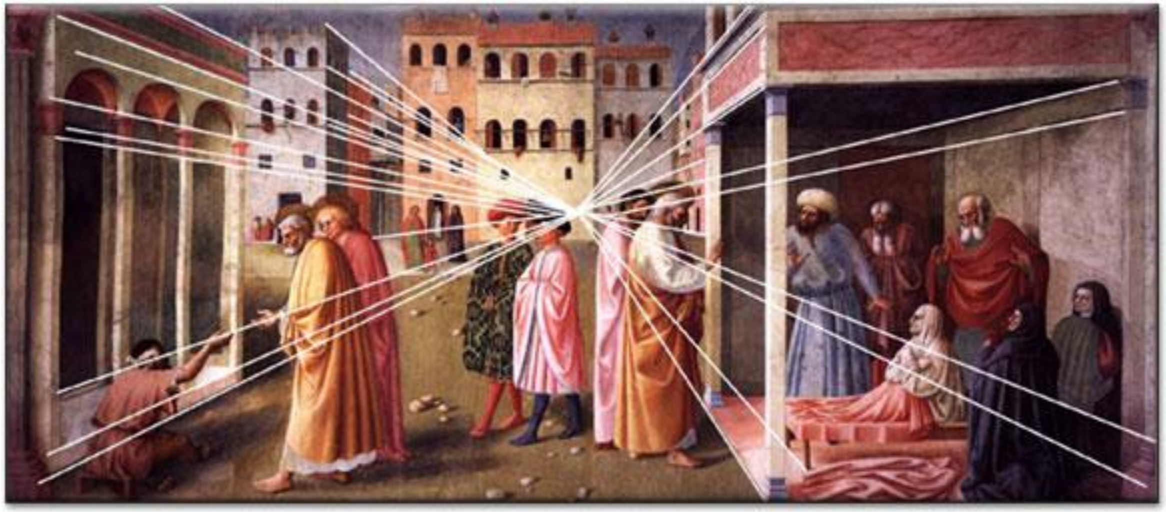
Microsoft Research

GRAPHICS & MEDIA LAB
VISION GROUP

Graphical models in image understanding



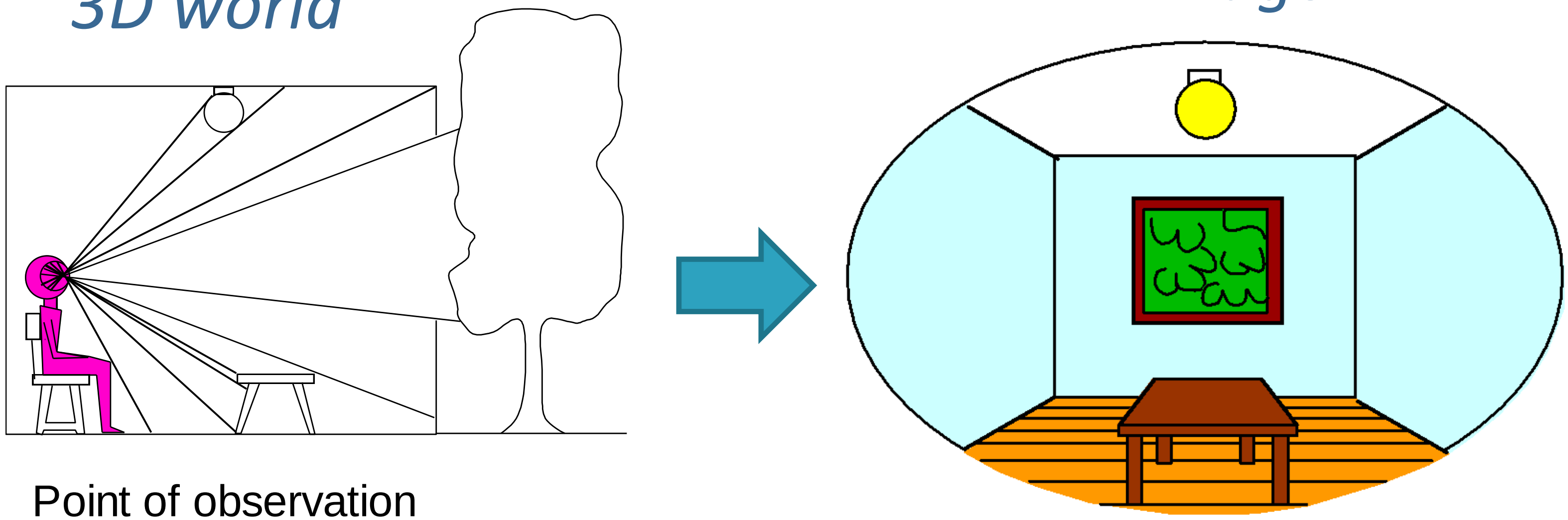
Olga Barinova

Lomonosov Moscow State University

Thanks to Anton Konushin for some slides

Why is image understanding difficult?

- Dimensionality reduction



What do we lose in perspective projection?

- Angles
- Distances and lengths

Figures © Stephen E. Palmer, 2002

Why is image understanding difficult?

- Dimensionality reduction

  The same 2D image can correspond to different 3D scenes

Image source: F. Durand

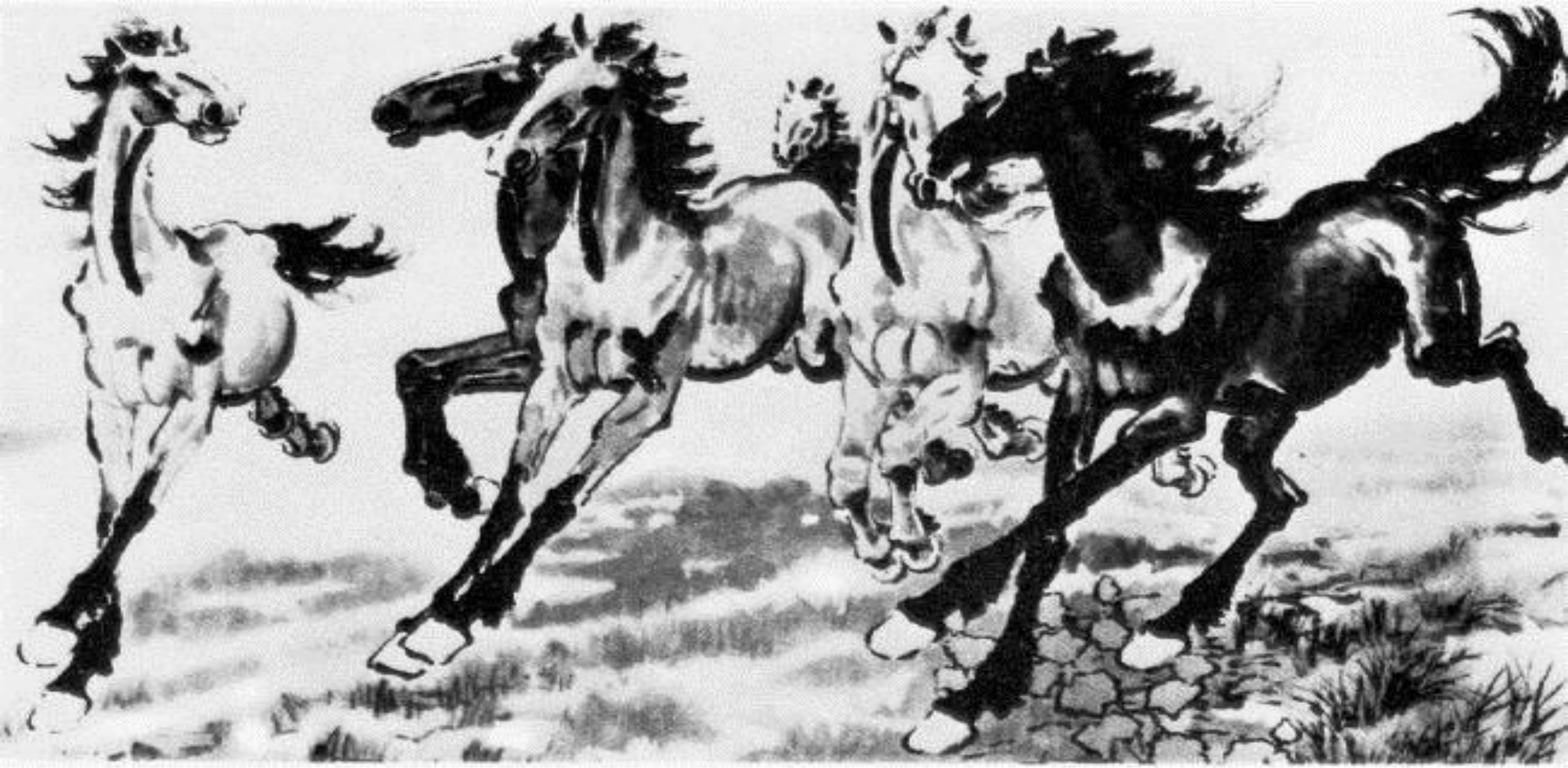
Microsoft Research

Why is image understanding difficult?

•Variability: interclass variability of appearance

Why is image understanding difficult?

- Variability: deformations and occlusions



Xu, Beihong 1943

Slide credit: Fei-Fei, Fergus & Torralba

Microsoft Research

Why is image understanding difficult?

•Variability: different viewpoints

Michelangelo 1475-1564

slide credit: Fei-Fei, Fergus & Torralba

Why is image understanding difficult?

- Variability: different lighting

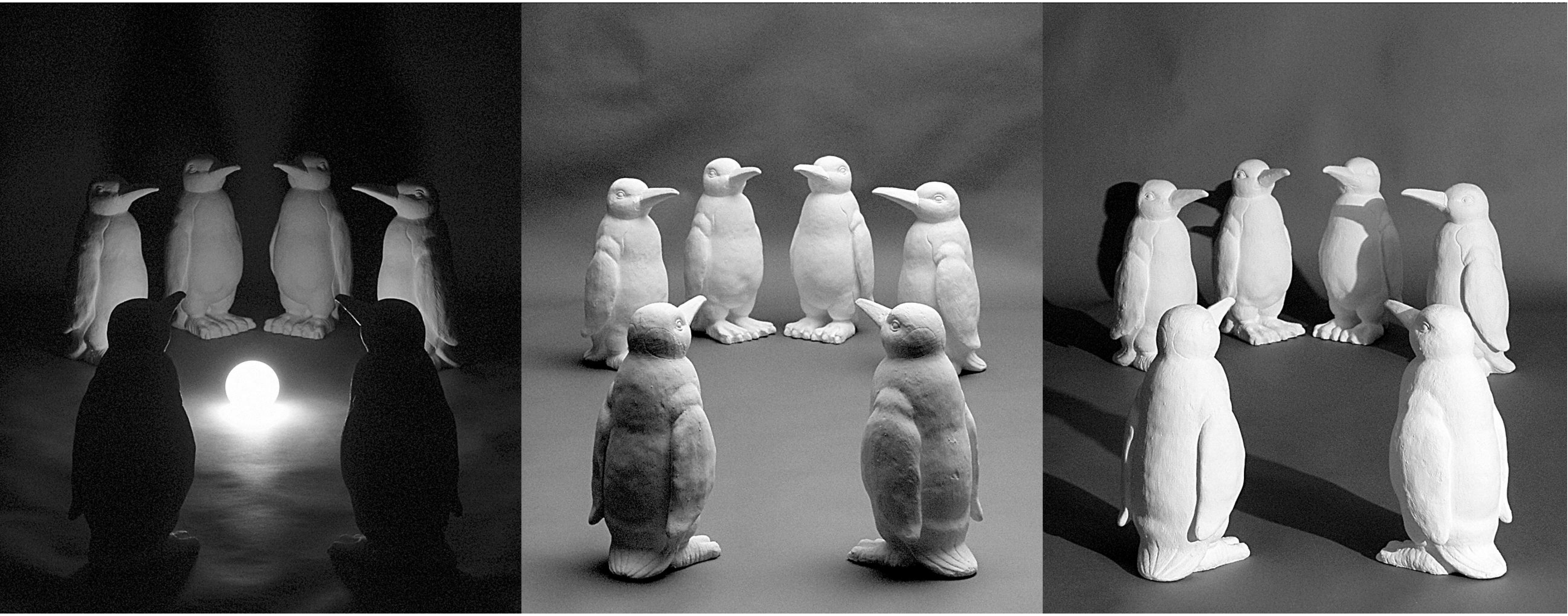


image credit: J. Koenderink

Microsoft Research

GRAPHICS & MEDIA LAB VISION GROUP

Reducing variability by using local cues

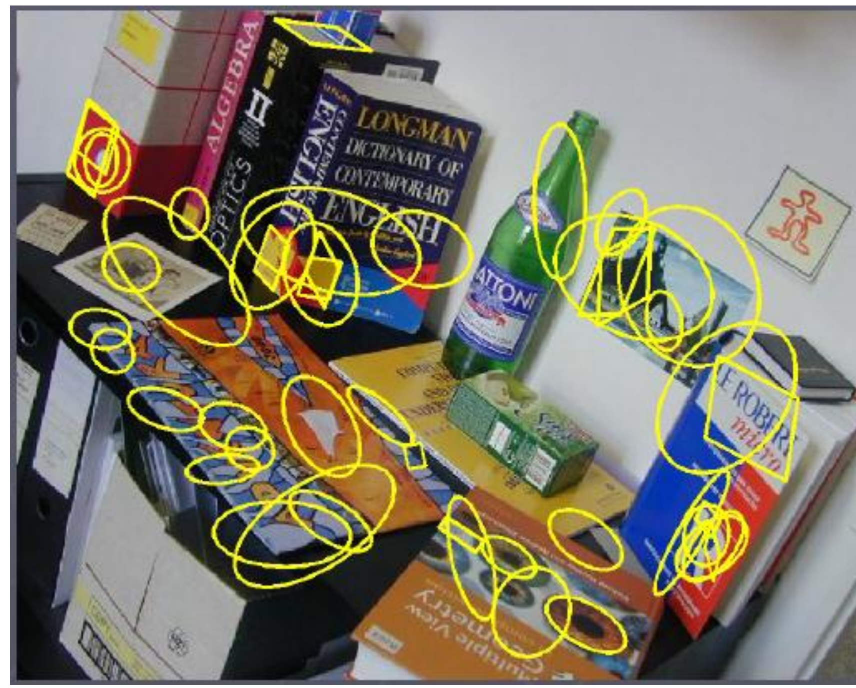
- Motivation: stitching panoramas
  - Find distinctive points
  - Find an alignment that matches these points

Microsoft Research

GRAPHICS & MEDIA LAB VISION GROUP

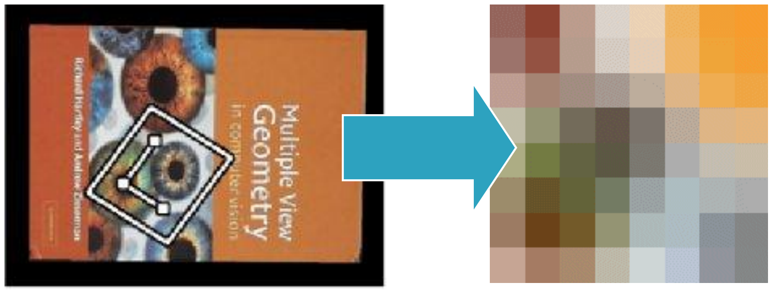
Reducing variability by using local cues

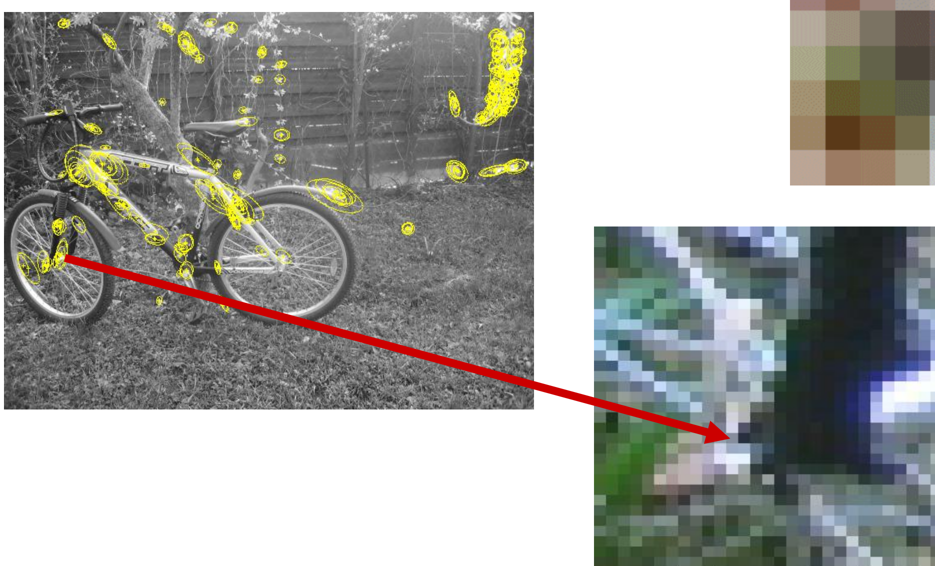
- Motivation: stereo matching



Reducing variability by using local cues

- Motivation: image retrieval object detection

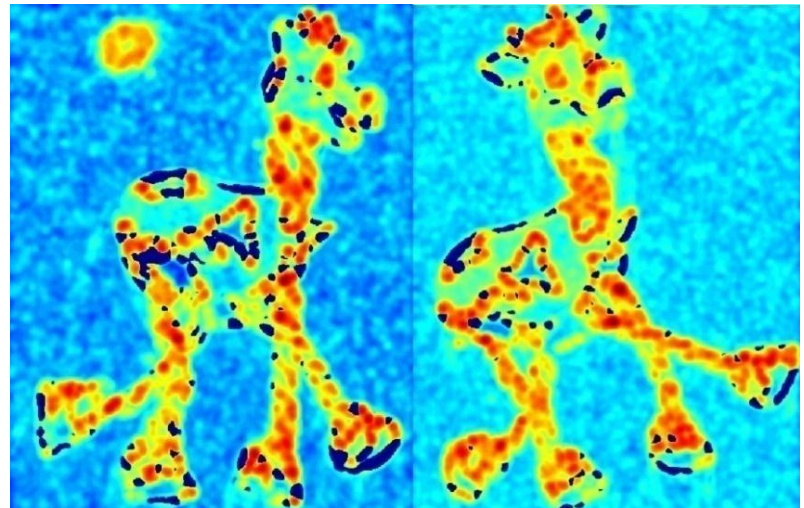


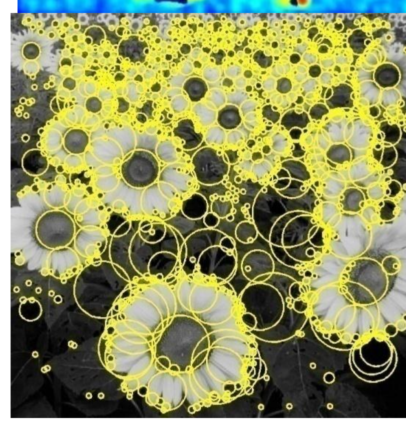


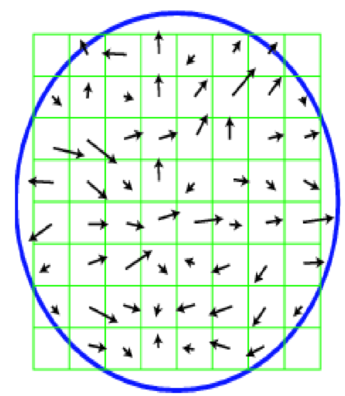
Slide creit: S. Lazebnik

Learning local models from local cues

- Local features and descriptors
  - Feature detectors
    - Harris-Laplace
    - LoG
    - DoG
    - Dense sampling
  - Descriptors
    - SIFT
    - Shape context
    - HOG
    - Pixel comparison

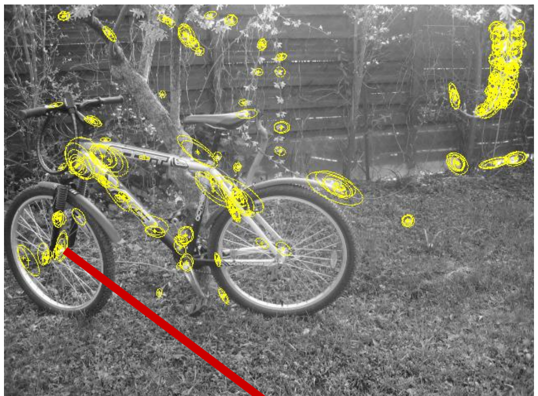


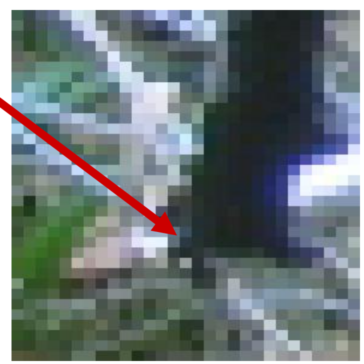


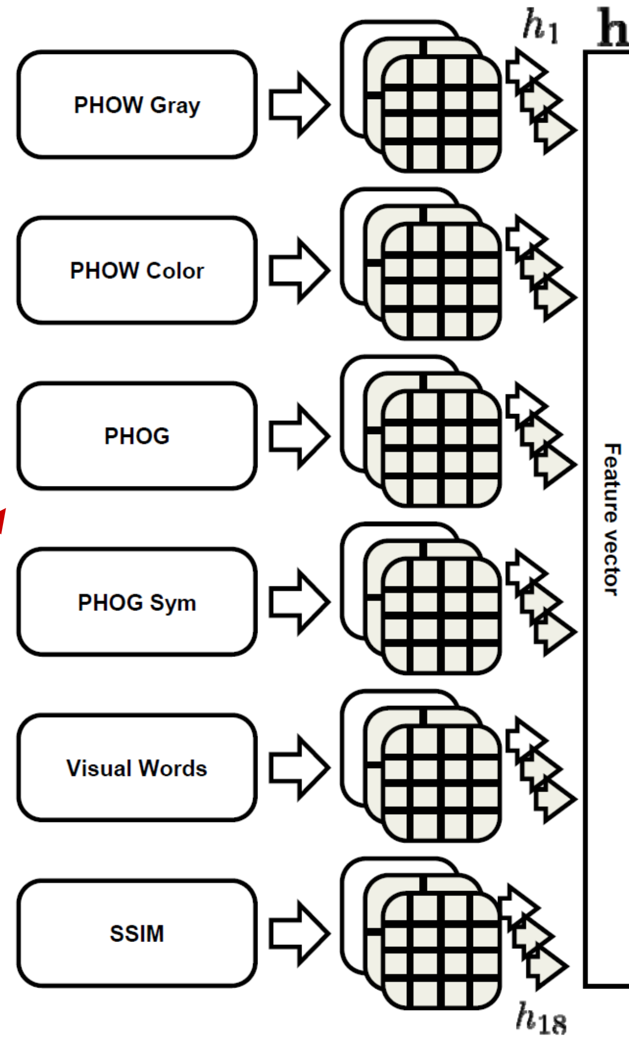


Learning local models from local cues

- Combining different descriptors

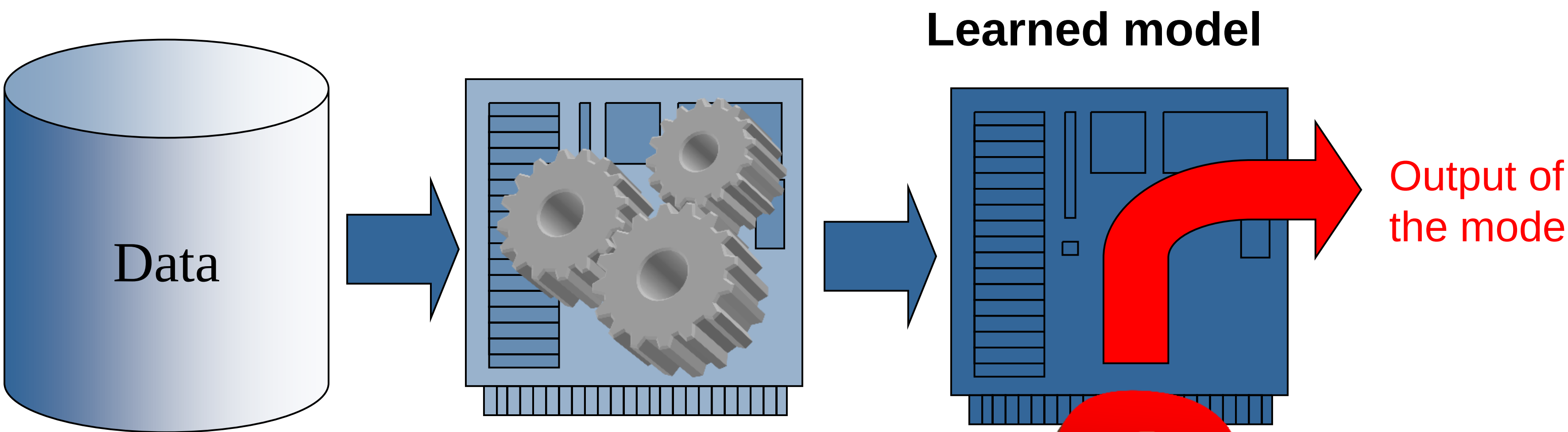






Learning local models from local cues

- Learning models from the data



Learning methods:

- SVM
- Boosting
- Random Forests
- ……

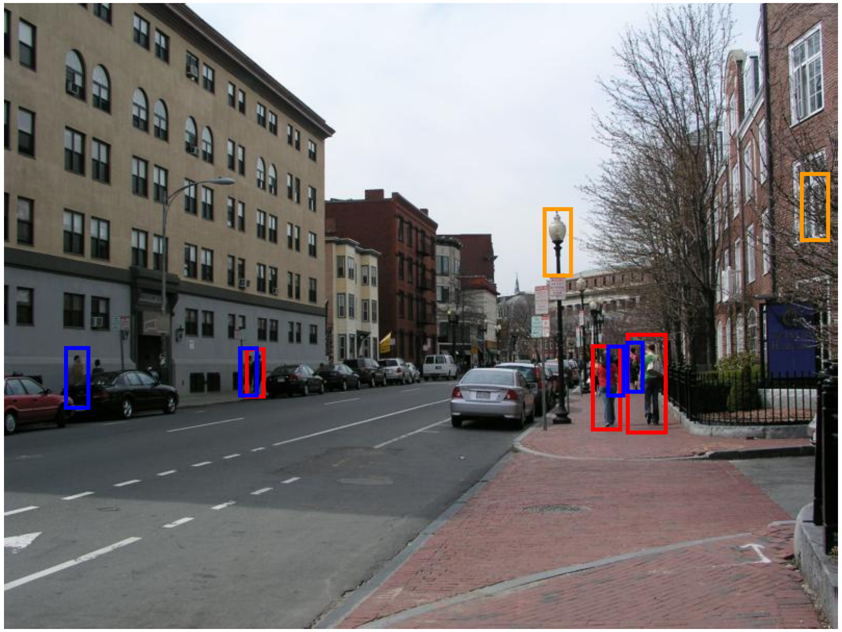
Microsoft Research

Learning local models from local cues

- Example: object detection using sliding window

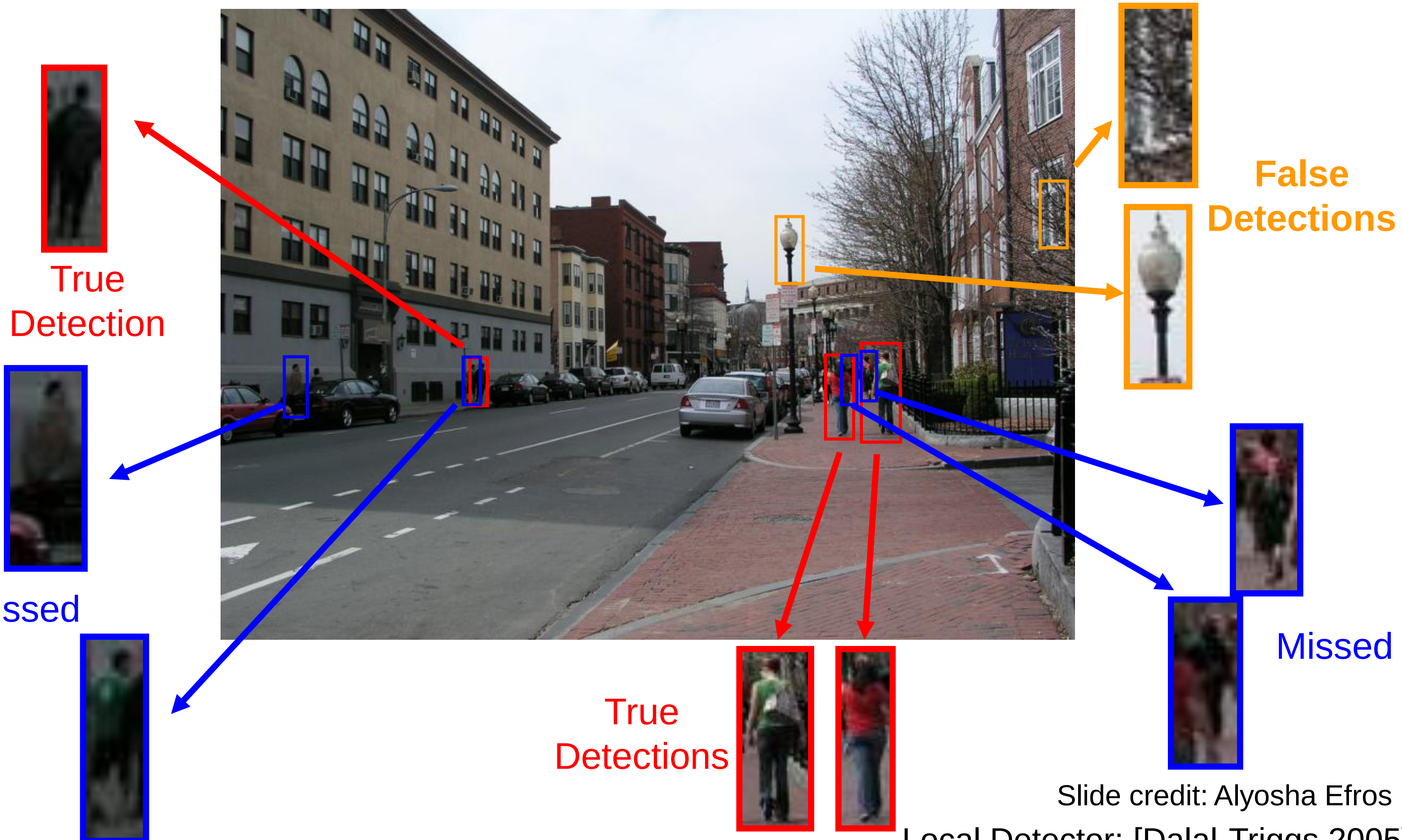
  - 'Local' has been the dominant paradigm in computer vision till the 2000s

  - Works notoriously well for detection of rigid objects, e. g. faces
    [Viola, Jones, 2001],
    [Dalal, Triggs, 2005]



Learning local models from local cues

- Let's have a closer look at the results



Slide credit: Alyosha Efros

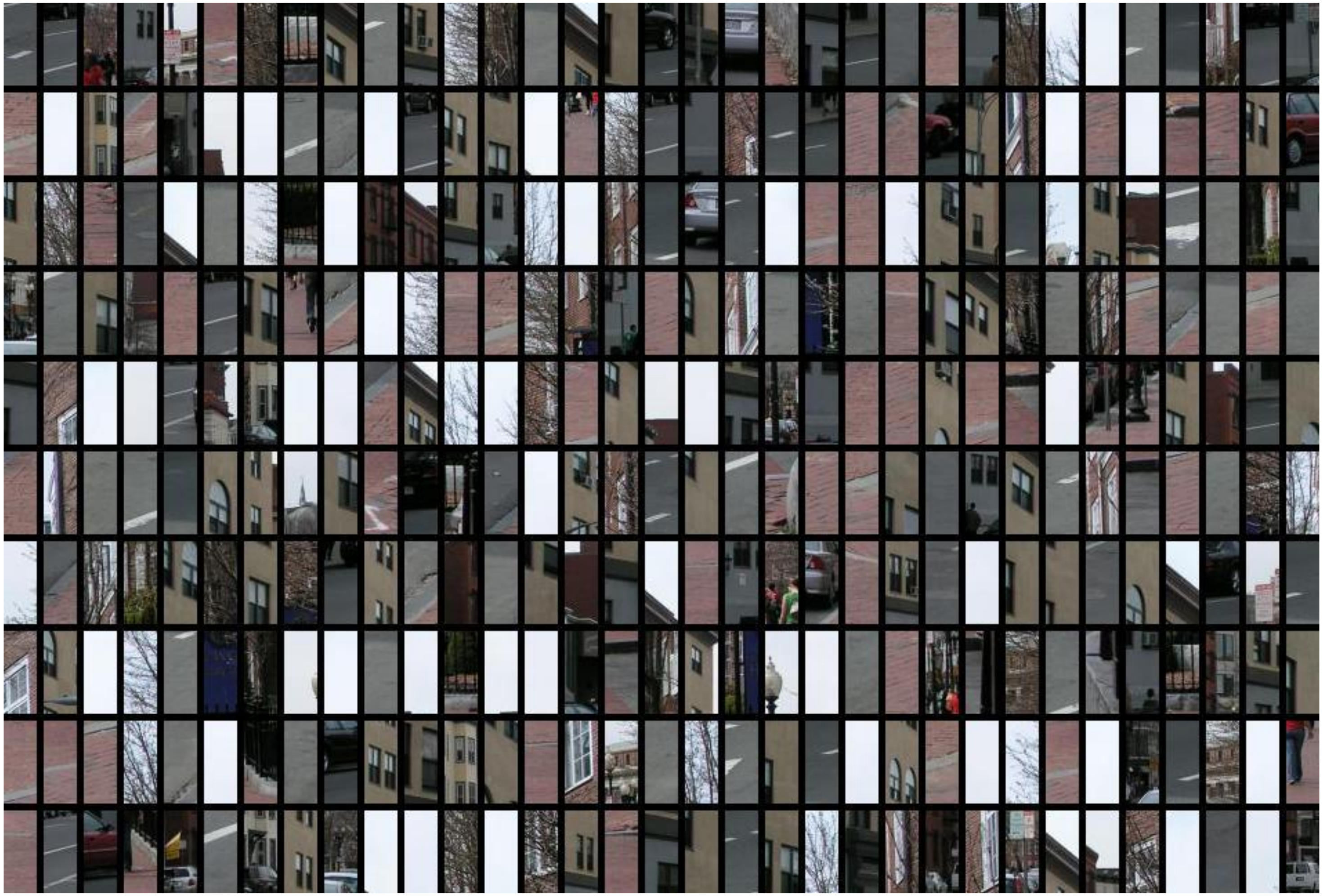
Local Detector: [Dalal-Triggs 2005]

Microsoft Research

GRAPHICS & MEDIA LAB VISION GROUP

Learning local models from local cues

- What the detector sees



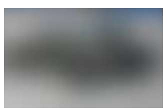
Slide credit: Alyosha Efros

Microsoft Research

GRAPHICS & MEDIA LAB VISION GROUP

Learning local models from local cues

- Local ambiguity



Slide credit: Fei-Fei, Fergus & Torralba

Microsoft Research

Learning local models from local cues

- The role of context

Slide credit: Fei-Fei, Fergus & Torralba

Microsoft Research

GRAPHICS & MEDIA LAB VISION GROUP

Learning local models from local cues

- The role of context

Slide credit: Fei-Fei, Fergus & Torralba

Constraints of the world

- The world is structured, not everything is possible

Local cues
- Similar appearance of similar objects

Global constraints
- Limited number of allowed deformations of the objects in 3d
- Depth ordering and occlusions
- Rules of perspective projection
- ….

Chaotic world

Structured world

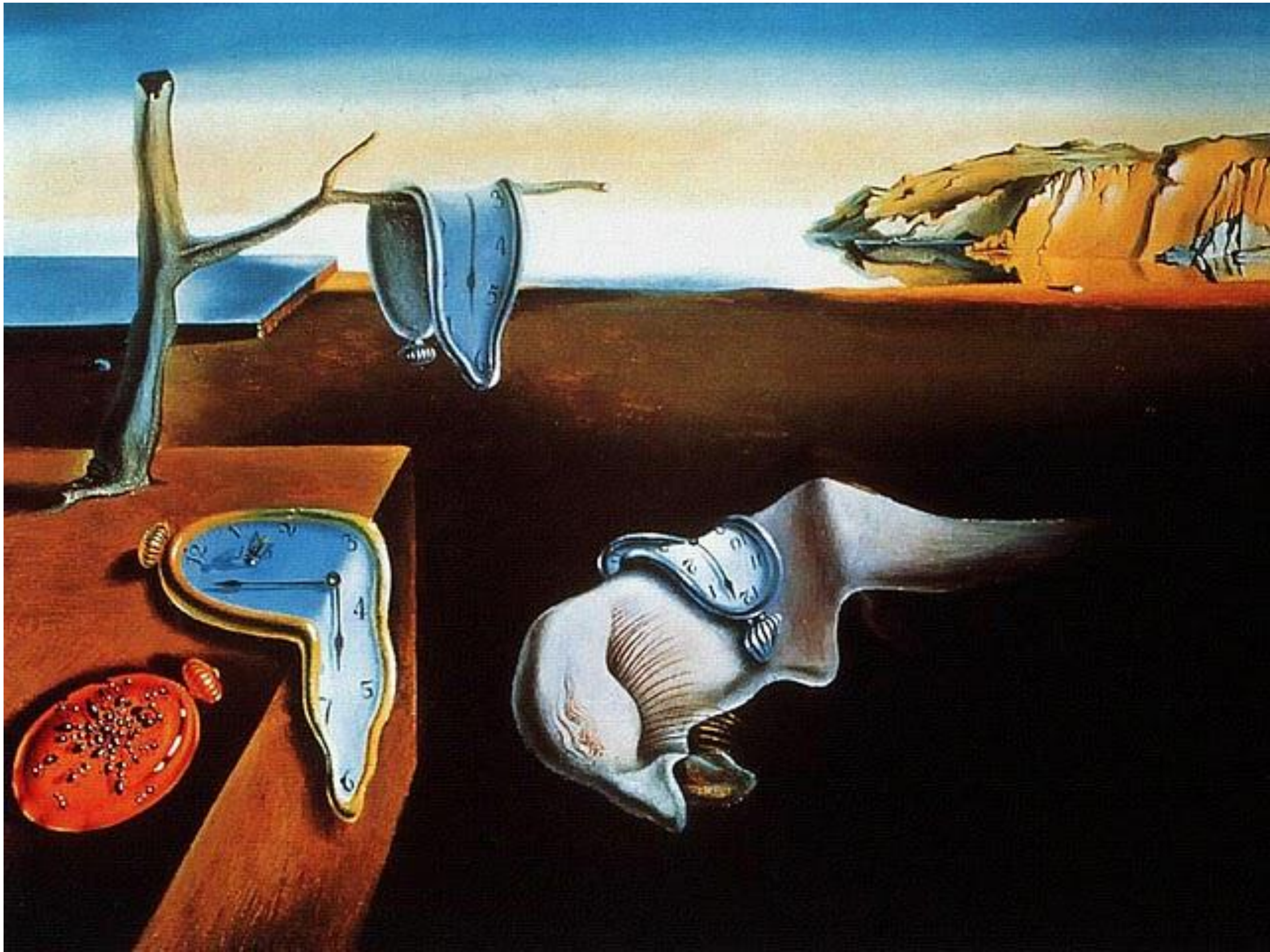
Microsoft Research

GRAPHICS & MEDIA LAB VISION GROUP

Constraints of the world

- Limited set of allowed deformations for the objects

Dali, 1931

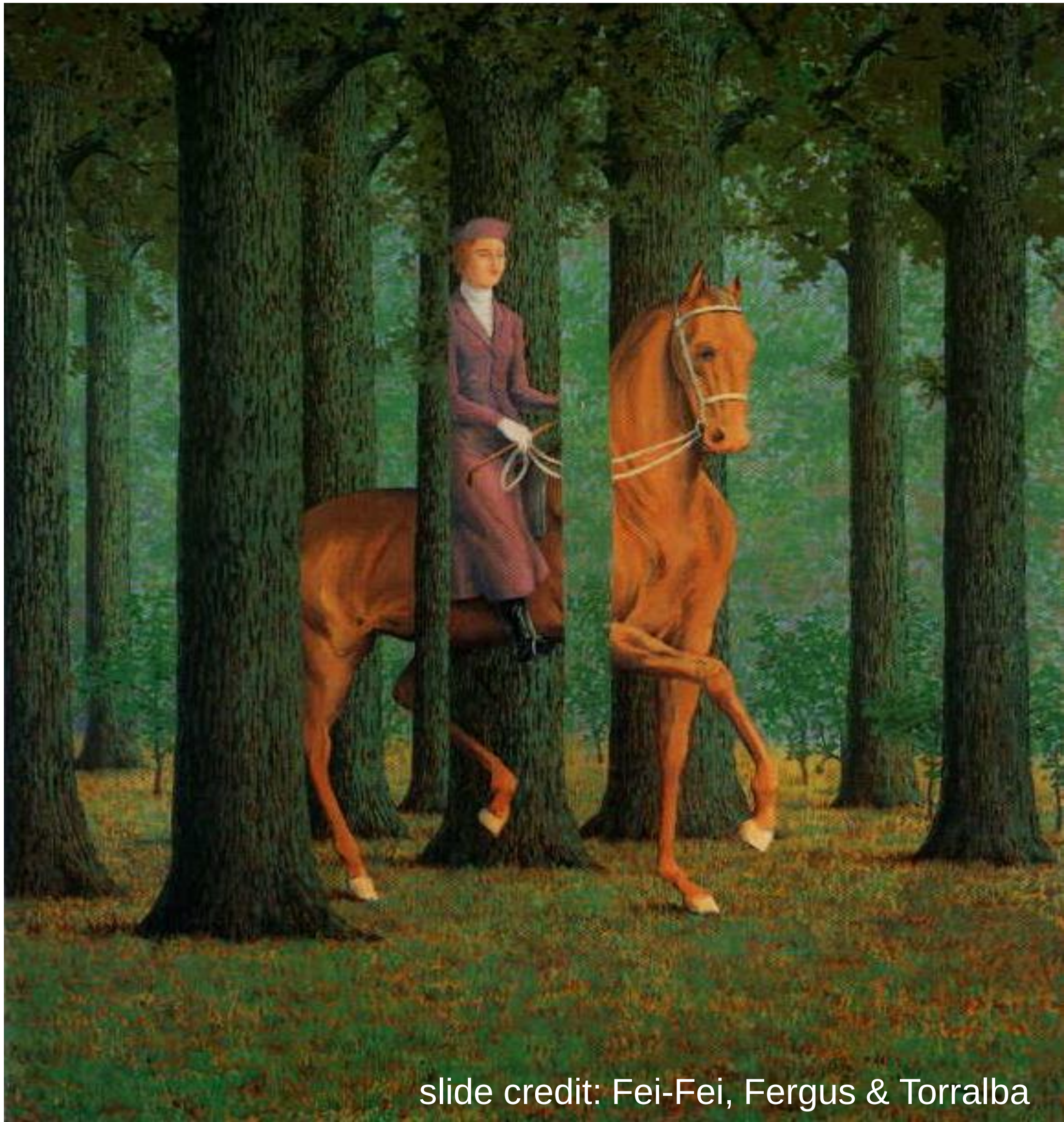


Microsoft Research

Constraints of the world

- Occlusions

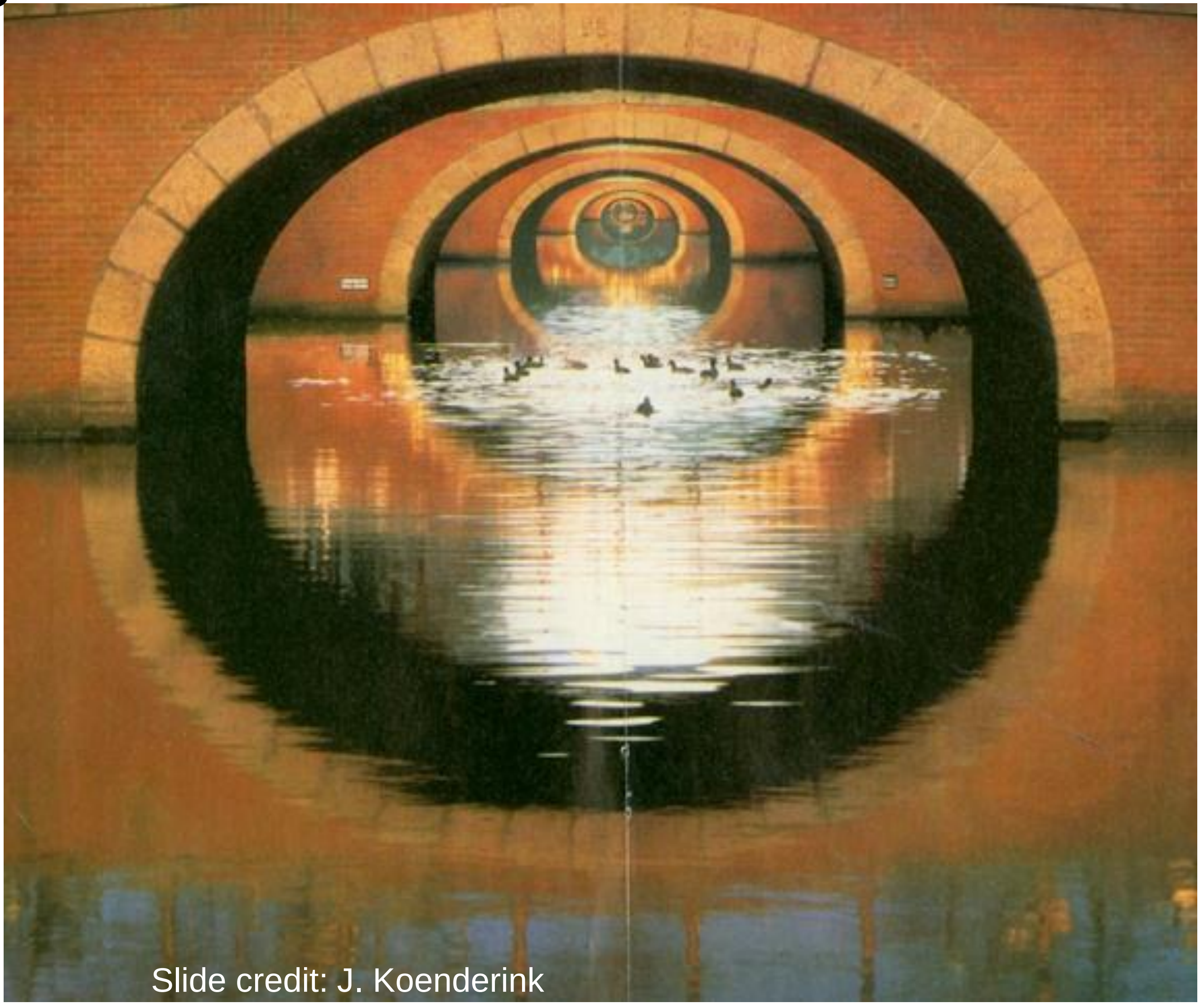
Magritte, 1957



slide credit: Fei-Fei, Fergus & Torralba

Constraints of the world

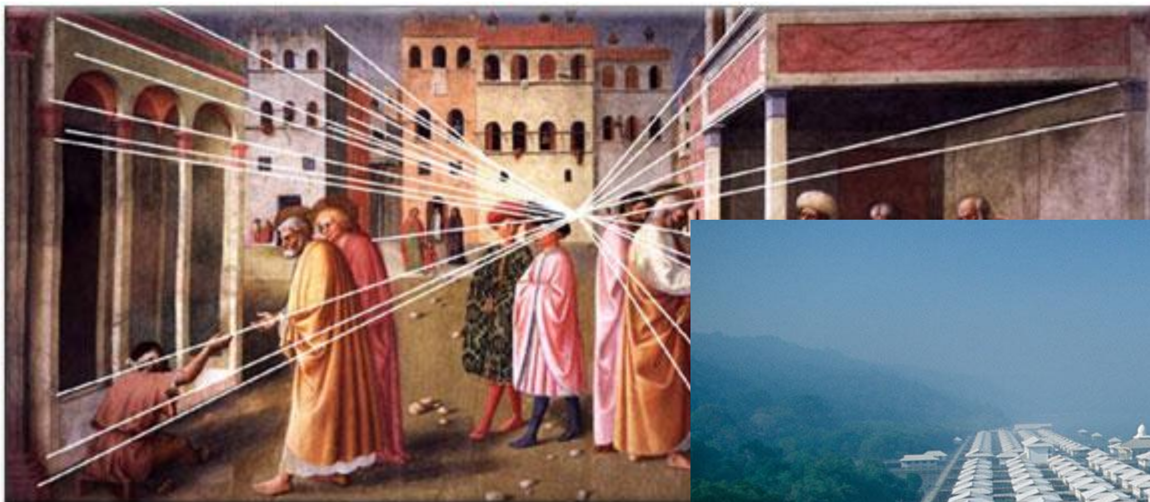
- Depth ordering



Slide credit: J. Koenderink

Constraints of the world

- Rules of perspective geometry

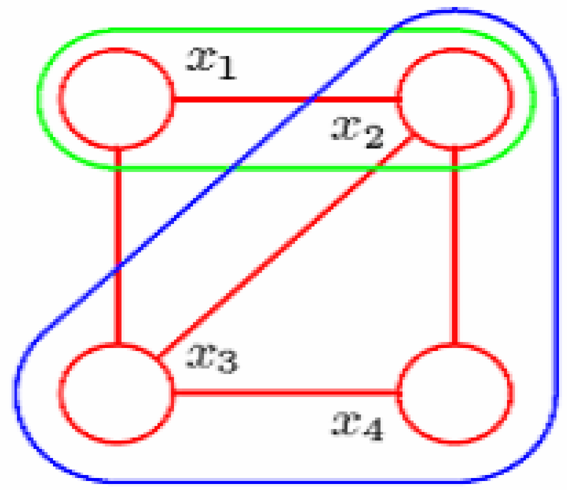


Expressing constraints with graphical models

- Outline of the talk
  - The idea of graphical models
  - Examples:
    - Limiting the set of allowed deformations
    - Occlusion constraint
    - Depth ordering constraint
    - Modeling the rules of perspective geometry

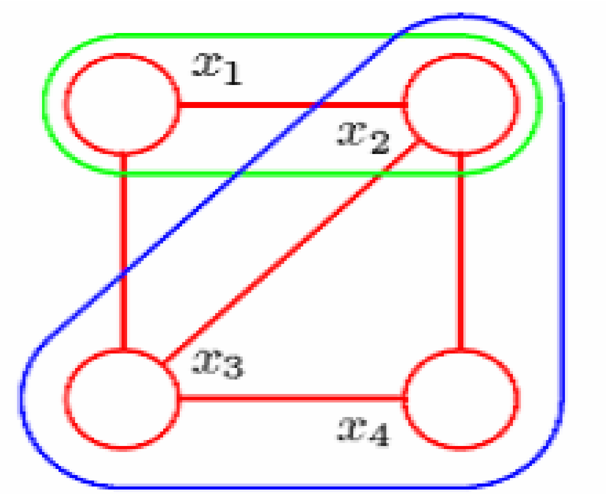
Expressing constraints with graphical models

- Outline of the talk
  - **The idea of graphical models**
  - Examples:
    - Limiting the set of allowed deformations
    - Occlusion constraint
    - Depth ordering constraint
    - Modeling the rules of perspective geometry



Expressing constraints with graphical models

- Graphical models
  - Graphical representation of probability distributions
  - Graph-based algorithms for calculation and computation
  - Capture both local cues and global constraints by modeling dependencies between random variables



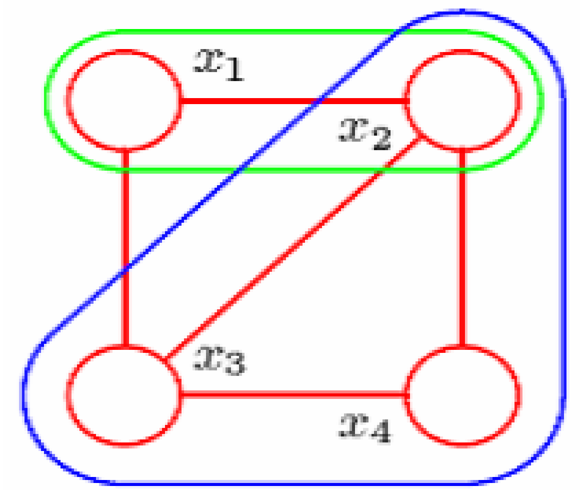
Picture credit: C. Bishop

Microsoft Research

Graphical models

- Graph representation

  - Each node corresponds to a **random variable**
  - **Dependent variables** are connected with edges



  - **Clique** - fully connected set of nodes in the graph
  - **Maximal clique** - a clique that is not a subset of any other cliques

$$p(x_1, x_4 | x_2, x_3) =$$
$$= p(x_1 | x_2, x_3)\, p(x_4 | x_2, x_3)$$

Picture credit: C. Bishop

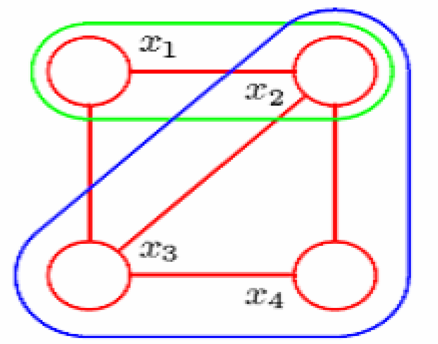
Graphical models

- Joint distribution and potentials

Joint distribution of all random variables can be written as a product of nonnegative **potentials** defined on maximal cliques:

$$p(X) = \frac{1}{Z}\prod_{C}\psi_C(X_C) \qquad Z = \sum_X \prod_C \psi_C(X_C), \qquad \psi_C(X_C) \geq 0$$

$$p(X) = \frac{1}{Z}\psi_1(x_1, x_2, x_3)\psi_2(x_2, x_3, x_4)$$



Microsoft Research

Graphical models

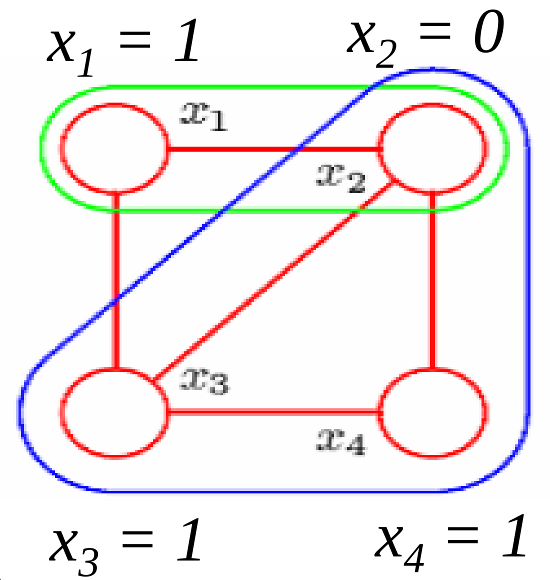
- MAP-inference and energy function

Maximum a-posteriori (MAP) inference - find the values of all variables in the graphical model that maximize the joint probability

$$\arg\max p(X) = \arg\max \frac{1}{Z}\prod_C \psi_C(X_C) =$$

$$= \arg\max \exp\left(-\sum_C E_C(X_C)\right) =$$

$$= \arg\min \sum_C E_C(X_C)$$



Energy function: $E(X) = logP(x) = \sum_C E_C(X_C)$

MAP-inference = energy minimization

Graphical models

- Methods for MAP-inference
  - Many computationally efficient methods for inference in graphical models have been developed:
    - graph cuts
    - TRW
    - belief propagation
    - expectation propagation
    - MCMC
    - ….
  - All these methods have limitations and can be used to minimize energy functions of specific forms → the art is to find tradeof between flexibility and tractability