Local cues and global constraints in image understanding

Olga Barinova

Lomonosov Moscow State University

*Many slides adopted from the courses of Anton Konushin

6th Microsoft PhD Summer School, Cambridge, UK, 27 June – 1 July 2011.

Microsoft Research

Image understanding

- «To see means to know what is where by looking»
  - *David Marr, **Vision**, 1982*

We see

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

Computer sees

Slide credit S. Narasimhan

Image understanding

Slide credit: Fei-Fei, Fergus & Torralba

Microsoft Research

Image understanding

•Scene interpretation

- outdoors
- urban
- street
- Beijing, China

slide credit: Fei-Fei, Fergus & Torralba

Microsoft Research

Image understanding

- Object detection

slide credit: Fei-Fei, Fergus & Torralba

Microsoft Research

Image understanding

GRAPHICS & MEDIA LAB

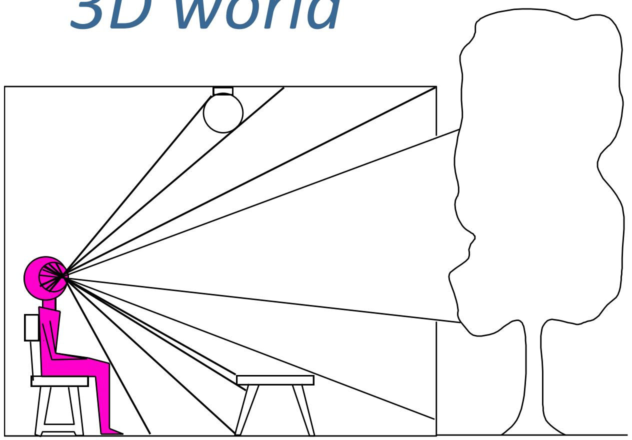
- Extracting details about the objects

slide credit: Fei-Fei, Fergus & Torralba

Why is image understanding difficult?

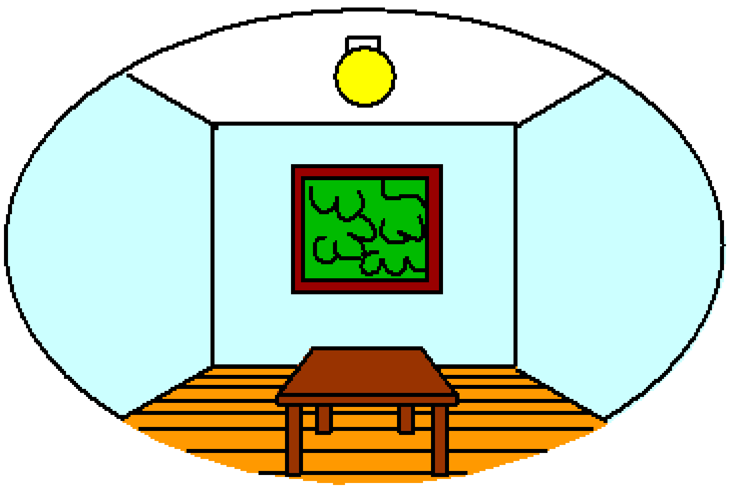
- Dimensionality reduction

3D world



Point of observation

2D image



What do we lose in perspective projection?

- Angles
- Distances and lengths

Why is image understanding difficult?

- Variability: interclass variability of appearance

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

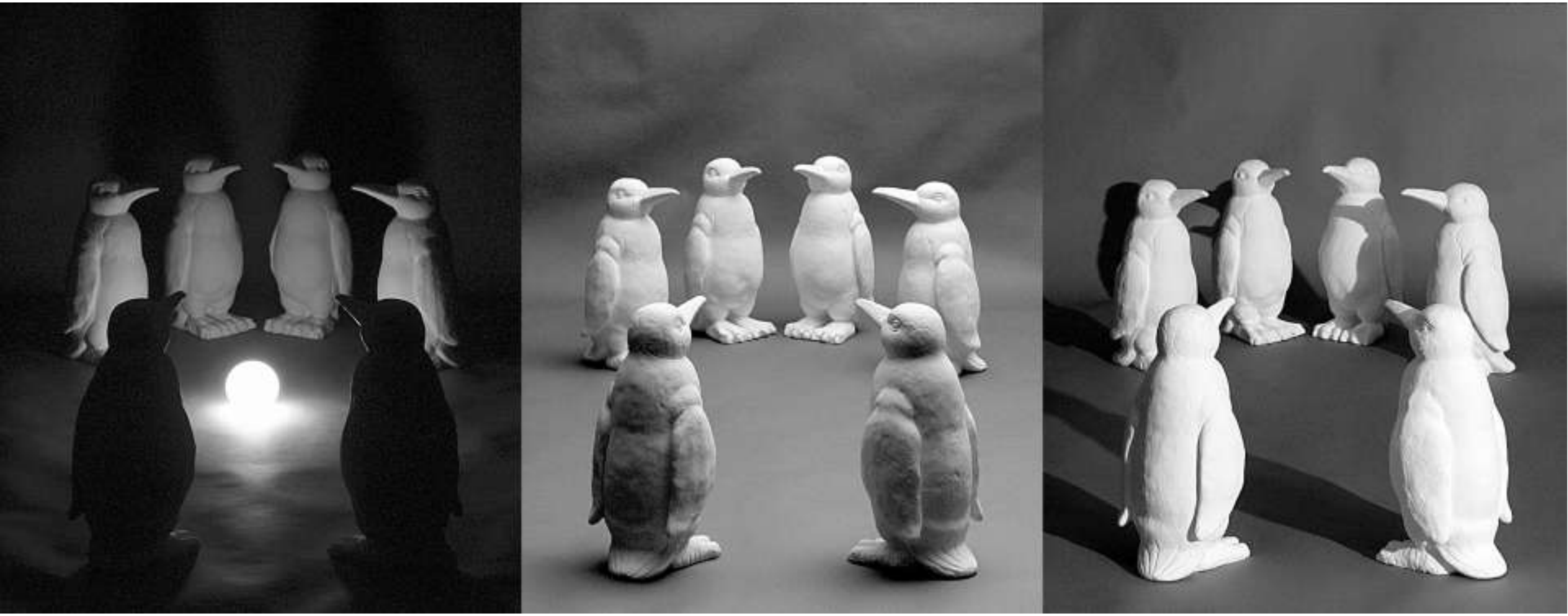
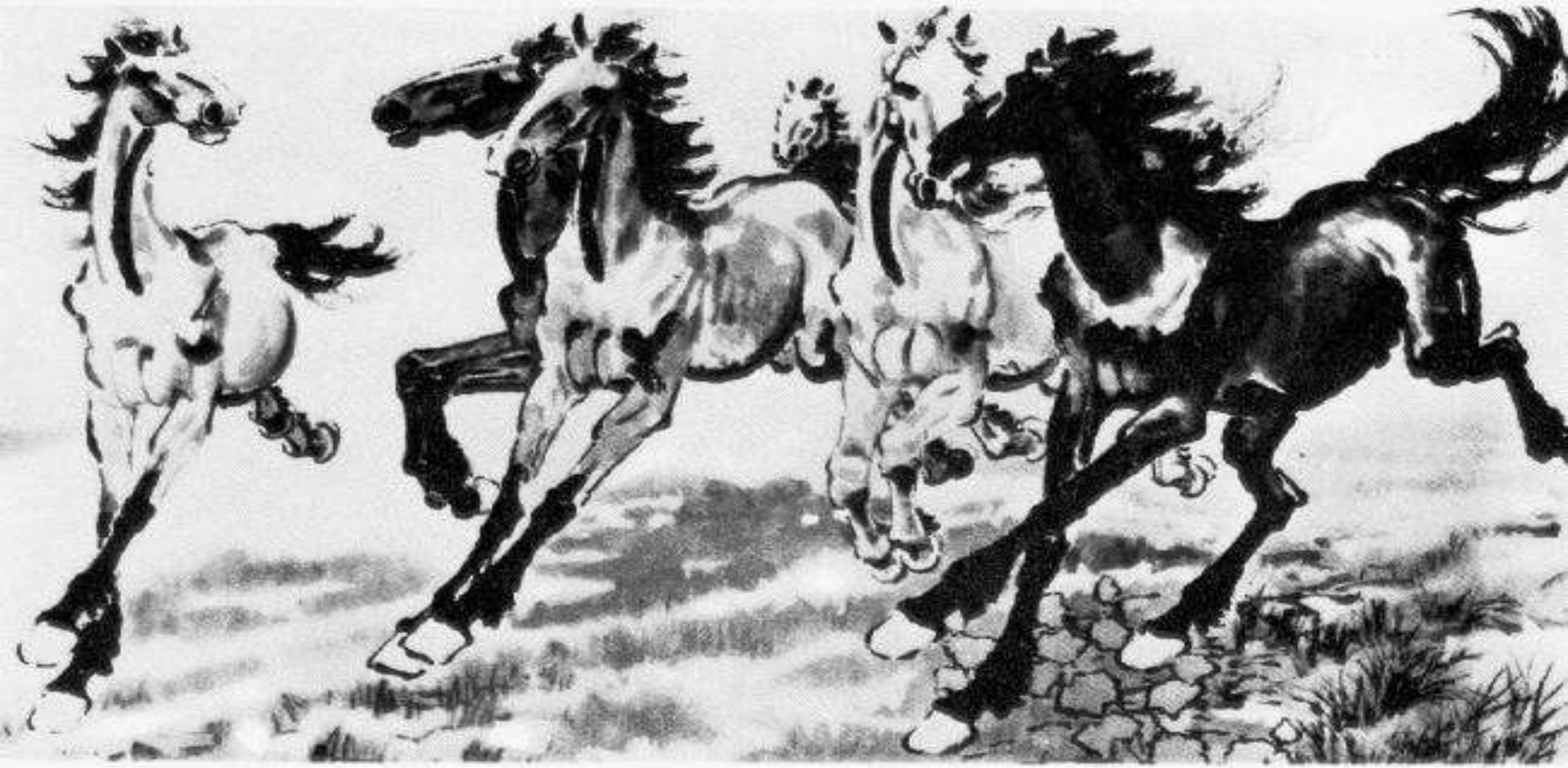


image credit: J. Koenderink

Why is image understanding difficult?

•Variability: deformations and occlusions



Xu, Beihong 1943

Slide credit: Fei-Fei, Fergus & Torralba

Reducing variability by using local cues

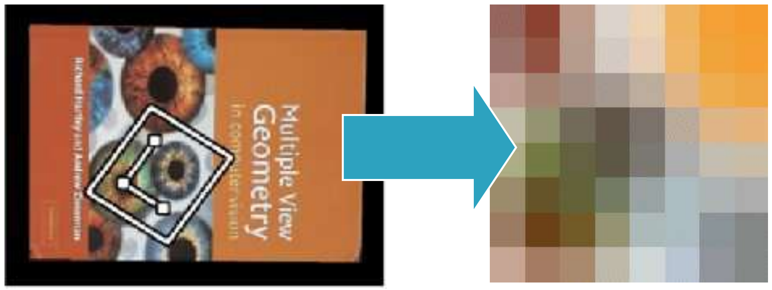
- Motivation: stitching panoramas
  - Find distinctive points
  - Find an alignment that matches these points

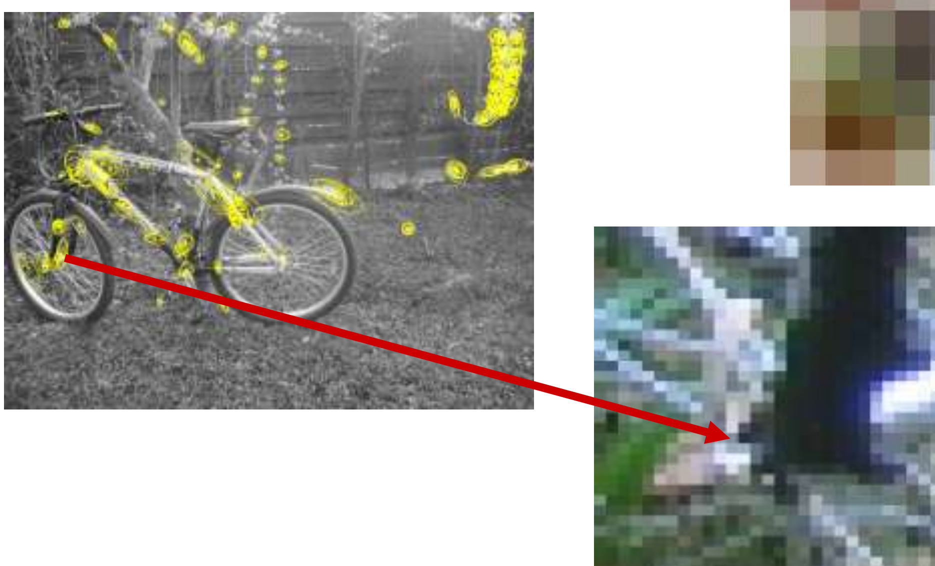
Reducing variability by using local cues

- Motivation: stereo matching

Reducing variability by using local cues

• Motivation: image retrieval object detection

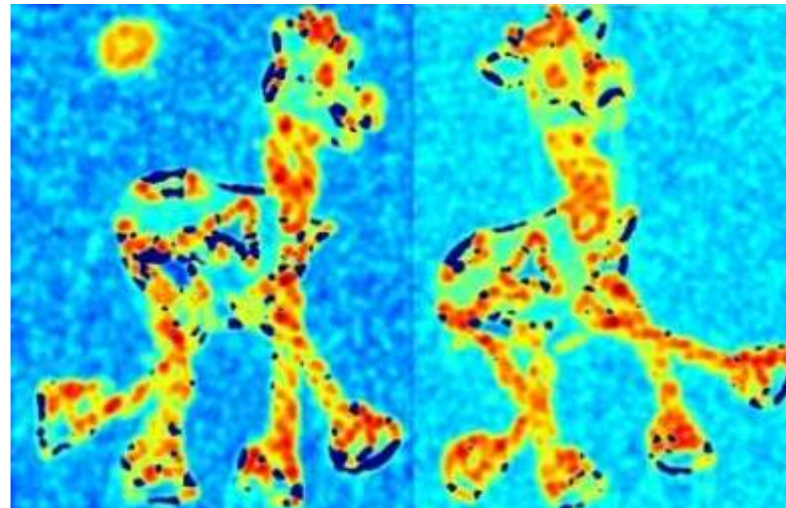


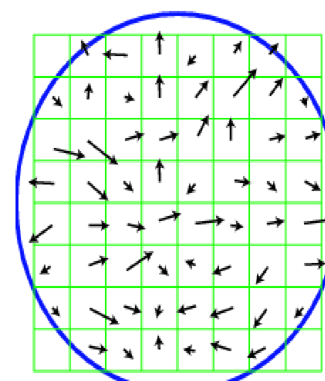


Slide creit: S. Lazebnik

Learning local models from local cues

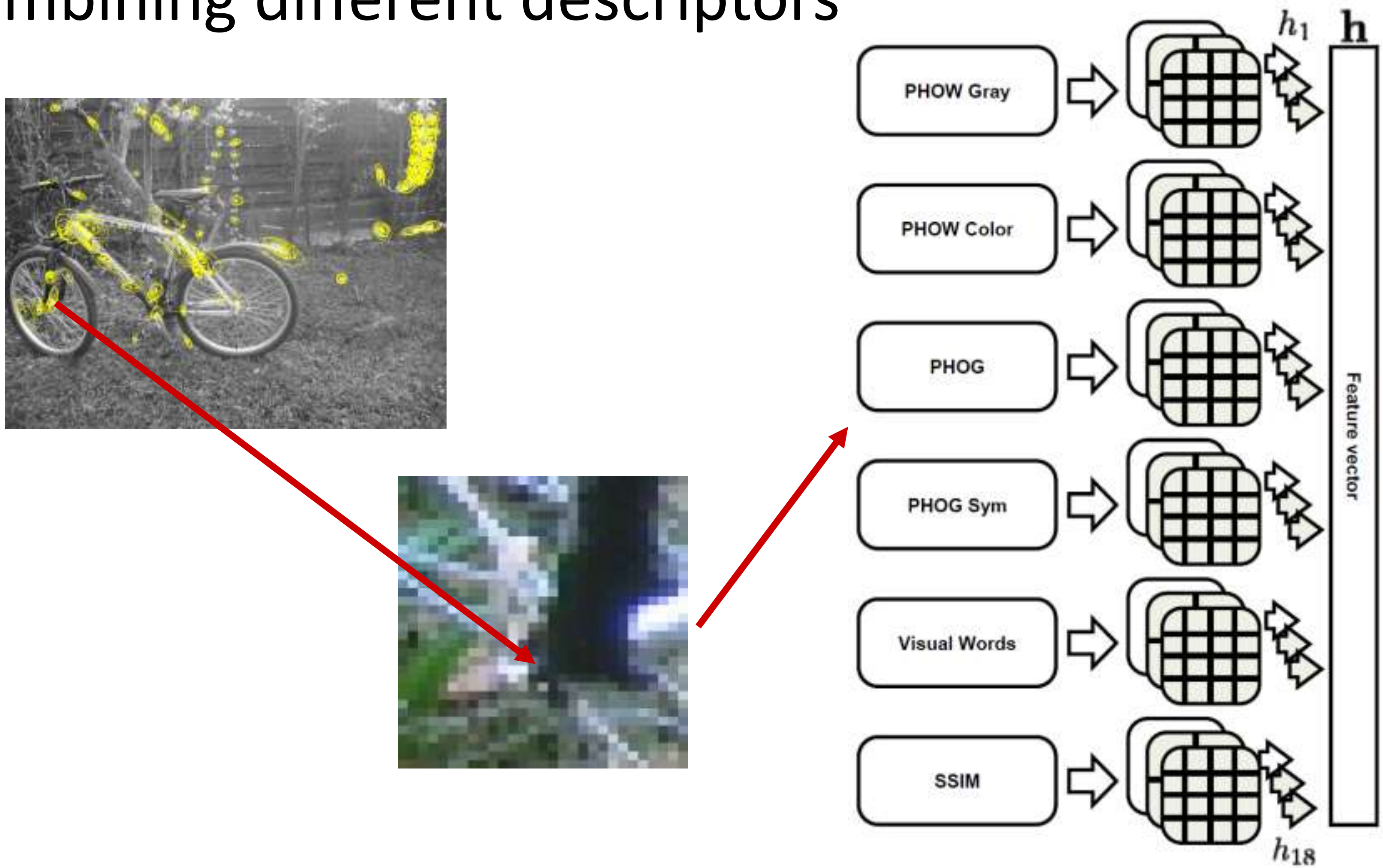
- Local features and descriptors
  - Feature detectors
    - Harris-Laplace
    - LoG
    - DoG
    - Dense sampling
  - Descriptors
    - SIFT
    - Shape context
    - HOG
    - Pixel comparison





Learning local models from local cues

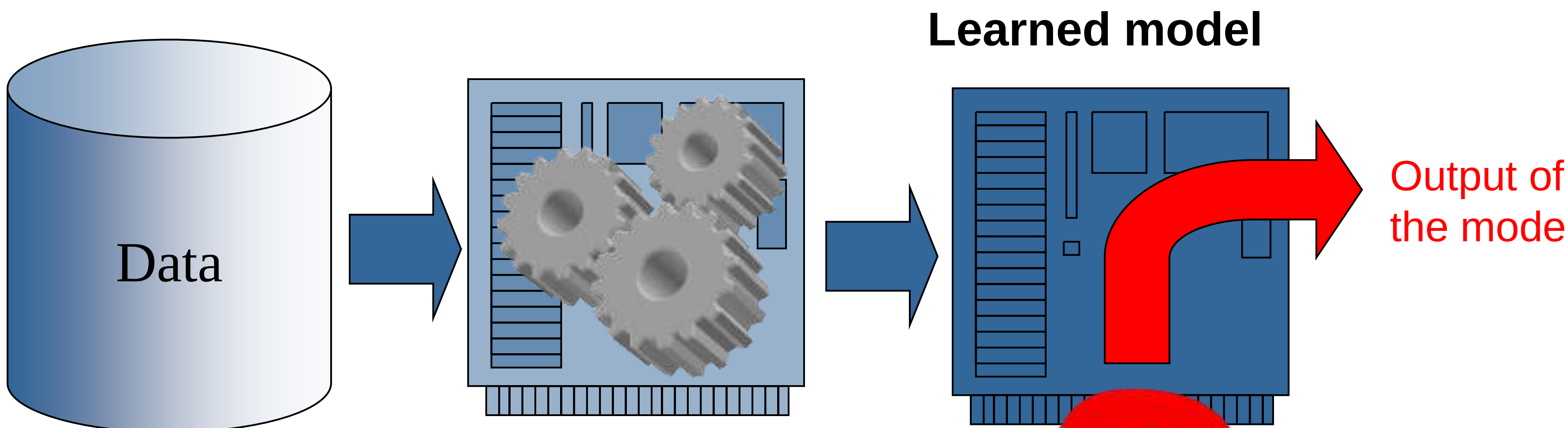
- Combining different descriptors

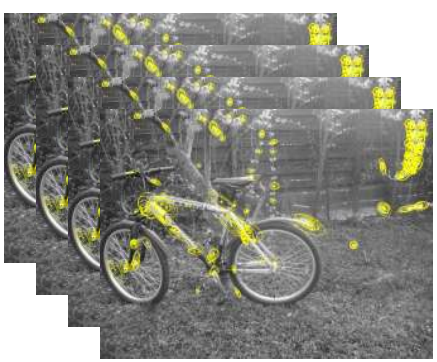


Microsoft Research

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

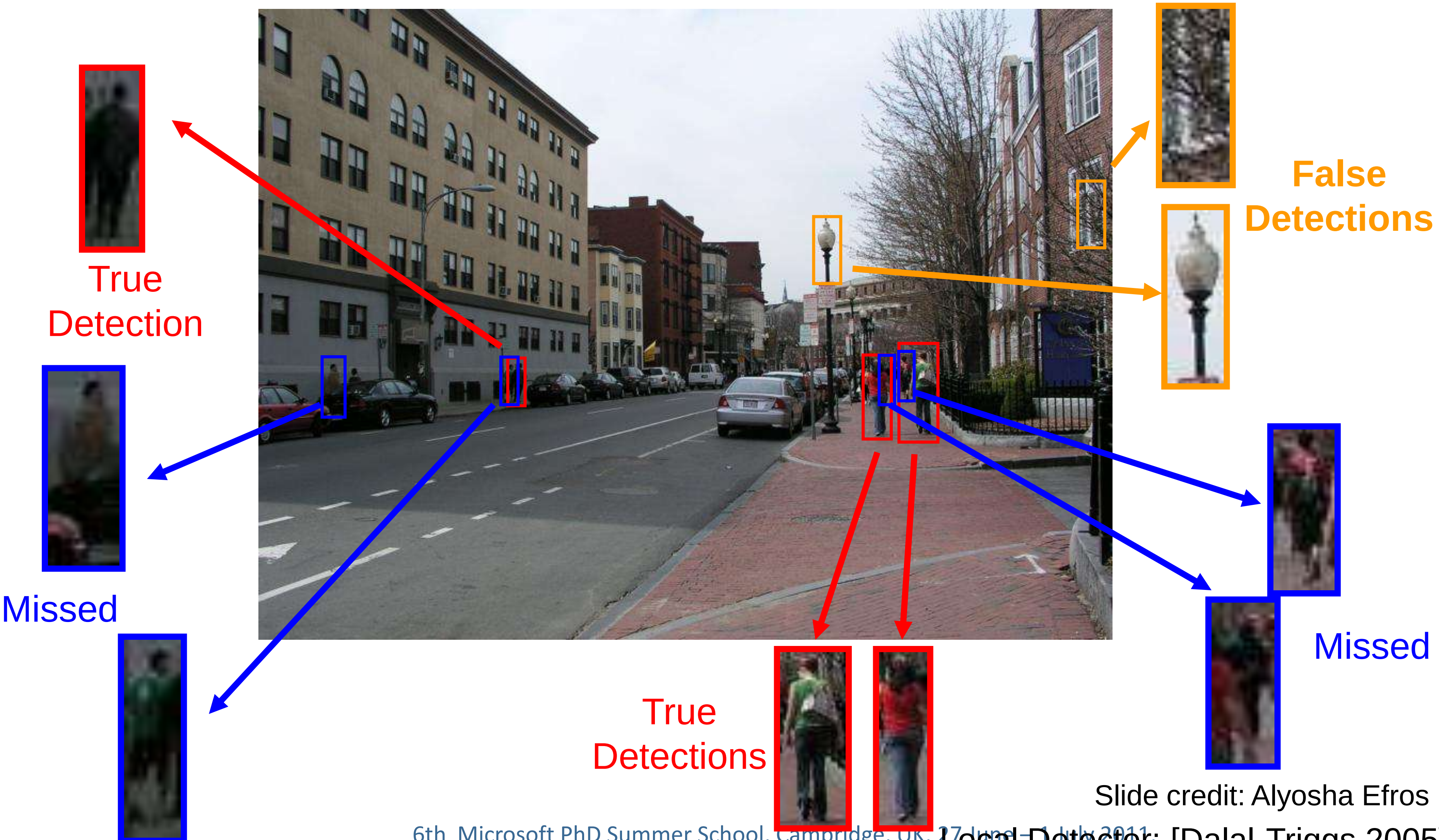
- Works notoriously well for detection of rigid objects, e. g. faces

  [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

Learning local models from local cues

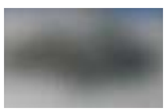
- What the detector sees

Slide credit: Alyosha Efros

Microsoft Research

Learning local models from local cues

- Local ambiguity



Slide credit: Fei-Fei, Fergus & Torralba

Learning local models from local cues

- The role of context

Slide credit: Fei-Fei, Fergus & Torralba

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
- ….

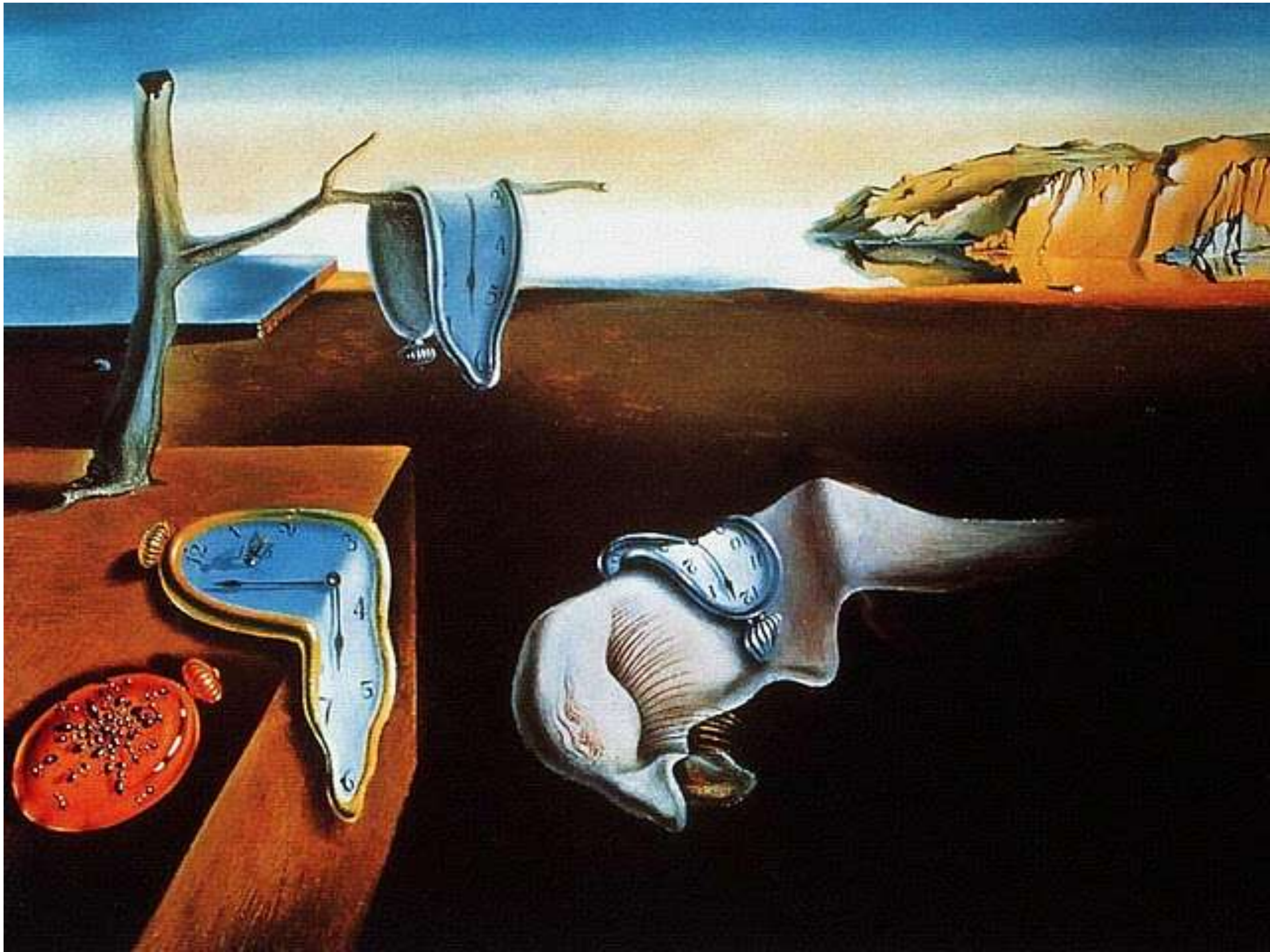
Chaotic world

Structured world

Constraints of the world

- Limited set of allowed deformations for the objects

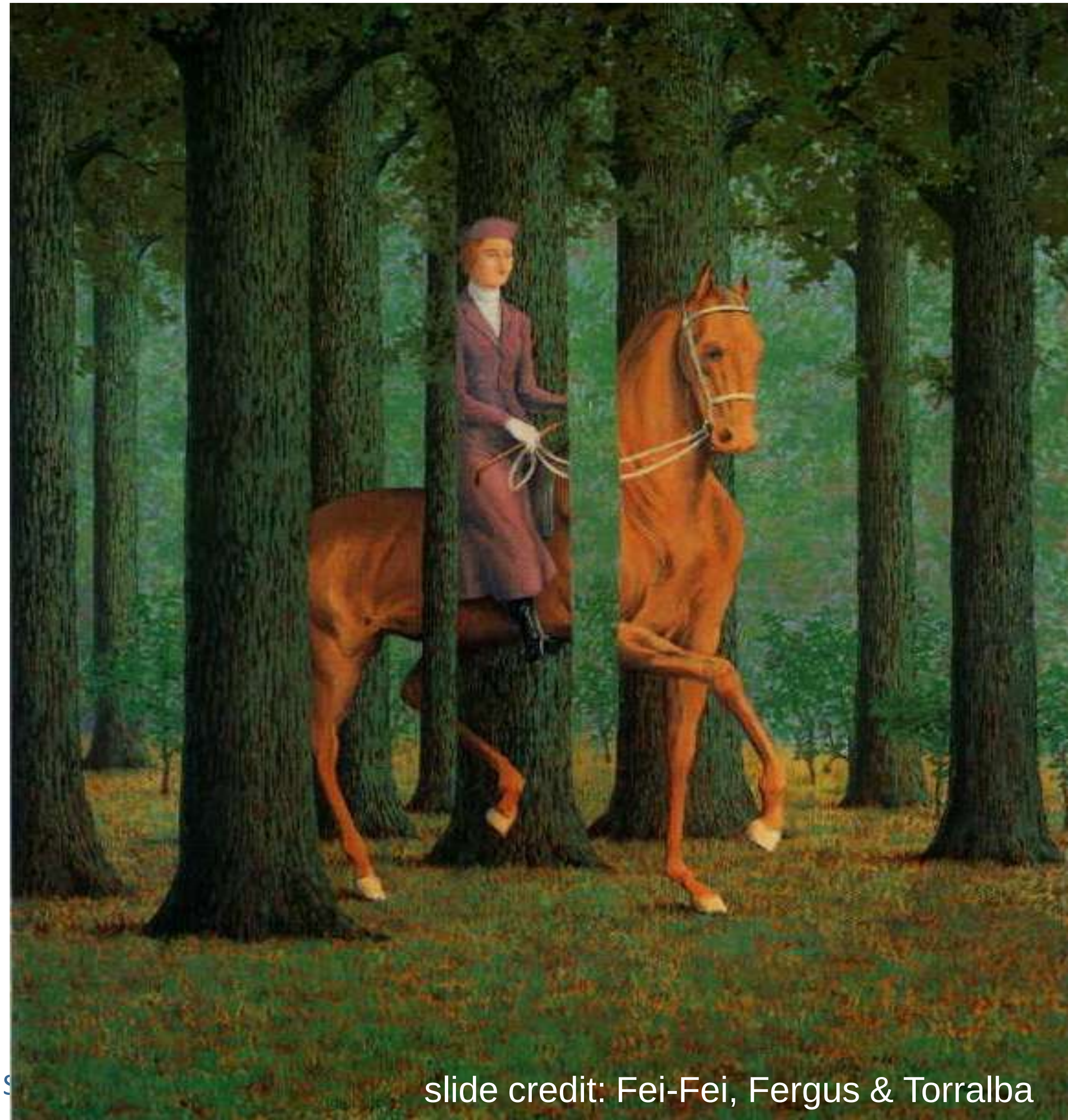
Dali, 1931



Constraints of the world

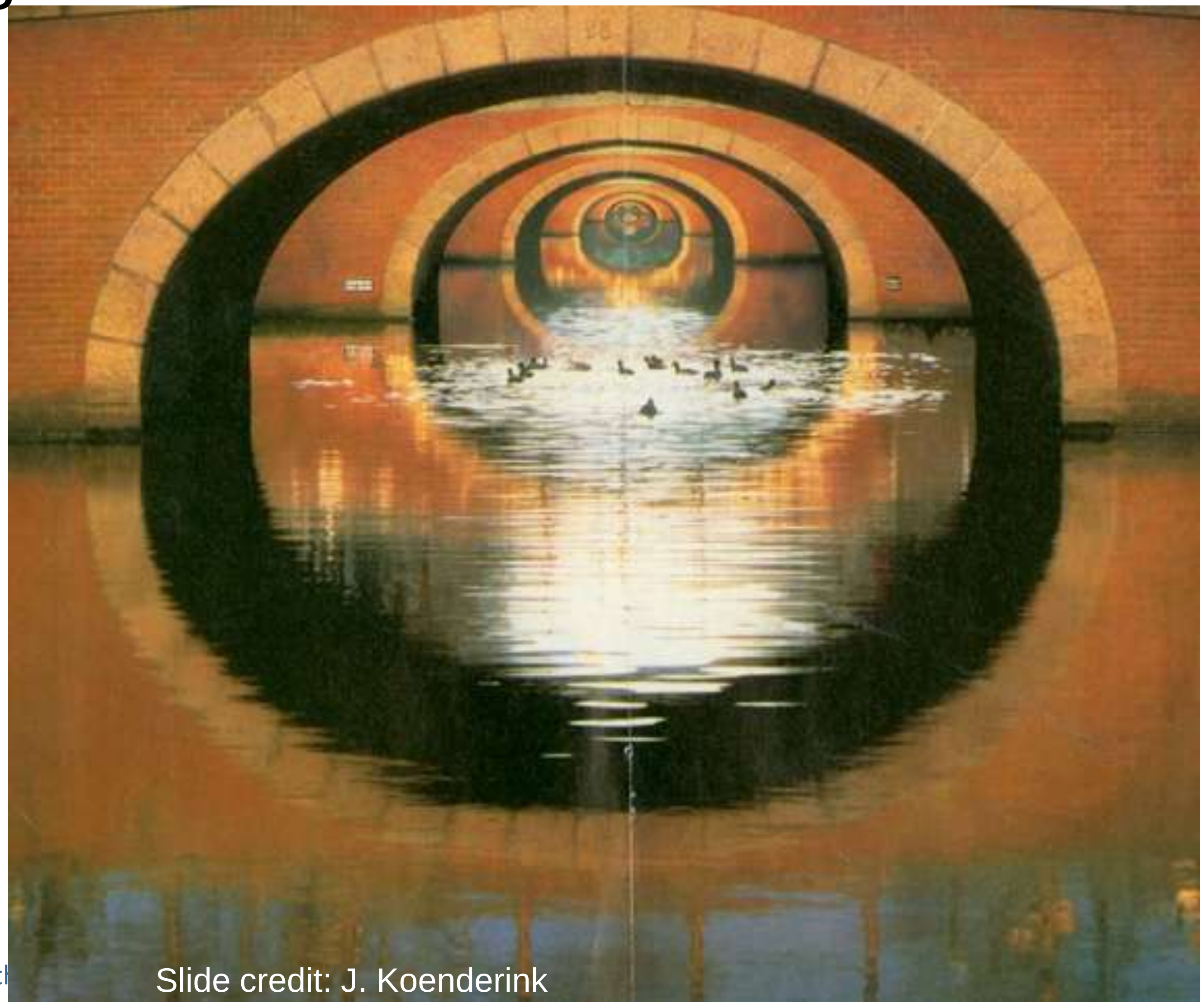
- Occlusions

Magritte, 1957



Constraints of the world

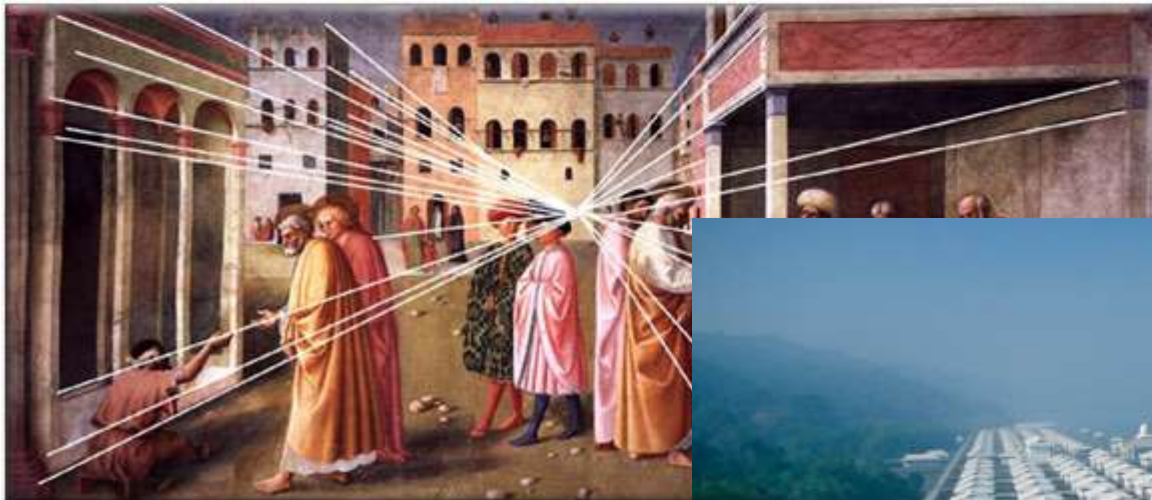
- Depth ordering



Slide credit: J. Koenderink

Constraints of the world

- Rules of perspective geometry

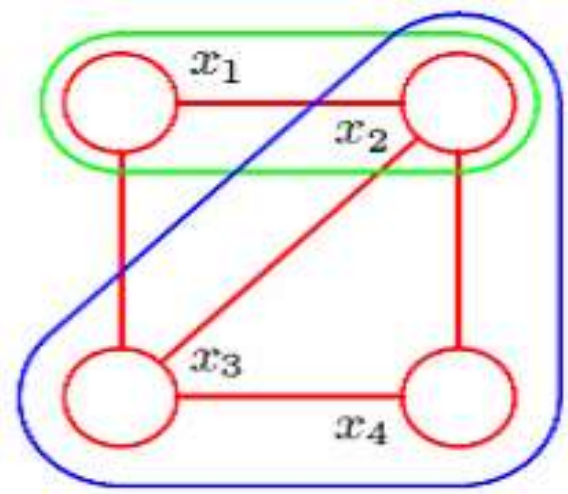


Expressing constraints with graphical models

- 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

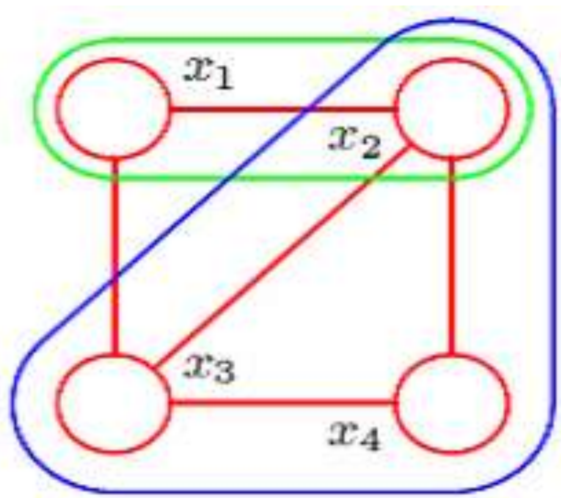
Expressing constraints with graphical models

- 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

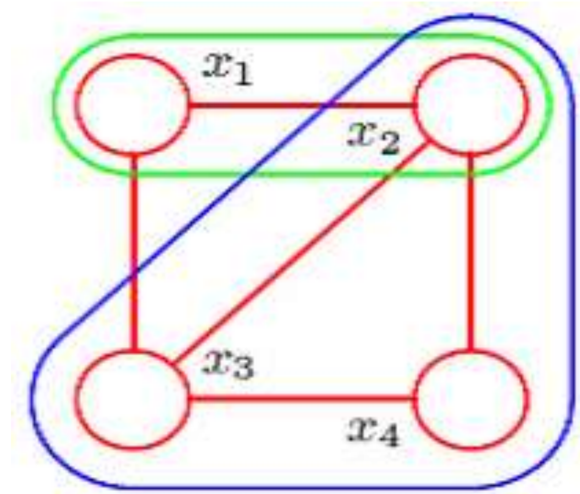


Picture credit: C. Bishop

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

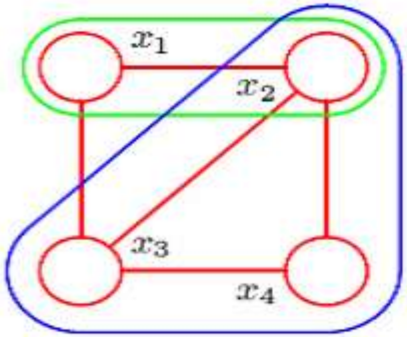
Graphical models

- Joint distribution and potentials

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

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

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



Graphical models

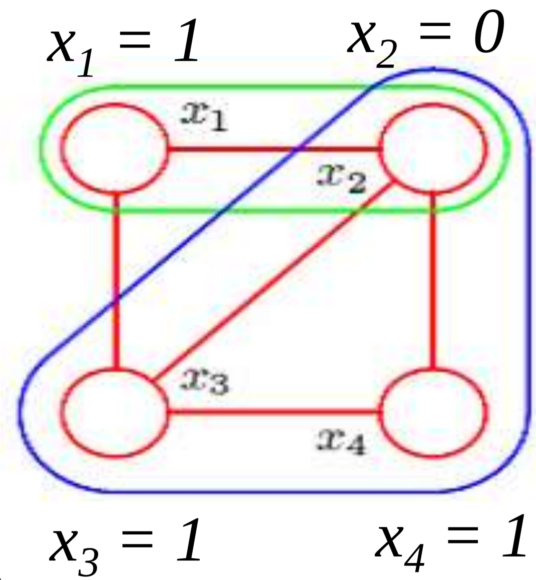
- MAP-inference and energy function

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

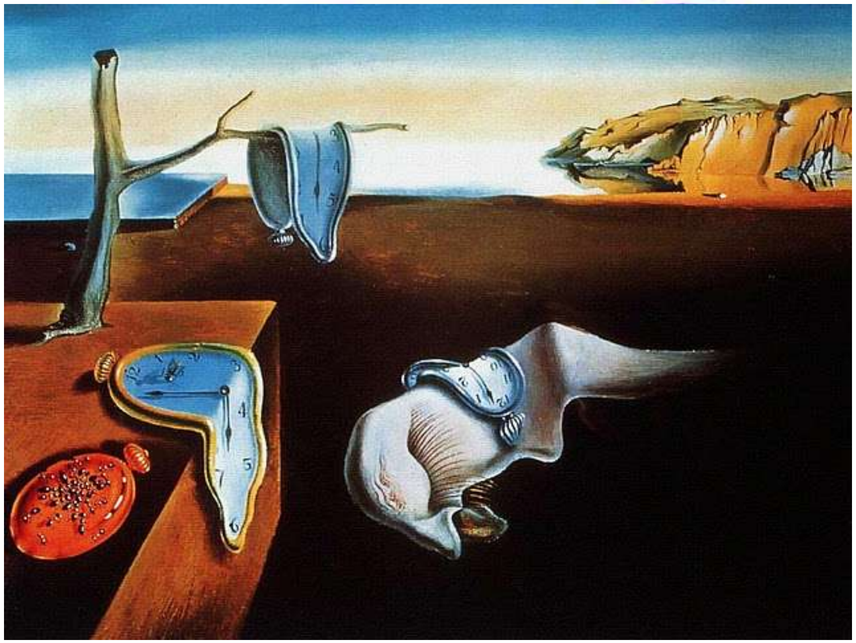
MAP-inference = energy minimization

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

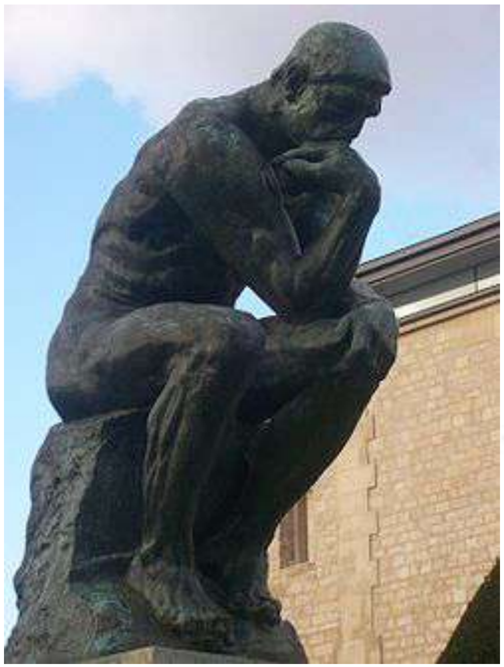
Expressing constraints with graphical models

- 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

- Limiting the set of allowed deformations



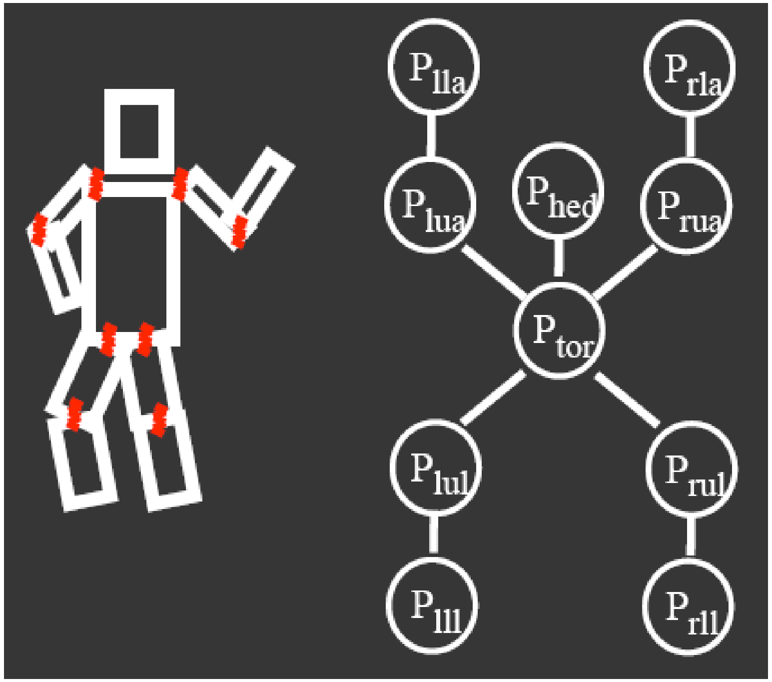
- Model should be flexible enough, but constrain the allowed deformations of an object

Limiting the set of allowed deformations

- Pictorial structures model

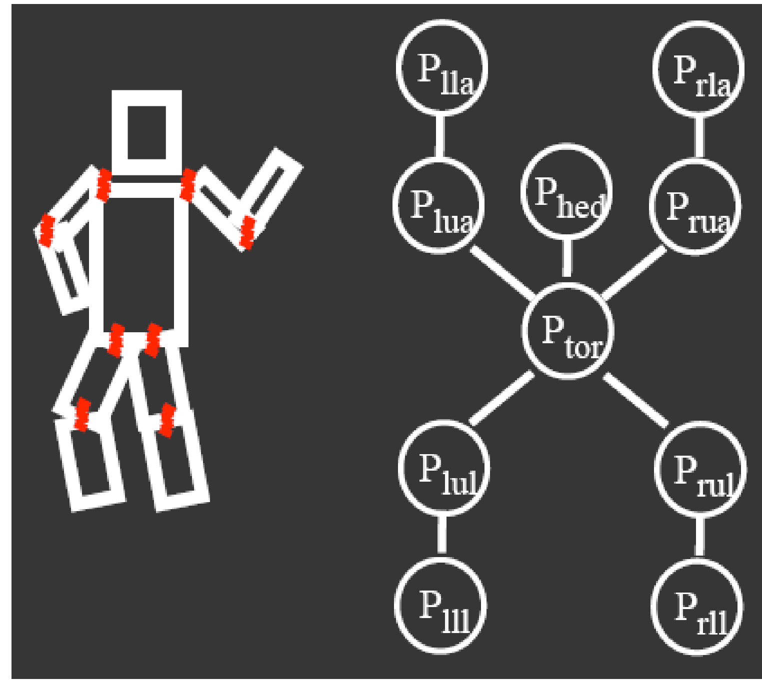
  - **Pictorial structures** strike a good balance between flexibility and tractability

  [Fischler & Elschlager 73],
  [Felzenshwalb & Huttenlocher 00]



Limiting the set of allowed deformations

- Pictorial structures model
  - Each vertex corresponds to a part of a person: *'Head', 'Torso', 'Legs', 'Arms'*
  - Edges form a tree
  - Person detector - for each vertex find a corresponding position from the set of valid positions

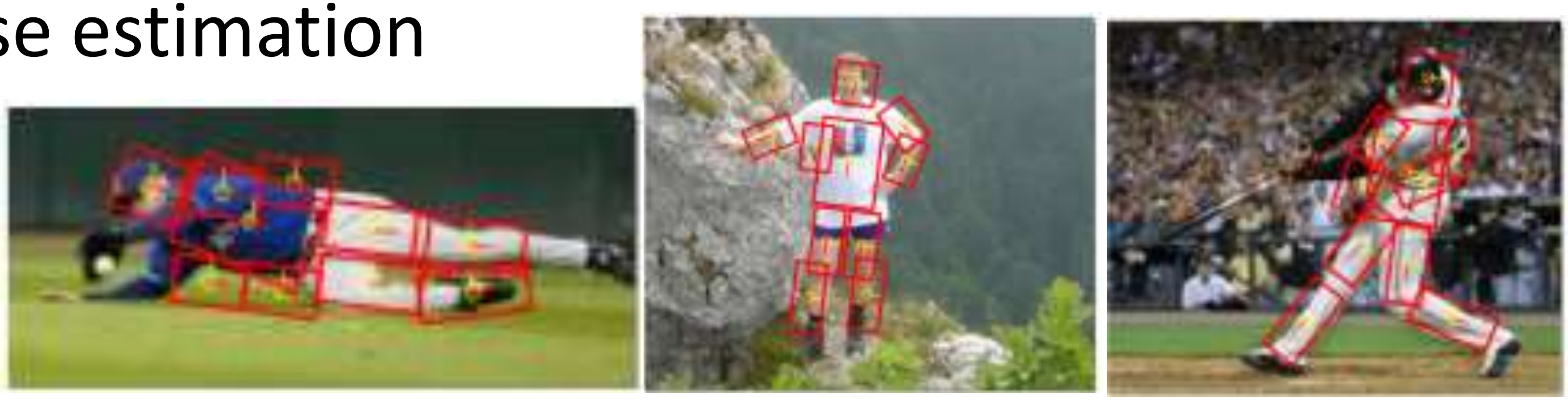


$$\Pr(\mathrm{P}_{tor}, \mathrm{P}_{arm}, \ldots | \mathrm{Im}) \propto \prod_{i,j} \underbrace{\Pr(P_i \mid P_j)}_{\text{part geometry}} \prod_{i} \underbrace{\Pr(\mathrm{Im}(P_i))}_{\text{part appearance}}$$

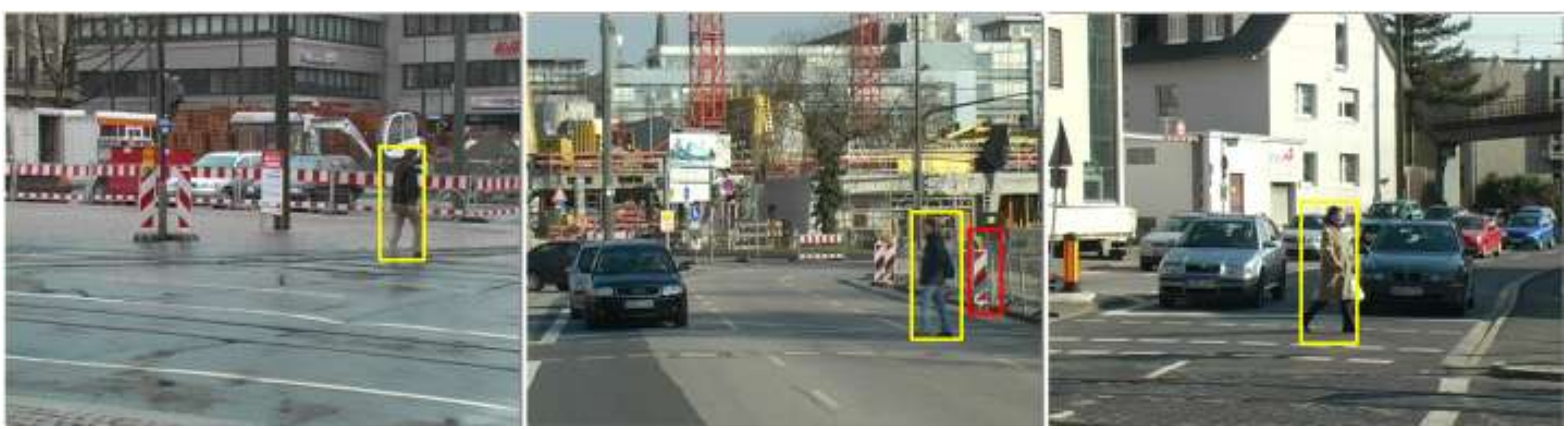
Microsoft Research

Limiting the set of allowed deformations

- Pose estimation



- Pedestrian detection



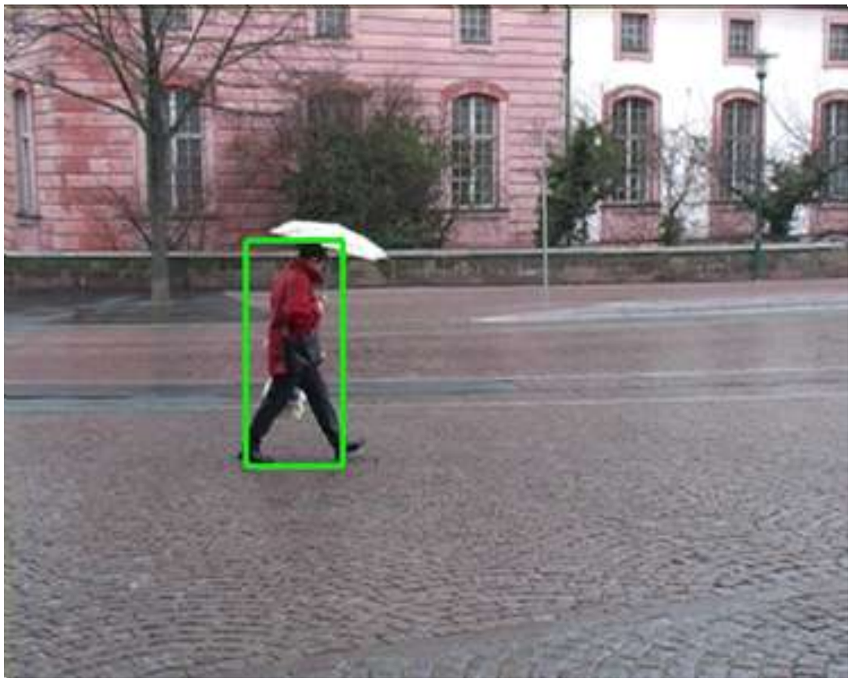
Expressing constraints with graphical models

- 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

Microsoft Research

Expressing constraints with graphical models

- Occlusions

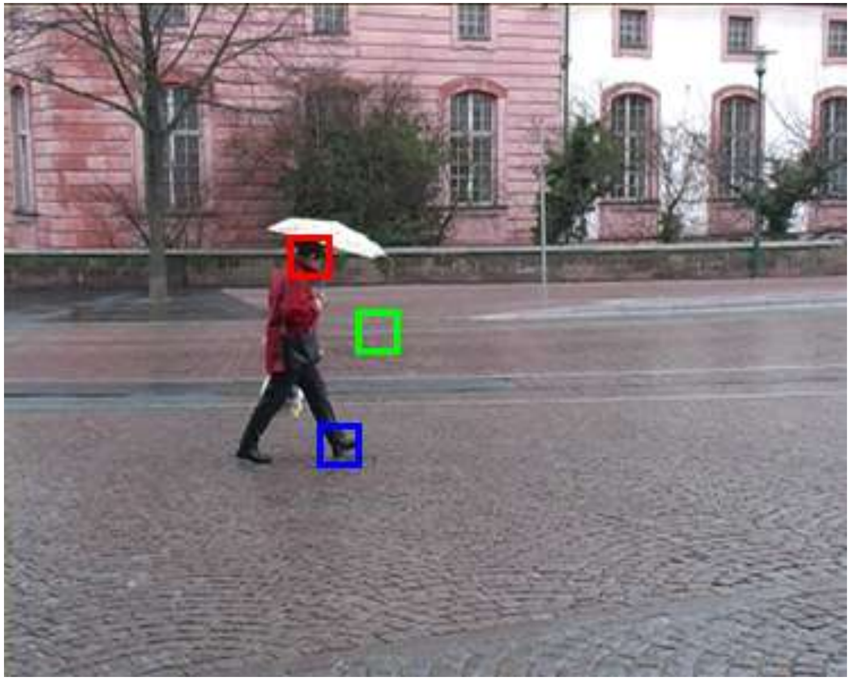


  - Occlusions and self occlusions make the task of object detection even harder

Joint work with Victor Lempitsky and Pushmeet Kohli, CVPR10

Occlusion constraint

- Local model

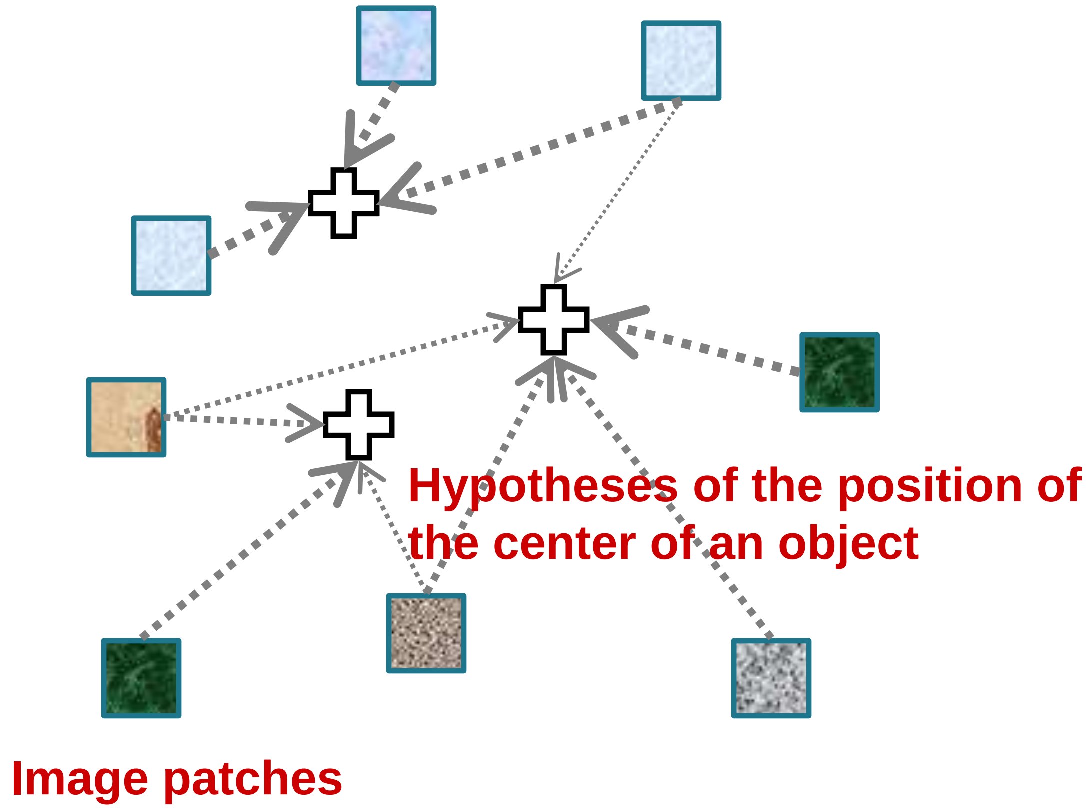


Model from [Gall & Lempitsky, CVPR09]

Joint work with Victor Lempitsky and Pushmeet Kohli, CVPR10

Occlusion constraint

- Each image pixel belong to no at most one object



Joint work with Victor Lempitsky and Pushmeet Kohli, CVPR10

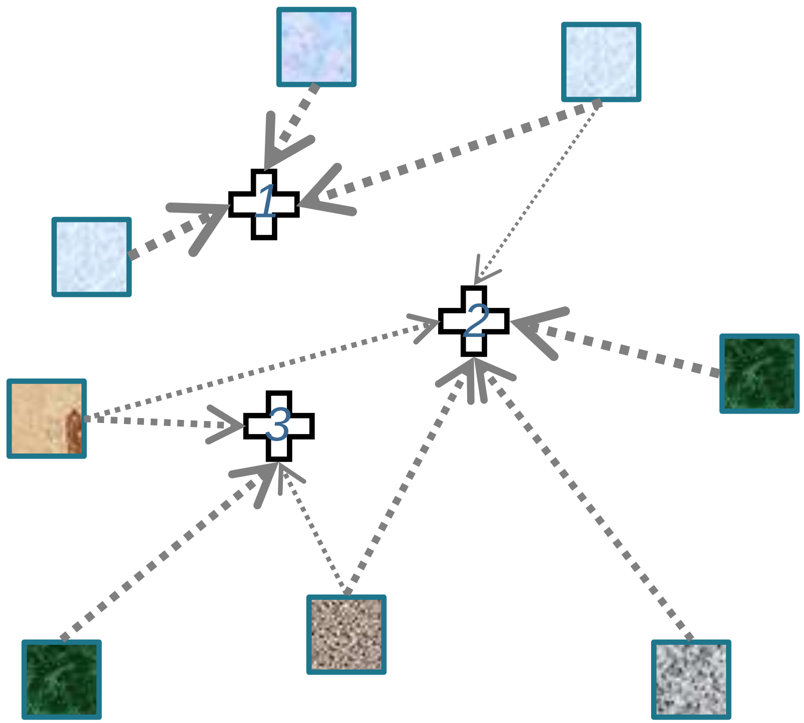
Occlusion constraint

- Modeling the occlusion constraint

x – labelling of image patches,

x_i = index of hypothesis, if the patch votes for hypothesis,

$x_i = 0$, if the patch votes for background



y – labelling of hypotheses, binary variables:

$y = 1$, if the object is present,

$y = 0$, otherwise

Joint work with Victor Lempitsky and Pushmeet Kohli, CVPR10

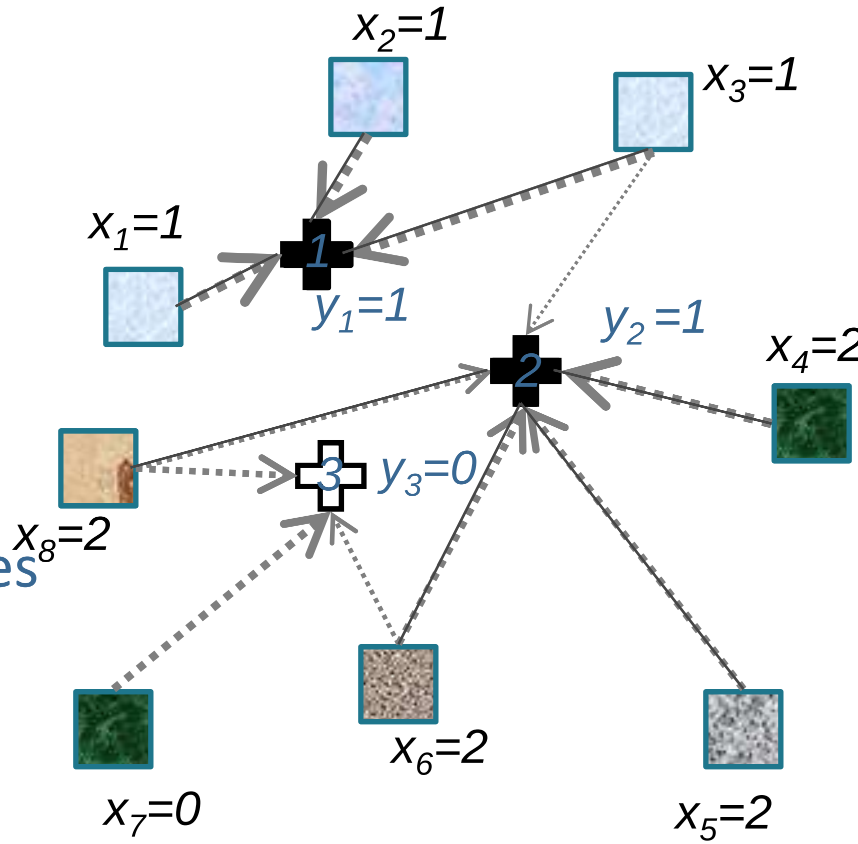
Occlusion constraint

- Modeling the occlusion constraint

$\boldsymbol{x}$ – labelling of image patches,

x_i = index of hypothesis, if the patch votes for hypothesis,

$x_i = 0$, if the patch votes for background



$\boldsymbol{y}$ – labelling of hypotheses, binary variables:

$y = 1$, if the object is present,

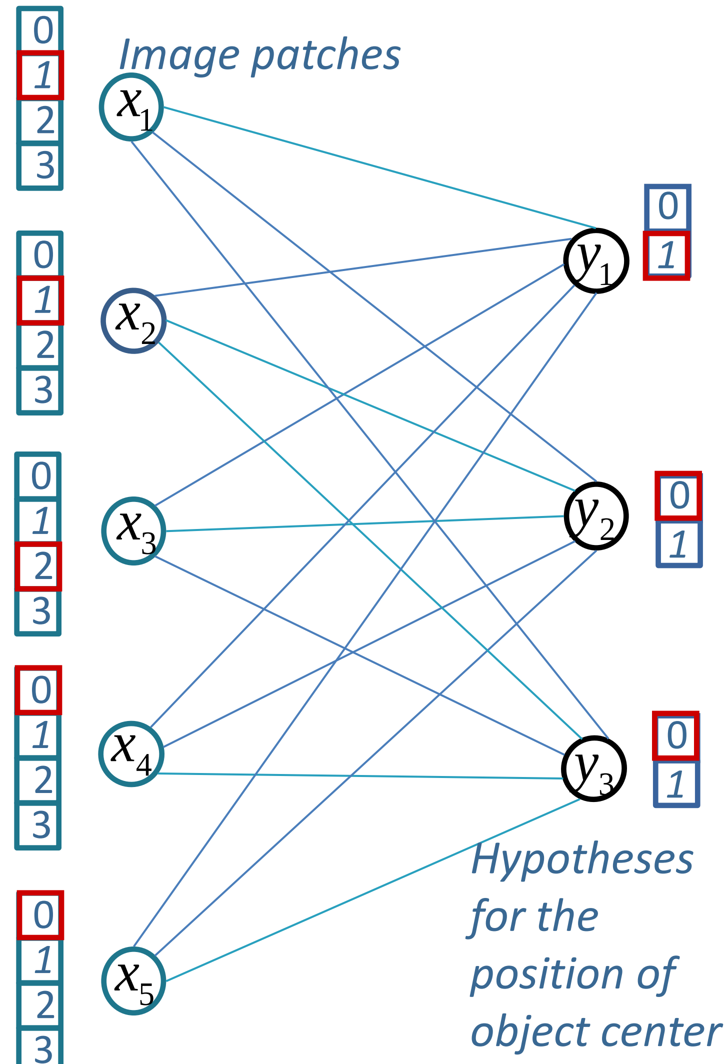
$y = 0$, otherwise

Key idea : joint MAP-inference in x and y

Joint work with Victor Lempitsky and Pushmeet Kohli, CVPR10

Occlusion constraint

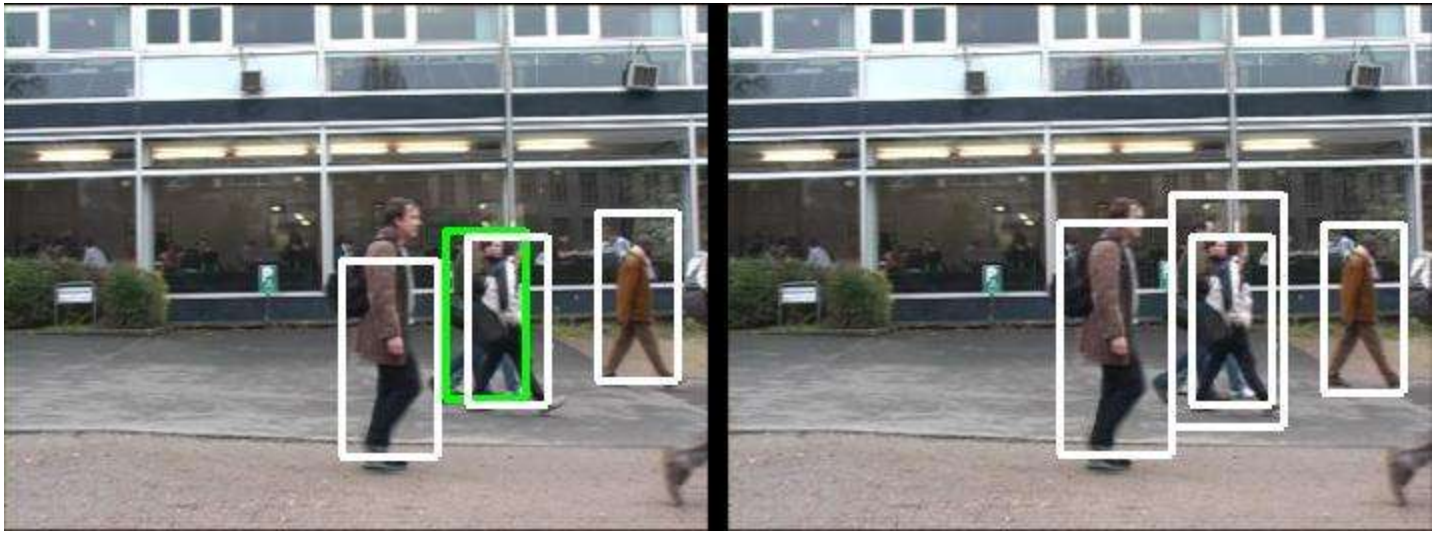
- Graphical model
  - If labeling of $\mathbf{y}$ is fixed, the values of x_i are independent
  - So we can maximize $\mathbf{x}$ out first and perform inference over $\mathbf{y}$



Joint work with Victor Lempitsky and Pushmeet Kohli, CVPR10

Occlusion constraint

- Comparison



Without occlusion constraint   Using occlusion constraint

White = correct detection

Green = missing object

Red = false positive

Code available online!!!

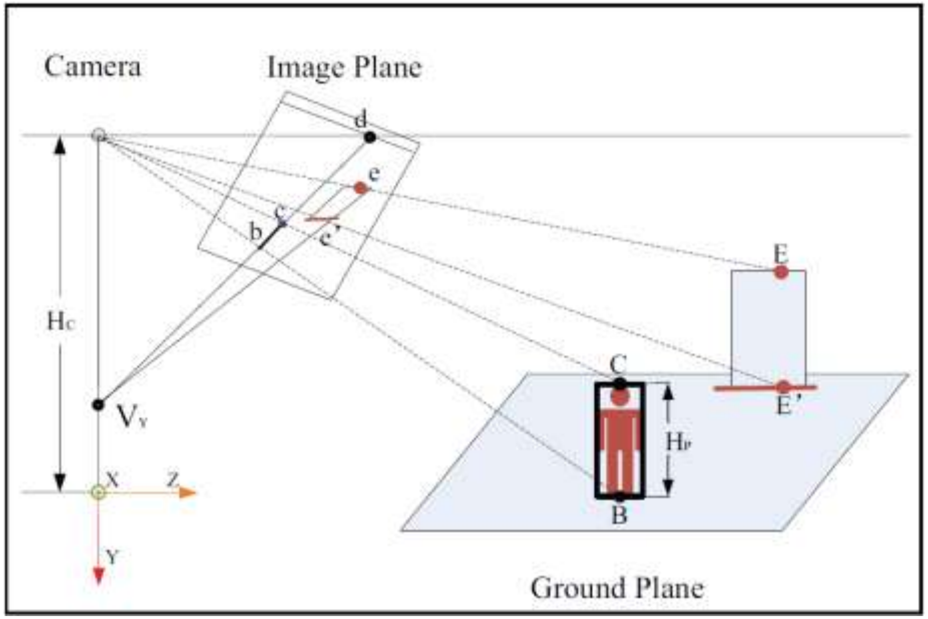
Joint work with Victor Lempitsky and Pushmeet Kohli, CVPR10

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

- Depth ordering

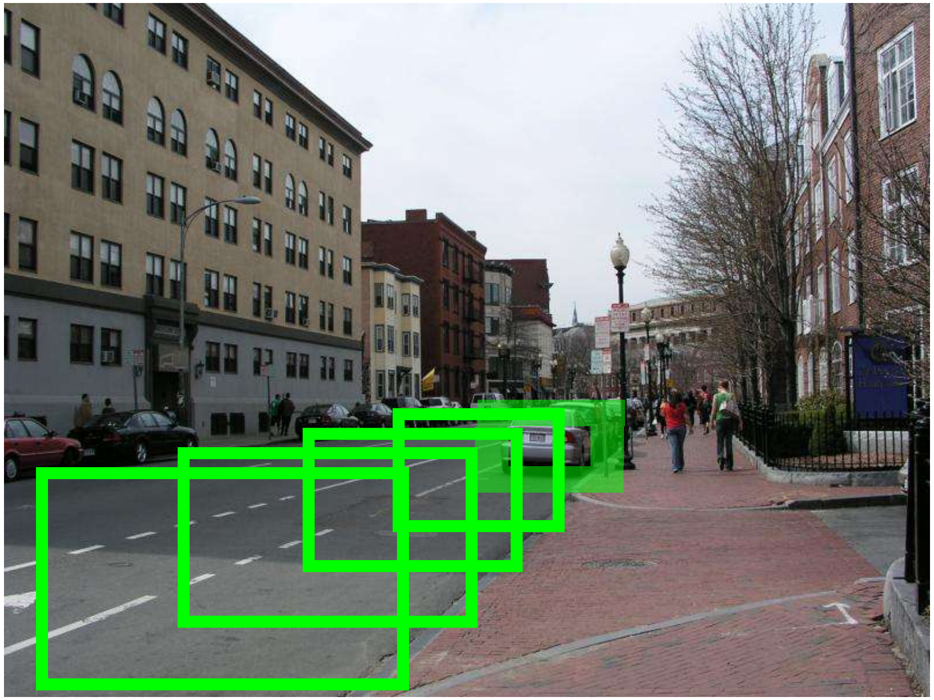


- The size of an object depends on the distance from the viewpoint
- Viewpoint is set by the position of horizon and ground plane in the image

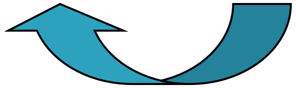
Depth ordering constraint

- Viewpoint → prior on the size of the objects

Object Position/Sizes



Viewpoint

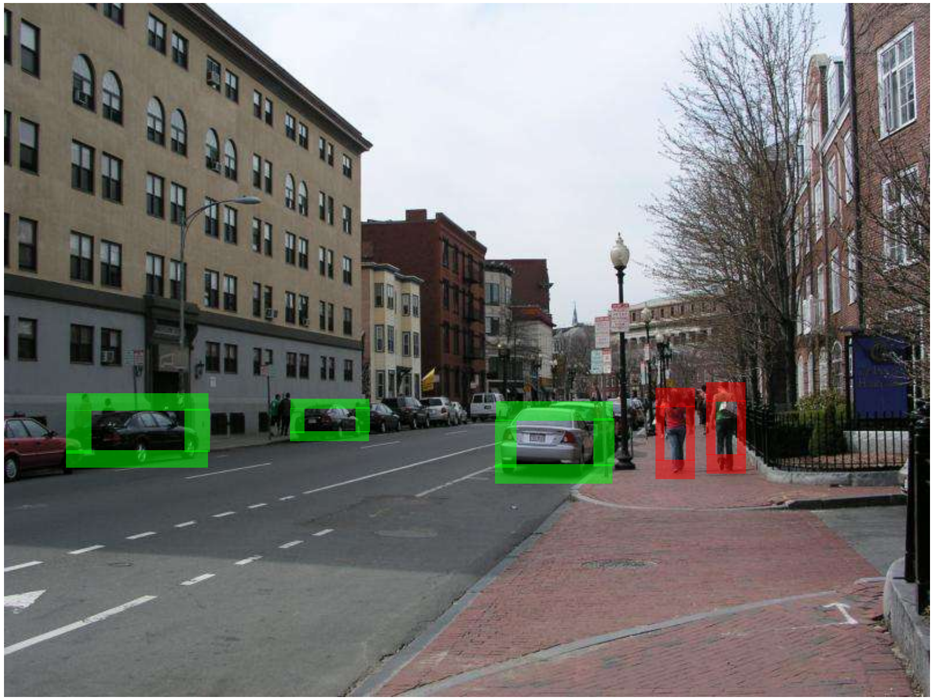


Hoiem et al. , CVPR2006

Depth ordering constraint

- Detected objects → viewpoint

Object Position/Sizes

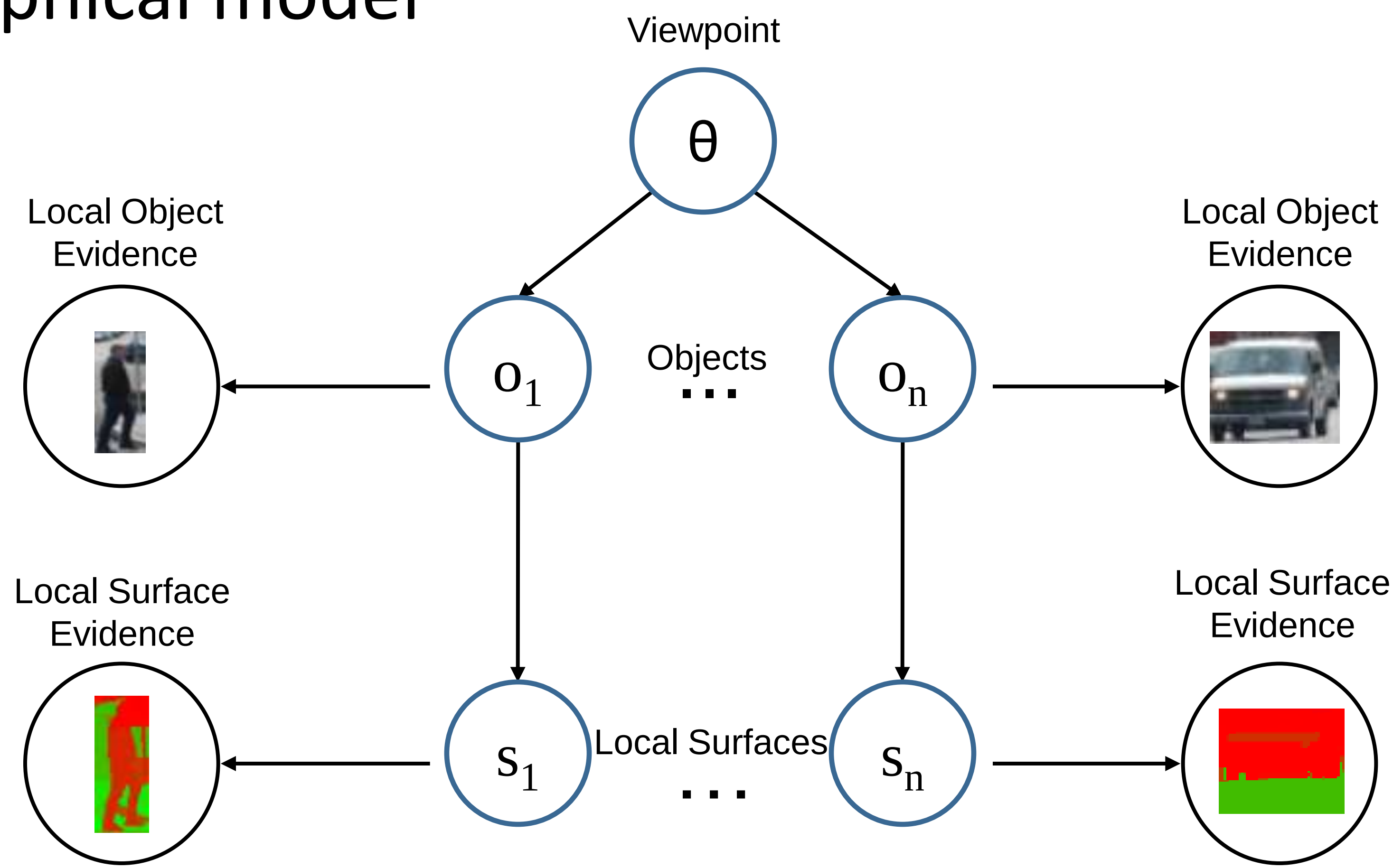


Viewpoint

Hoiem et al. , CVPR2006

Depth ordering constraint

- Graphical model



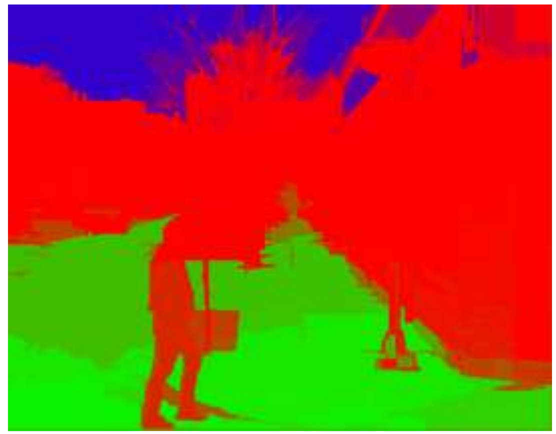
Hoiem et al. , CVPR2006

Microsoft Research

Depth ordering constraint

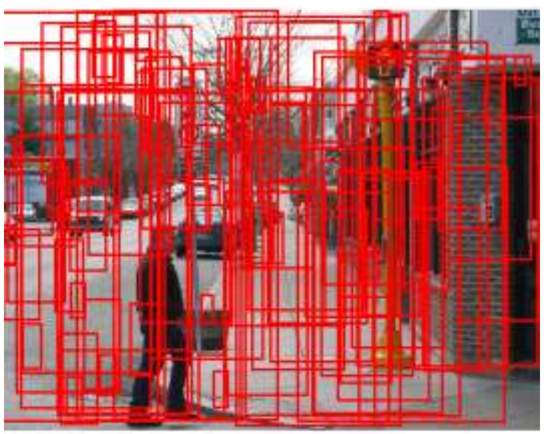
- Prior on the size an position of the objects

Image



P(surfaces)

P(viewpoint)



P(object)

P(object | surfaces, viewpoint)

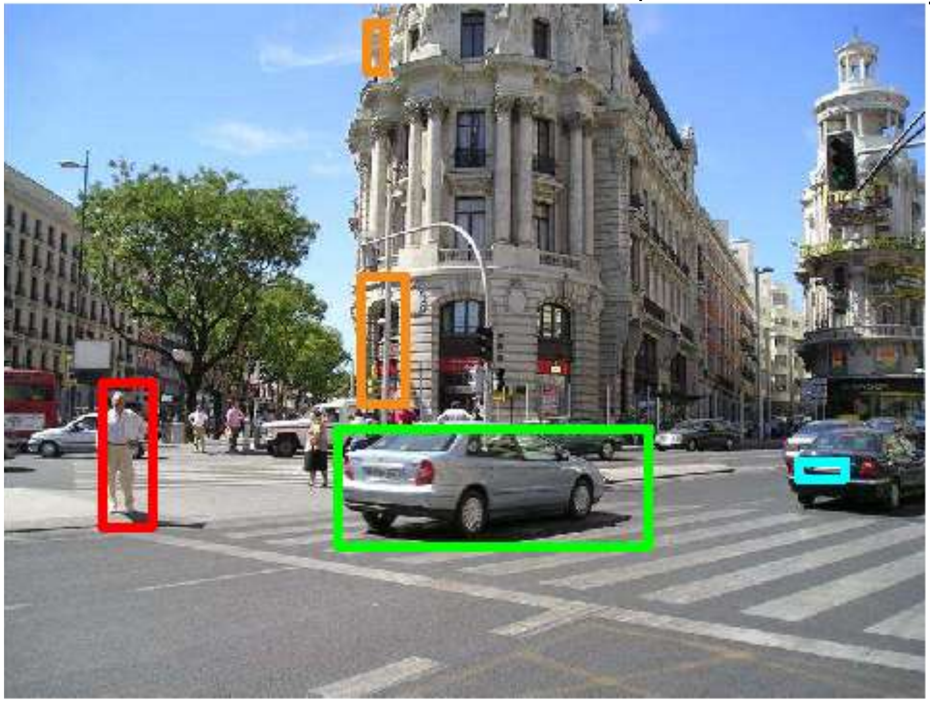
Hoiem et al. , CVPR2006

Microsoft Research

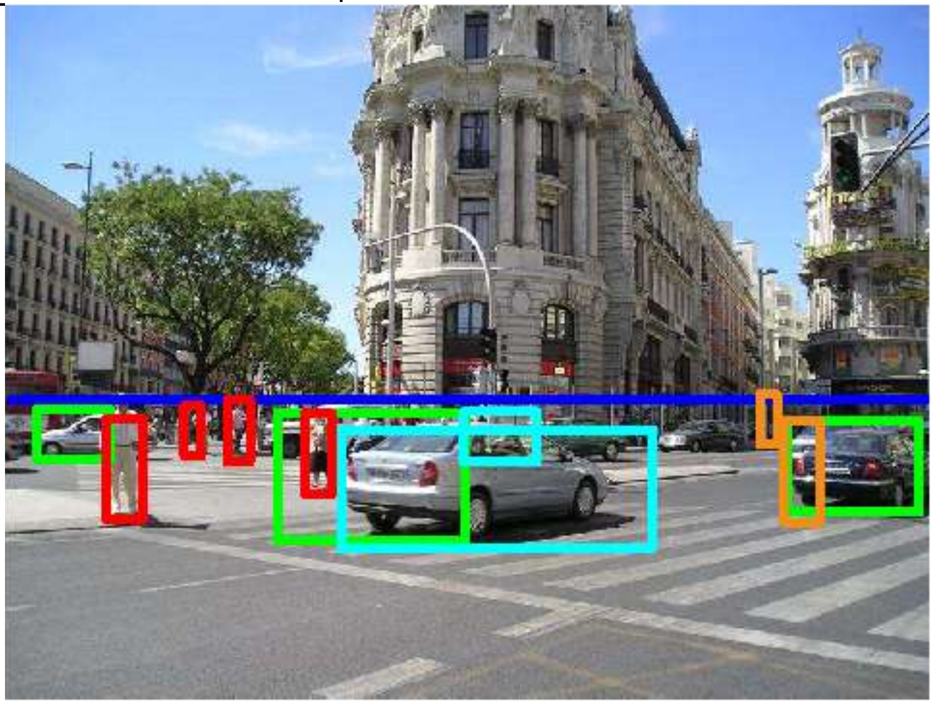
Depth ordering constraint

- Comparison

Car: TP / FP Ped: TP / FP



Initial: 2 TP / 3 FP



Final: 7 TP / 4 FP

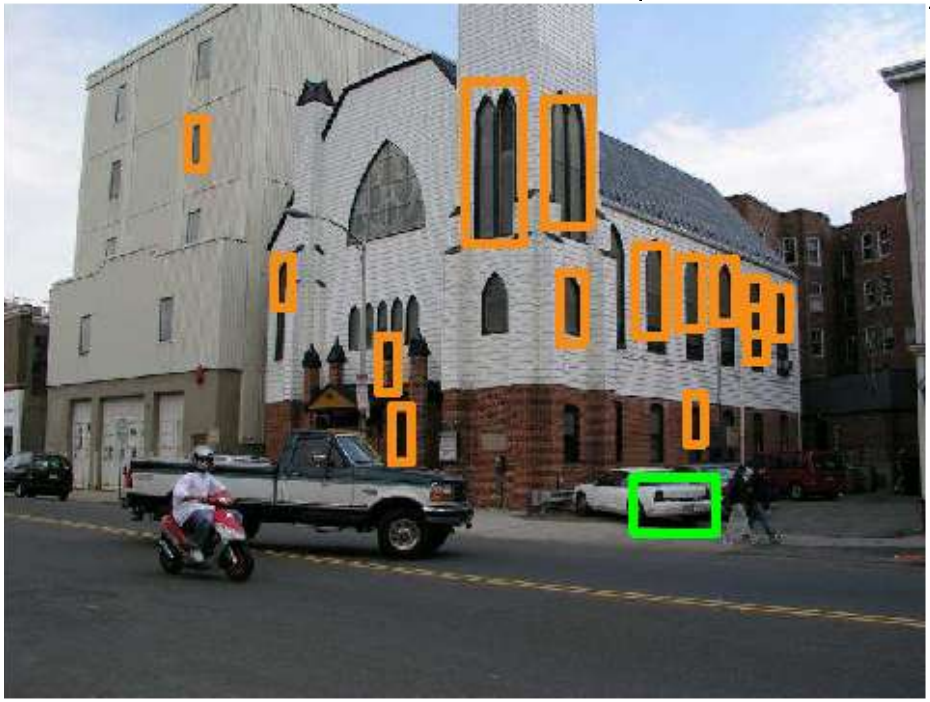
Hoiem et al. , CVPR2006

Local Detector from [Murphy-Torralba-Freeman 2003]

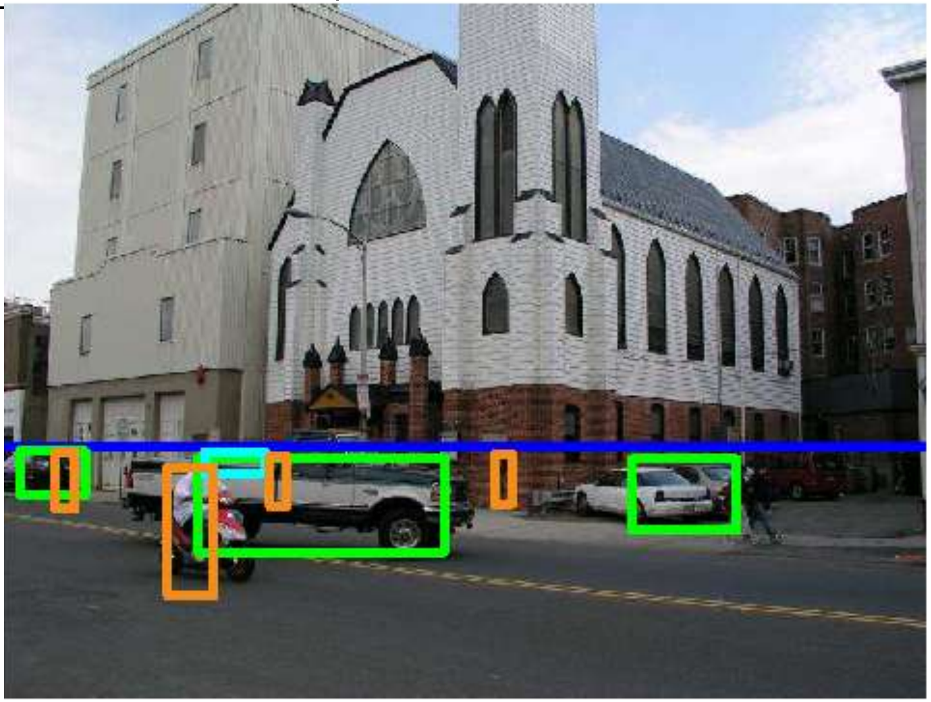
Depth ordering constraint

- Comparison

Car: TP / FP Ped: TP / FP



Initial: 1 TP / 14 FP



Final: 3 TP / 5 FP

Hoiem et al. , CVPR2006

Local Detector from [Murphy-Torralba-Freeman 2003]

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

- Rules of perspective geometry
  - Straight lines lying on parallel planes in 3D intersect in the image plane in **vanishing point**
  - Vanishing point corresponding to vertical lines is called **zenith**
  - All vanishing points which correspond to horizontal lines lie on the same straight line called **horizon**

Modeling the rules of perspective geometry

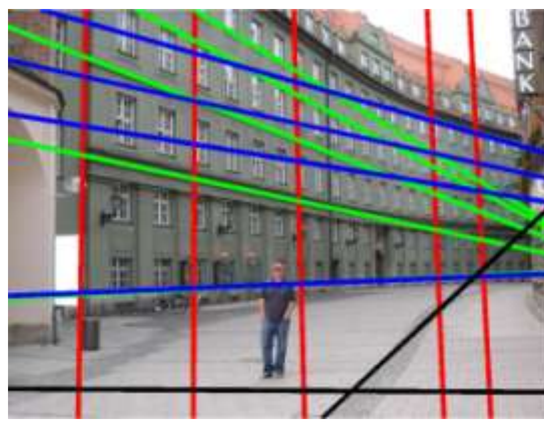
- Geometric primitives and geometric parsing

Image of man-made environment

Edge pixels

Line segments

Lines



Vanishing points

Zenith and horizon

Joint work with Elena Tretyak, Victor Lempitsky and Pushmeet Kohli, ECCV10

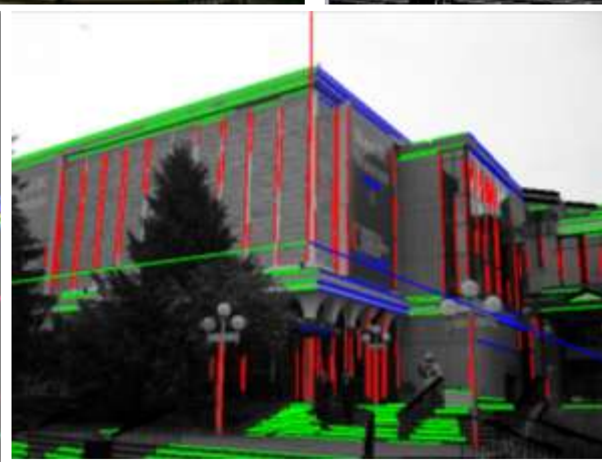
Modeling the rules of perspective geometry

- Traditional approach: bottom-up pipeline

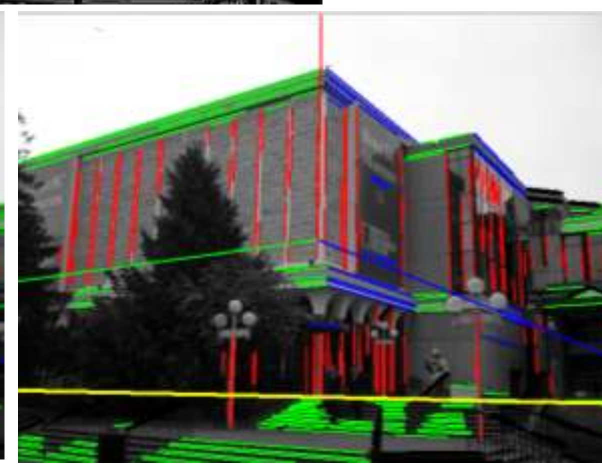
Input image

Edge map

1. *Grouping edge pixels into line segments*



2. *Grouping line segments and VPs estimation*



3. *Horizon and zenith estimation*

*Tuytelaars 1998, Antone 2000, Almansa 2003, Aguilera 2005, Boulanger 2006 , Tardif 2009

Microsoft Research

Modeling the rules of perspective geometry

- Energy function
  - All geometric primitives are detected in the simultaneously by energy minimization:

$$E_{total}(\mathbf{s},\mathbf{l},\mathbf{h},z|\mathbf{p}) = \sum_{i=1..P} E_{edge}(p_i|\mathbf{s}) + \sum_{i=1..S} E_{segment}(s_i|\mathbf{l}) + \sum_{i=1..L} E_{line}(l_i|\mathbf{h},z) + \sum_{1\le i<j\le H} E_{horizon}(h_i,h_j|z) + E_{prior}(\mathbf{s},\mathbf{l},\mathbf{h}),$$

p – edge pixels, ***s*** – line segments, ***l*** – lines,
z – zenith, ***h*** – horizontal vanishing points

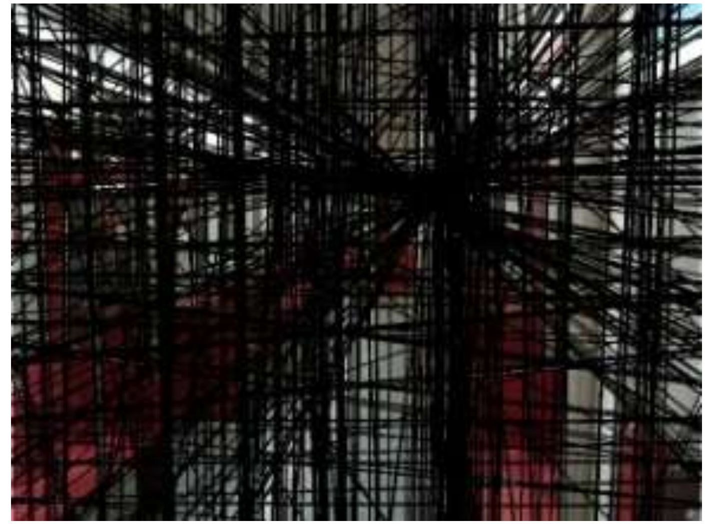
Joint work with Elena Tretyak, Victor Lempitsky and Pushmeet Kohli, ECCV10

Modeling the rules of perspective geometry

- Discretization of the model

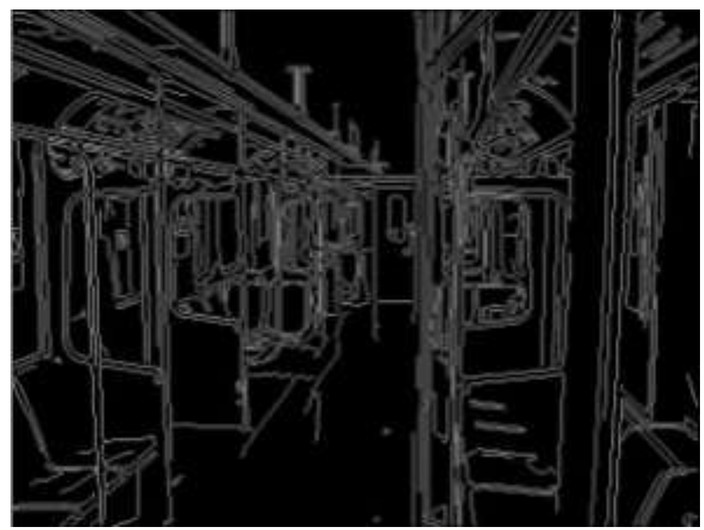
Candidate vanishing points

Candidate lines



Candidate line segments

Candidate edge pixels

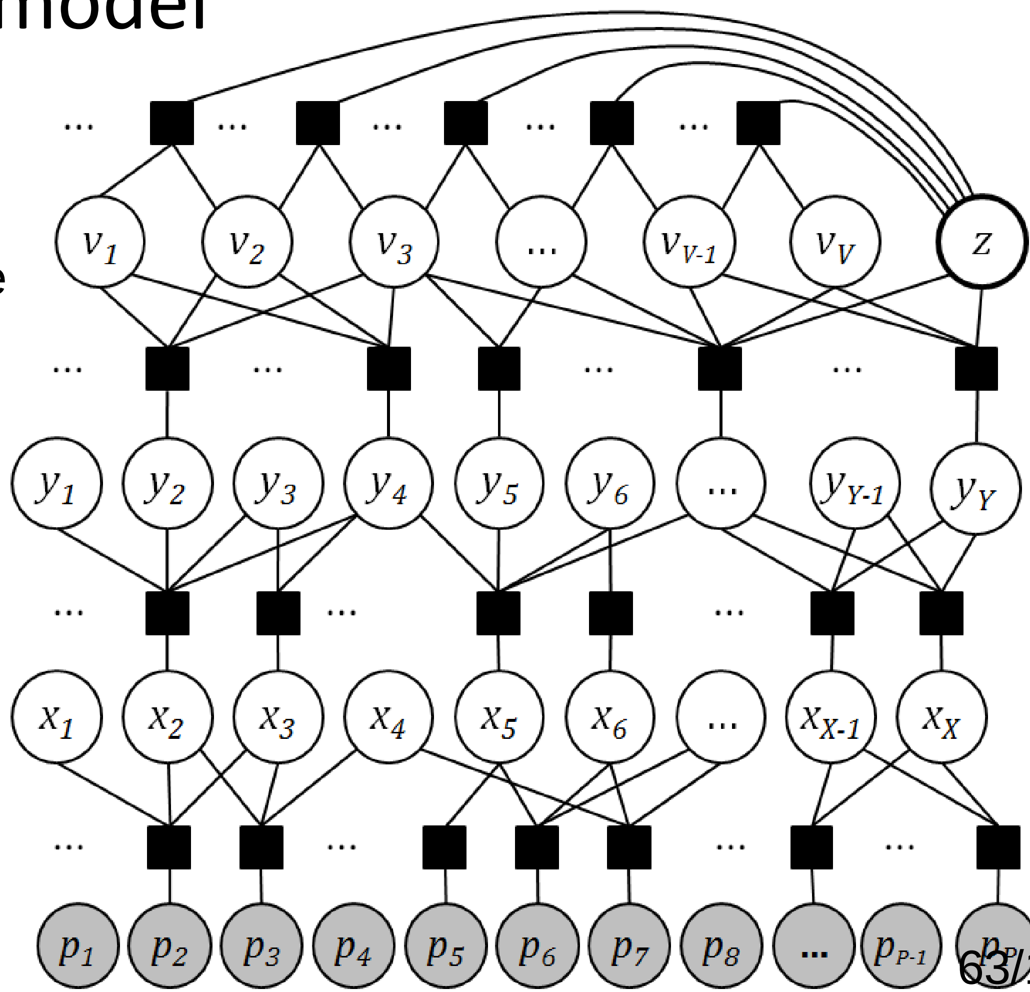


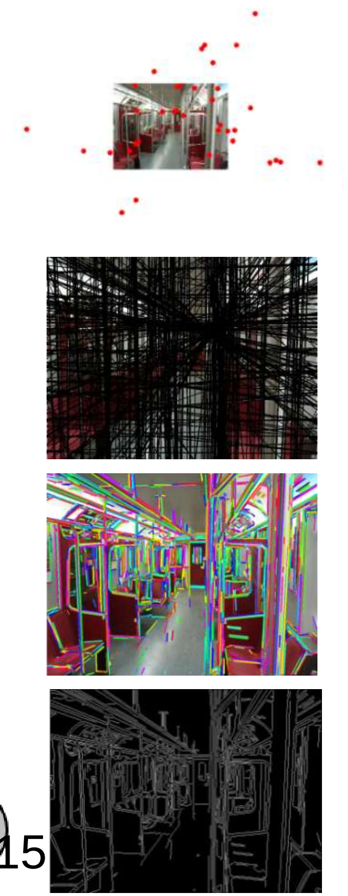
Modeling the rules of perspective geometry

- Graphical model

■ – factors correspond to the potentials of the energy function

○ – nodes correspond to geometric primitives





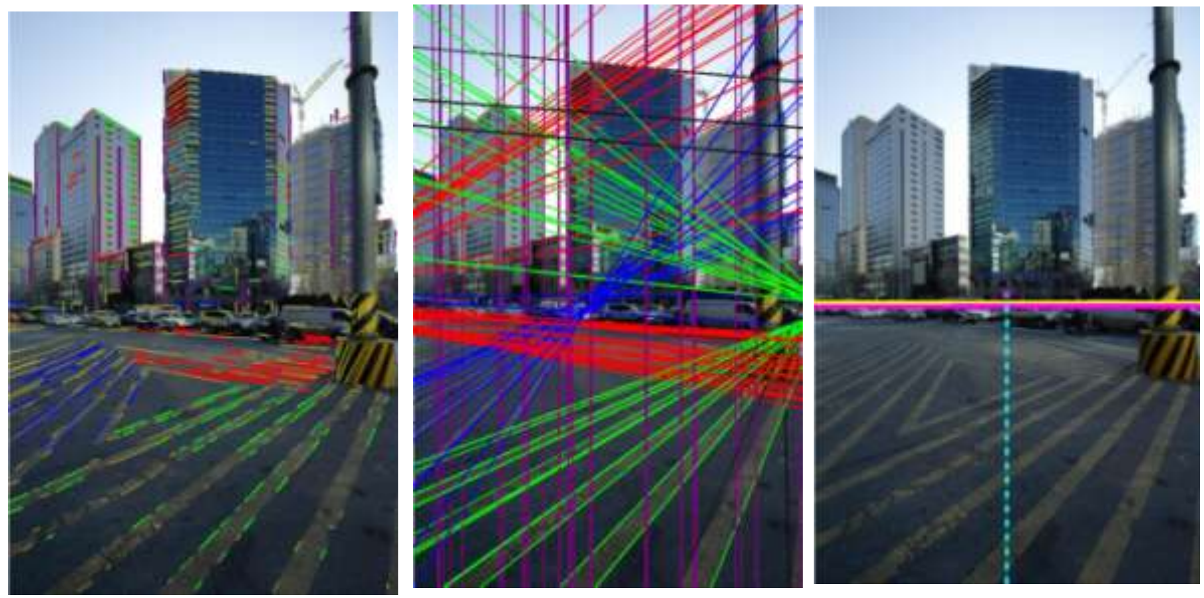
Joint work with Elena Tretyak, Victor Lempitsky and Pushmeet Kohli, ECCV10

Microsoft Research

Modeling the rules of perspective geometry

- Results of geometric parsing

Energy minimization



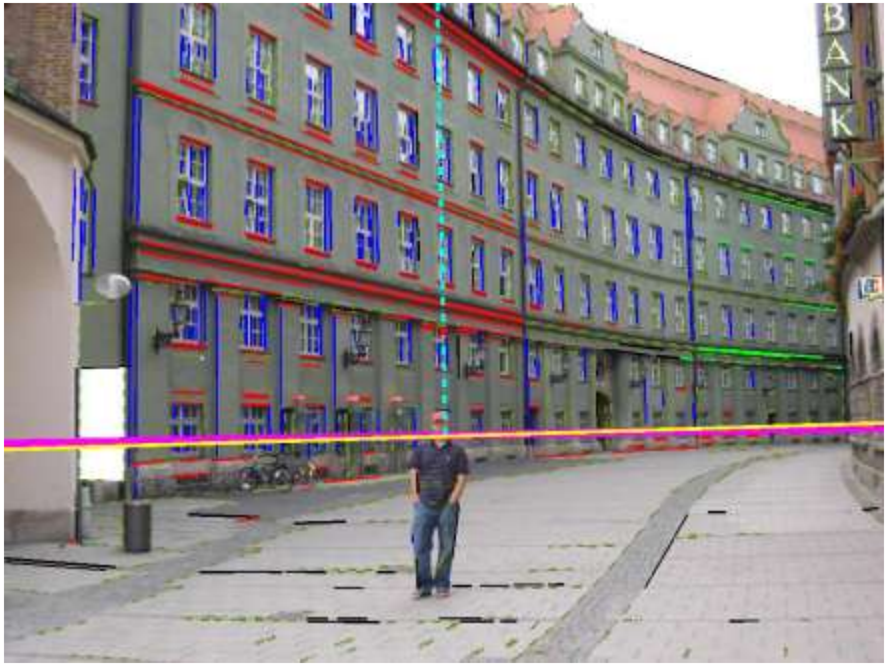
Input image

Detected geometric primitives

Joint work with Elena Tretyak, Victor Lempitsky and Pushmeet Kohli, ECCV10

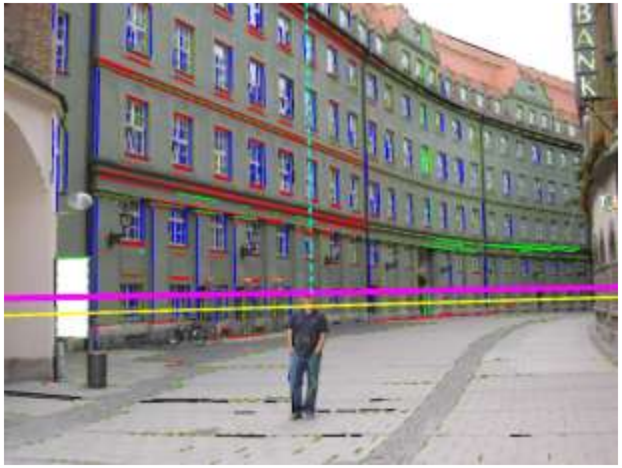
Modeling the rules of perspective geometry

- Comparison

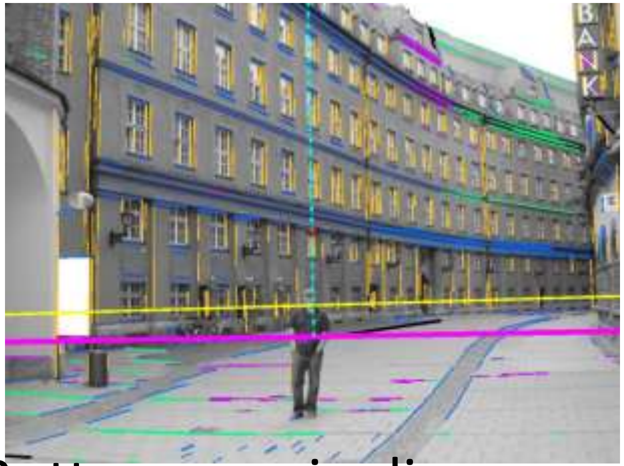


Result of geometric parsing
with full energy

Code available online!!!



Omitting horizon constraint

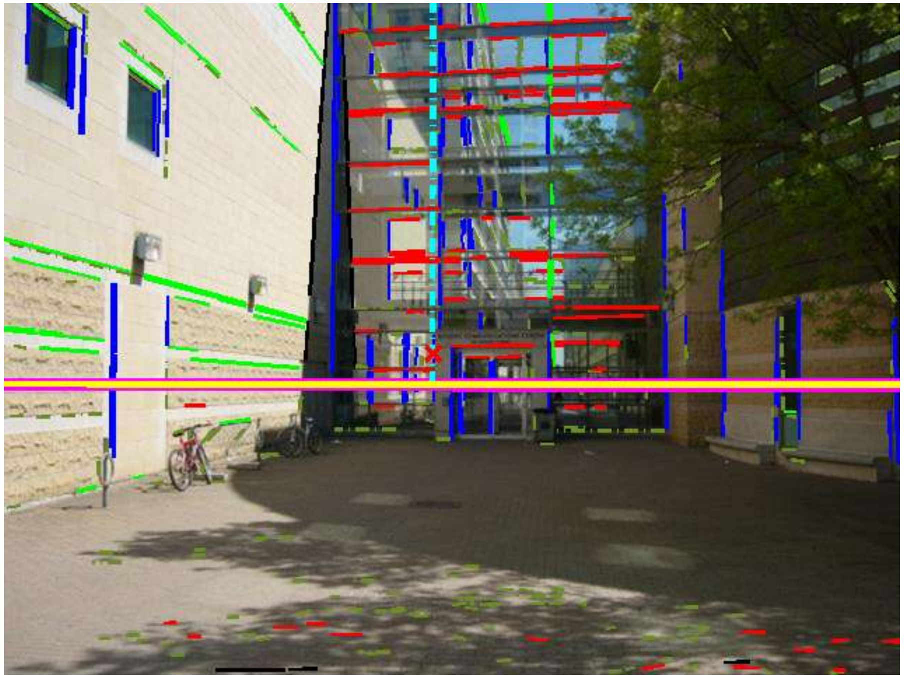


Bottom-up pipeline

Microsoft Research

Modeling the rules of perspective geometry

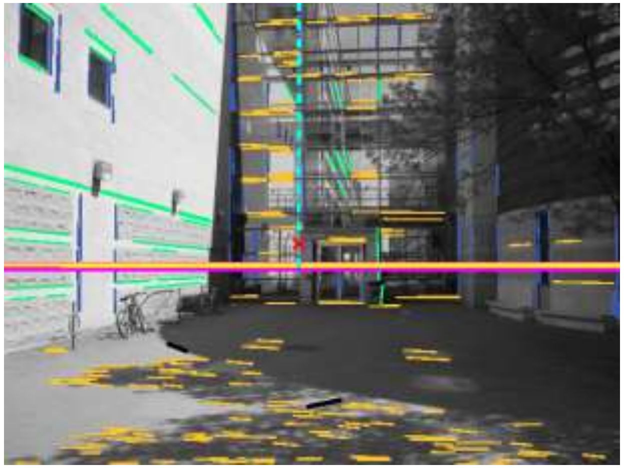
- Comparison



Result of geometric parsing with full energy

Code available online!!!

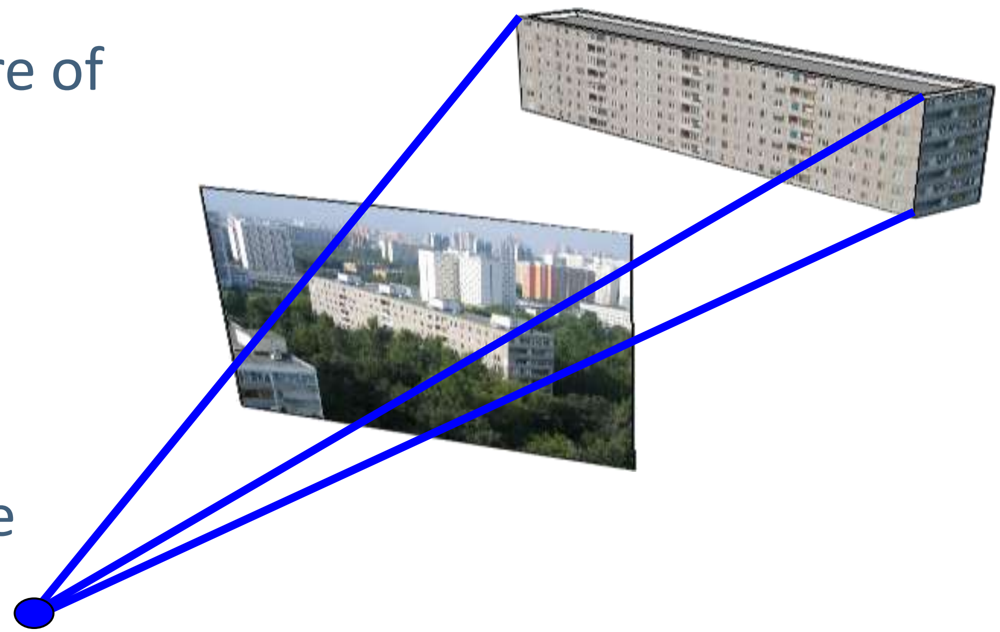
Omitting horizon constraint



Bottom-up pipeline

Modeling the rules of perspective geometry

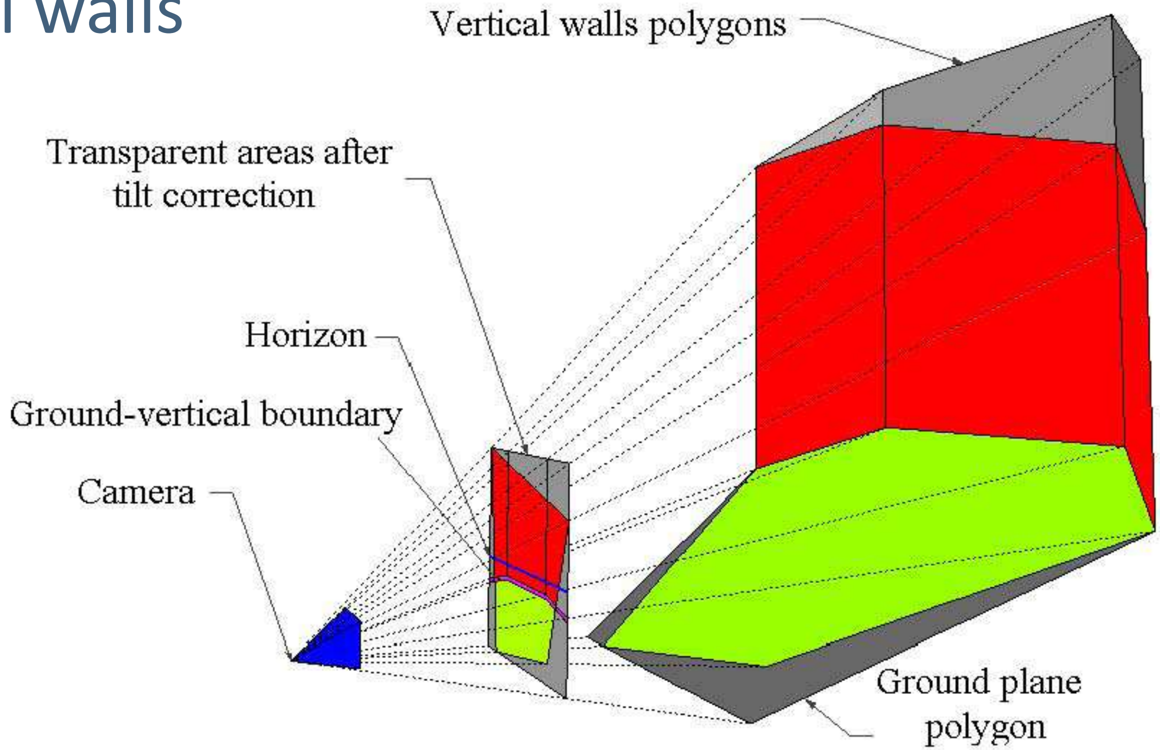
- Application: single-view geometry
  - 3-d reconstruction = recovering 3-d structure of the scene from its projection(s)
  - Single-view 3-d reconstruction is an Ill-posed problem, infinite number of possible solutions



Joint work with Vadim Konushin, Anton Yakubenko, Hwasup Lim, and Anton Konushin, ECCV08

Application: single-view geometry

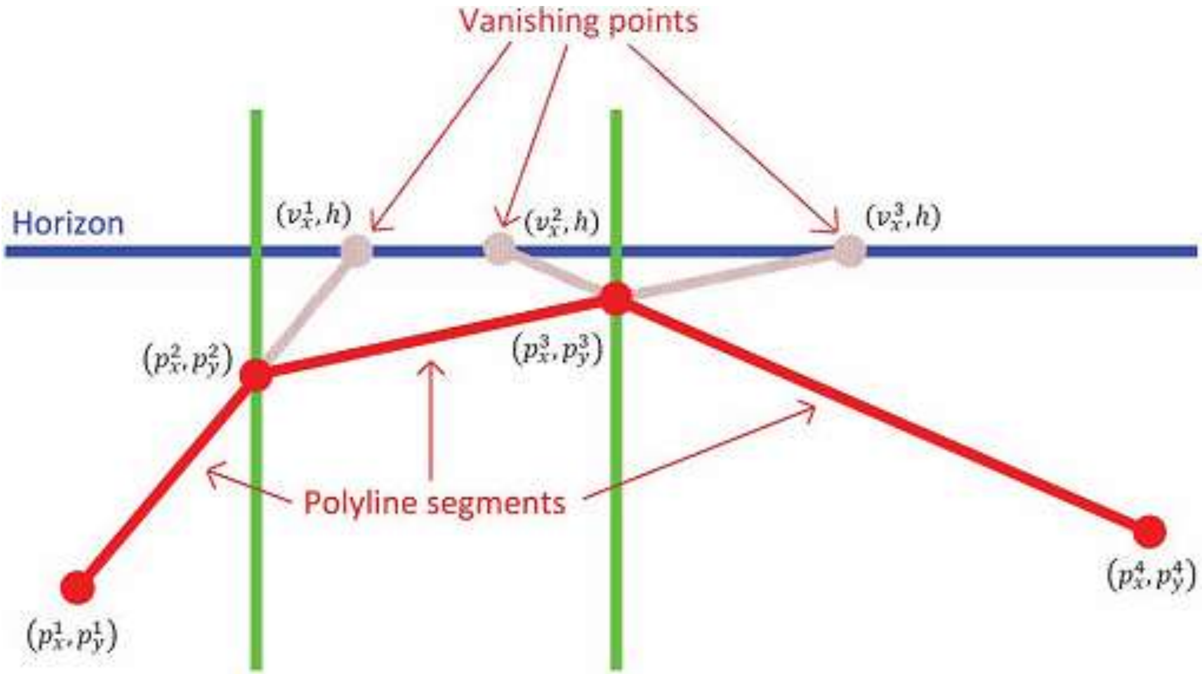
- The structure of 3d model
  - 3d model is composed of ground plane and a number of vertical walls



Joint work with Vadim Konushin, Anton Yakubenko, Hwasup Lim, and Anton Konushin, ECCV08

Application: single-view geometry

- The structure of 3d model



Parameterization:

- *(n-1) X-coordinates of polyline fractures* $p_x^i, (i = 2, \ldots, n)$
- *(n) X-coordinate of VP* $v_x^i, (i = 1, \ldots, n)$
- *Horizon level h*
- *Y-coordinate of polyline left end* p_y^1

Application: single-view geometry

- Demo

single image

3d model

Joint work with Vadim Konushin, Anton Yakubenko, Hwasup Lim, and Anton Konushin, ECCV08

Thank you for attention!

Thanks to Microsoft Research Connections
for the support of creating
courses on computer vision at MSU