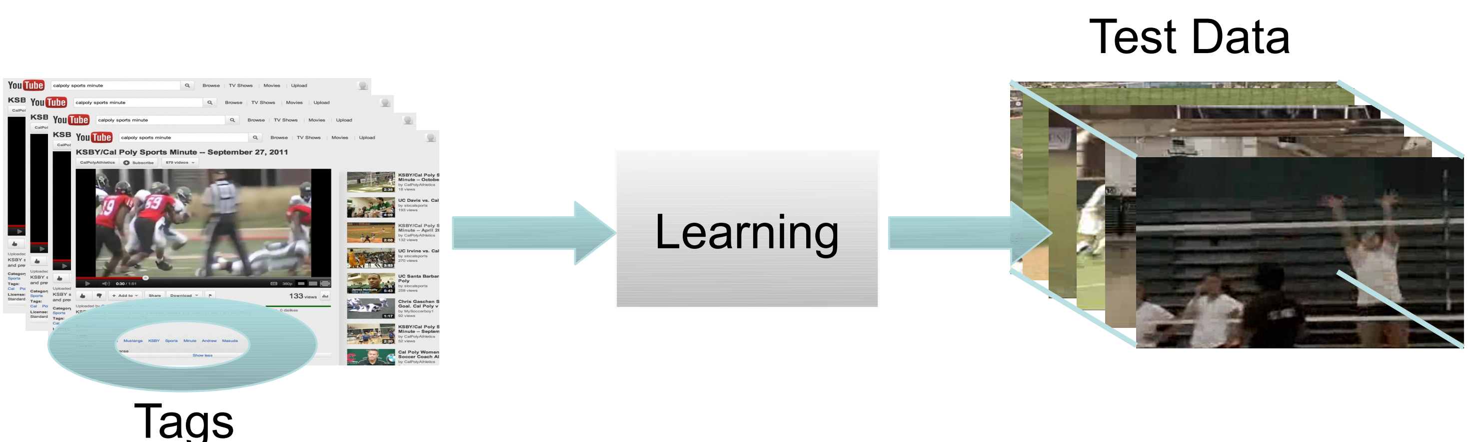
- Aim: Learn video annotation model from tags
  - Online video databases: Sports news



[ICDM'11]

Weakly Supervised Multi Label: Tagging

- Challenges:
  - Weak annotation
  - Multiple labels per instance
  - Huge intra-class variability



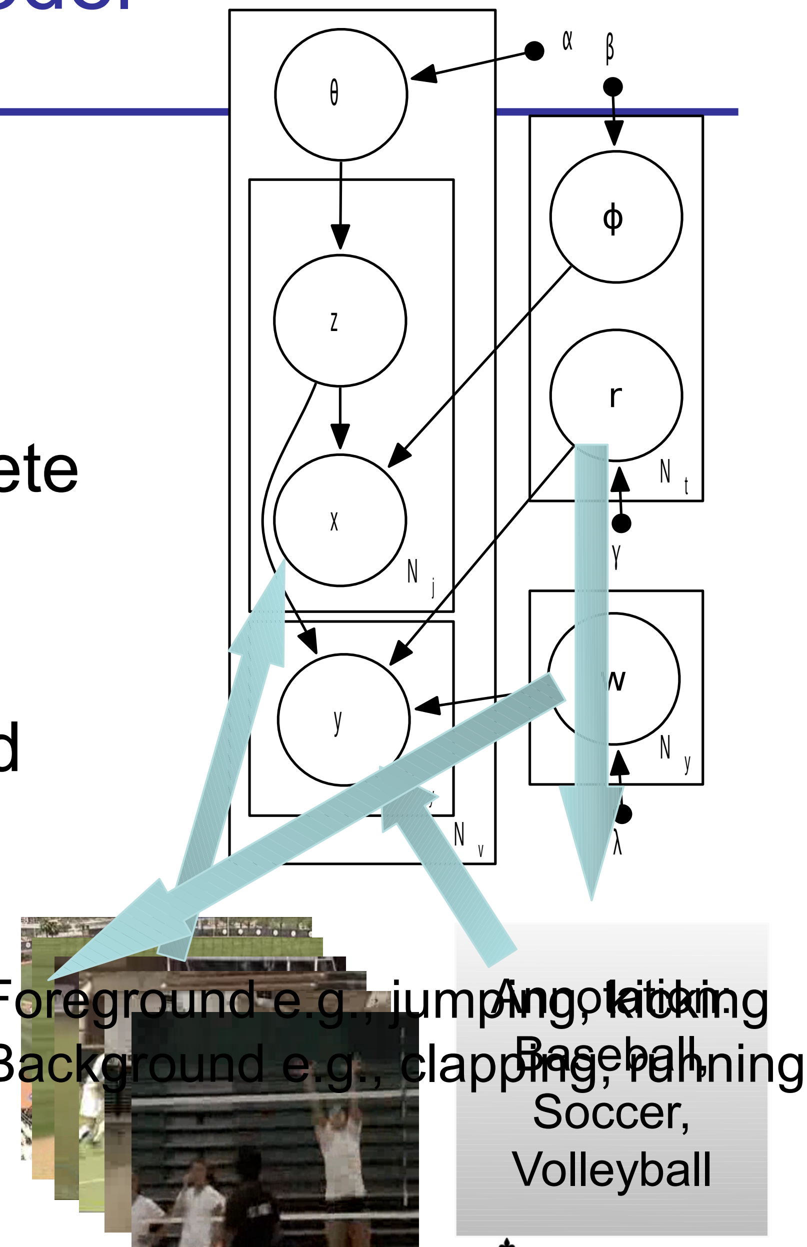
CIS centre for intelligent sensing

[ICDM'11]

Queen Mary University of London

Weak/Multi Label: VTT model

- Topic Model, e.g., LDA.
  - Density estimate for discrete data corpus p(X)
- Our VTT model
  - Joint estimate for data and tags p(X,Y)
  - Permits annotation p(y|x)



CIS centre for intelligent sensing

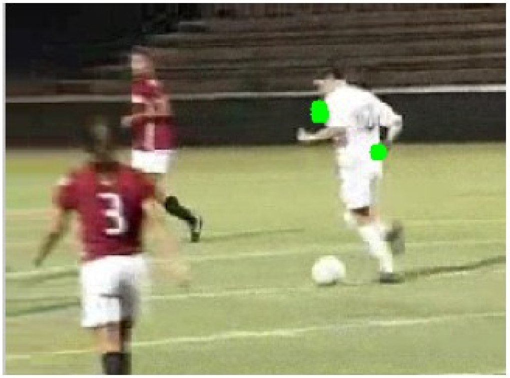
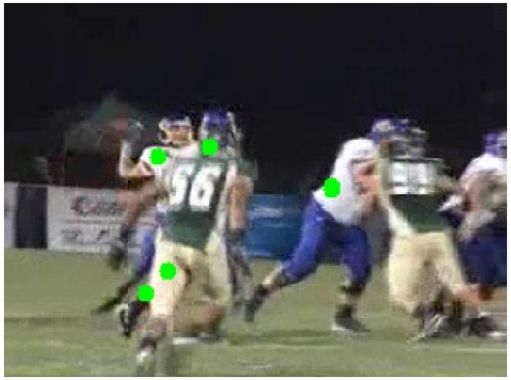
[ICDM'11]

Queen Mary University of London

Weakly Supervised Multi Label: Results

Weakly Supervised Multi Label: Insight

- Foreground Topics
  - Shared, e.g., running
  - Specific, e.g., pitching

EM Learning

Inference

[ICDM'11]

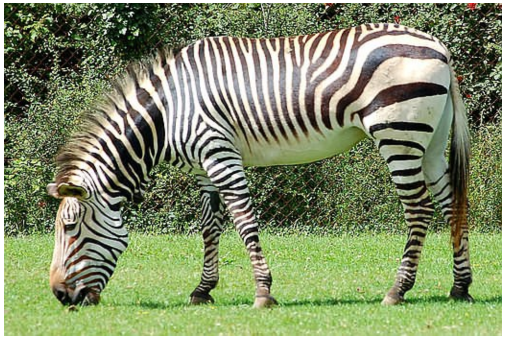
Weakly Supervised / Multi Label Learning

- Can we lean anything with no new data at all?
  - (Classification)

Zero-Shot Attribute Learning / Tag & Classify

- Aim: Given a set of Tags & Classes
  - Learn how to tag
  - Learn how to relate tags to classes
  - Zero-shot learning from tag description

Zero-Shot Learning (Look Mum! No Data!)

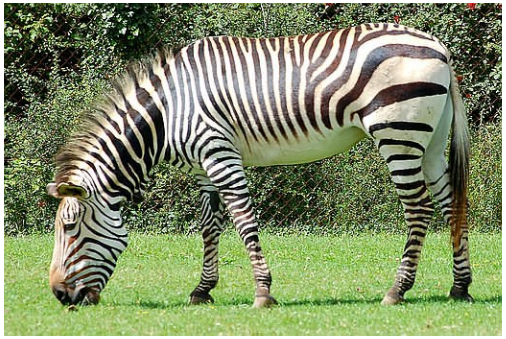


Stripes, Herbivore, Tail, Claws
➔Zebra

Lion!

Lion:= Stripes, Herbivore, Tail, Claws

Zero-Shot Learning (Look Mum! No Data!)



Stripes, Herbivore, Tail, Claws
➔Zebra

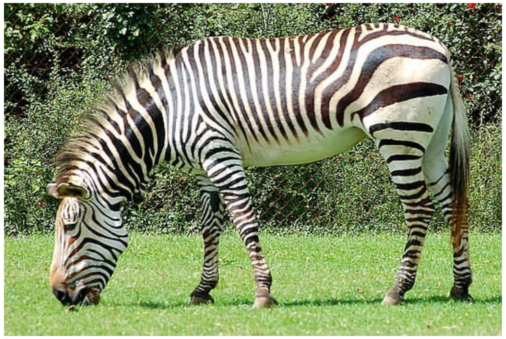
Lion!

Lion:= Stripes, Herbivore, Tail, Claws

Candles, Cake, Clapping, Dancing
➔Birthday Party

Wedding Dance:= Candles, Cake, Clapping, Dancing

Zero-Shot Learning (Look Mum! No Data!)



Stripes, Herbivore, Tail, Claws
➔Zebra

Lion!

Lion:= Stripes, Herbivore, Tail, Claws

Candles, Cake, Clapping, Dancing
➔Birthday Party

Wedding Dance!

Wedding Dance:= Candles, Cake, Clapping, Dancing

Latent Attribute Learning

- Aim: Given a set of Tags & Classes
  - Learn how to tag
  - Learn how to relate tags to classes
  - Zero-shot learning from tag description
- Challenge: Reduce human effort
  - Avoid annotating every attribute on every training image
  - Avoid specifying every attribute on every new class

Latent Attribute Learning

Classes

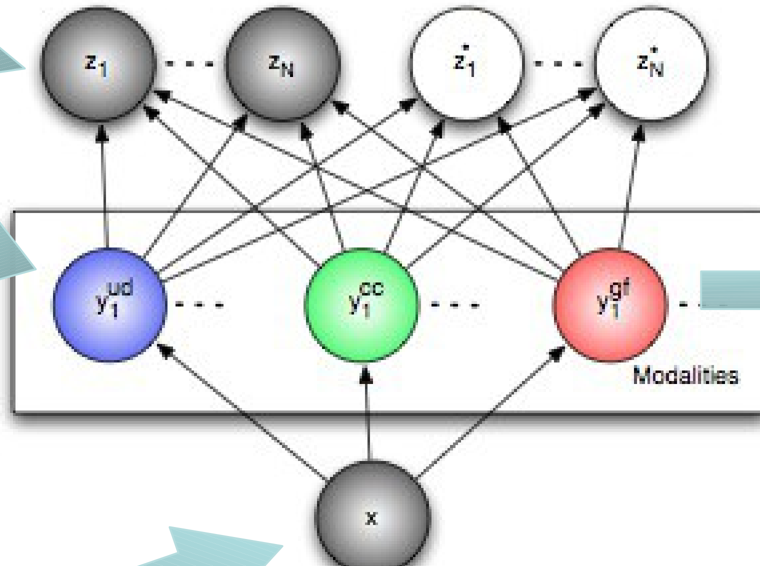
Lion
Zebra

Attributes

Stripes, Herbivore, Tail, Claws

Stripes, Herbivore, Tail, Claws

Image/Video



Latent Attributes

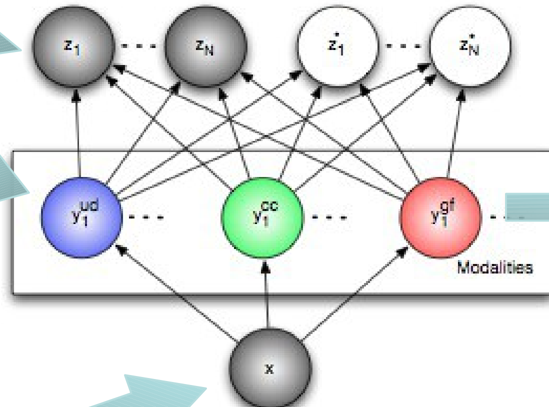
Teeth, Furry, Eats Grass

Teeth, Furry, Eats Grass

Latent Attribute Learning

Lion
Zebra

Stripes, Herbivore,
Tail, Claws
Stripes, Herbivore,
Tail, Claws



Teeth, Furry,
Eats Grass

Teeth, Furry,
Eats Grass

Bear := Eat Fish,
Claws, Tail

Teeth, Furry,
Eats Grass

Latent Attributes:

- Less annotation work
- Increased Classification Accuracy
  - Conventional
  - Zero-shot

EM Learning

[ECCV'12]

Zero-shot learning

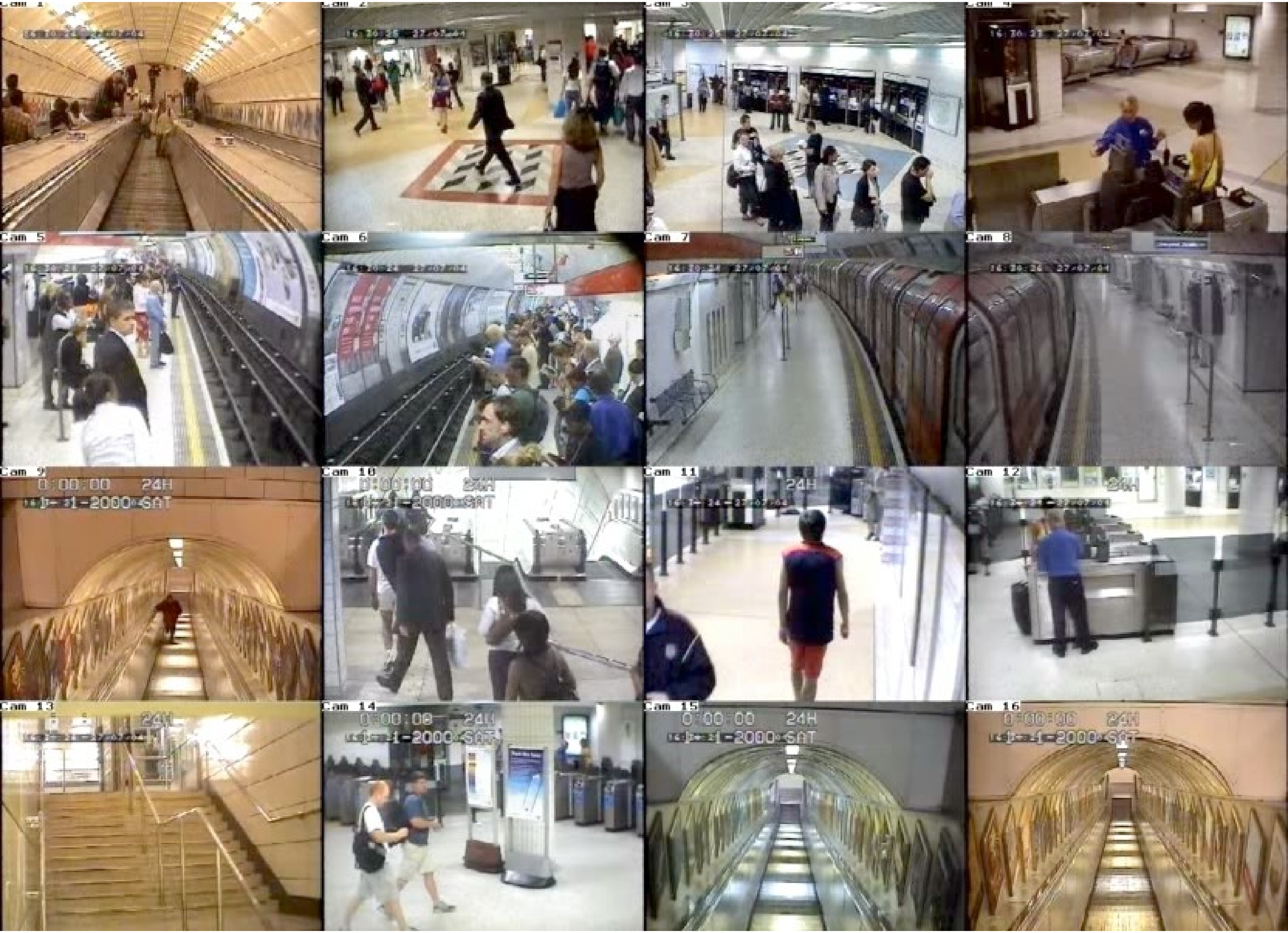
- We can also use attributes for sparse data learning
  - E.g., in re-identification.

More on attributes: Re-identification

CIS centre for intelligent sensing

[BMVC'12,ECCV'12]

Queen Mary University of London


Cam 5
Cam 6
Cam 7
Cam 8
Cam 9
Cam 10
Cam 11
Cam 12
Cam 13
Cam 14
Cam 15
Cam 16
24H

More on attributes: Re-identification

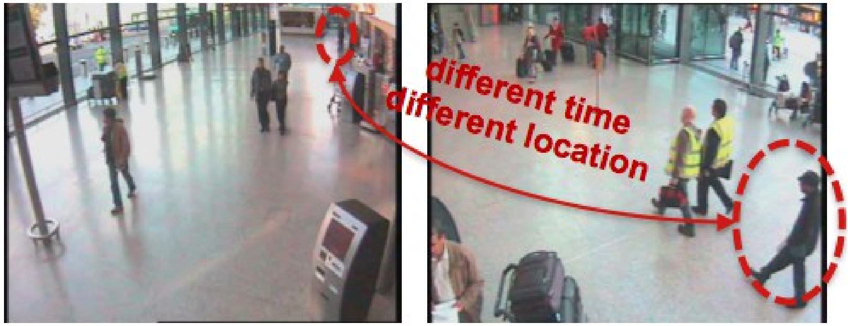
Aim: Re-identify across time, space and view

Solution?

- ~~Learn a recognizer~~

Challenge:

- Statistical Insignificance
- (One-shot learning)
- Huge intra-class variability



CIS centre for intelligent sensing

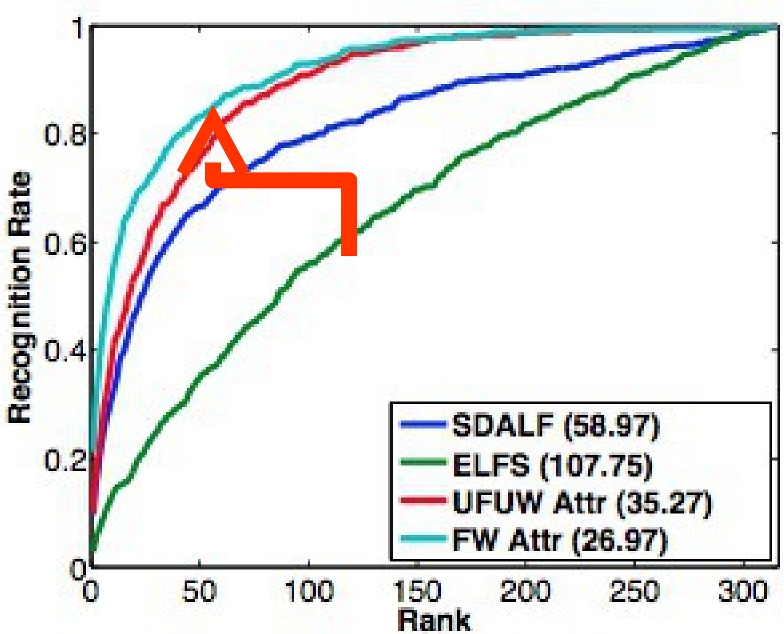
[BMVC'12,ECCV'12]

Queen Mary University of London

More on attributes: Re-identification

Aim: Re-identify across time, space and view

Solution: Attribute Transfer



- Target Person
  - ✔ Hat
  - ✗ Jeans
  - ✔ Male
  - ✔ Coat
  - ✗ Skirt
  - ✗ Tie
  - ✗ Sandals
  - ✗ Shorts
- Leverage lifetime of (attribute) experience

[BMVC’12,ECCV’12]

Queen Mary University of London

Sparse-data and re-identification

What if we have a lot of data but not much supervision?

- Active Learning

Outline

Intelligent Sensing

- Data Fusion
- Data and annotation efficiency
  - Unsupervised
  - Weakly-supervised
  - Semi-supervised
  - Multi-label/multi-instance
  - Zero-shot learning
- **Observation efficiency**
  - Active learning

Active Learning

Aim: Make query selection optimal

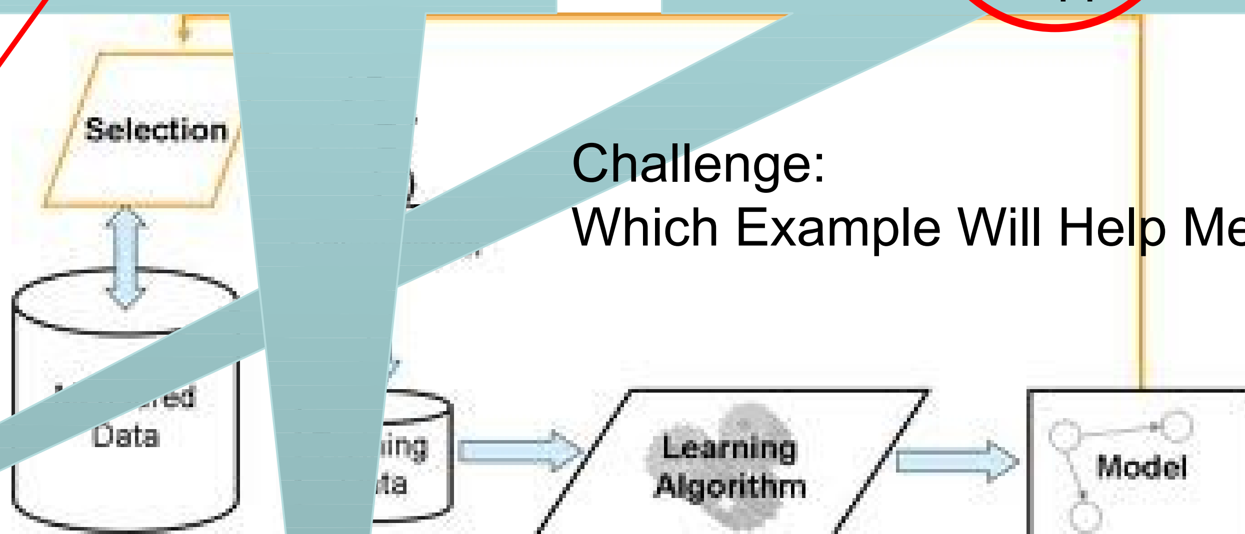
➔Minimize human annotation effort for given outcome

Cat

Dog

??

Challenge:
Which Example Will Help Me The Most?



Active Learning & Discovery

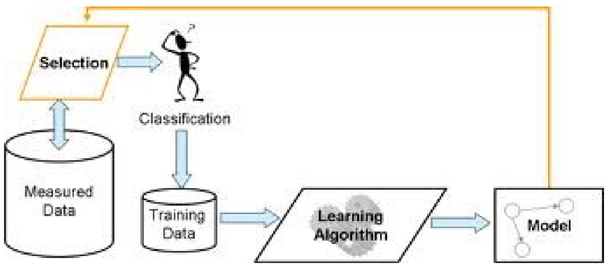
Aim: Make query selection optimal

➔Minimize human annotation effort for given outcome

Challenge:

How to do in a new domain with unknown class space?

Tractability?



[CVPR’12,ECCV’12]

Active Learning & Discovery

Solution:

- Bayesian non-parametrics
- Incremental Computation

Density Estimation
Bayes Inference

[CVPR'12,ECCV'12]

Active Learning & Discovery

Take Homes

- Be aware of the underlying ML of your intelligent sensing problem
  - Then you can find good techniques
- Separation of:
  - features, objectives, model/representation, optimizers
- Think of your data and annotation constraints
  - Can they be reduced to make your model more useful?
  - How does it depend on the strength of your annotation?
  - Could your model do better with more data (but same annotation?)
  - Would finding the right annotations help?