Systems Engineering Meets Life Sciences: Review (2nd Lecture)

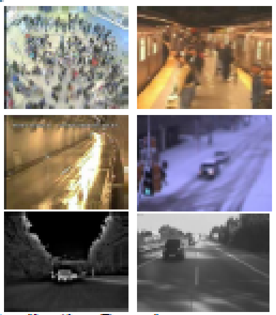
Prepared by:
Prof. Dr. Visvanathan Ramesh

Systems Engineering Methodology Summary

FIAS Frankfurt Institute for Advanced Studies

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

User View Of System



Application Domains:
- Automated Video Surveillance for Safety and Security
- Vision for Autonomous driving – comfort and safety

Task Specification:

e.g. A situational aware system that: parses video from indoor and outdoor environments and generates descriptions of objects, events and is able to alert to a user when event descriptions match user specified queries.

Performance Specs:

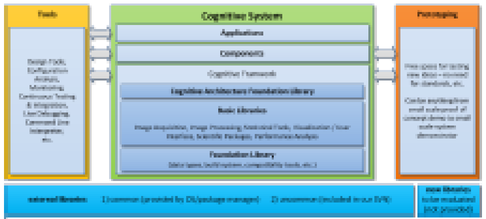
(Functional as well as non-functional):
For each task: probability of error (e.g. probability of miss detection, probablity of false alarm in two class situations)
- Accuracy in Parameter estimates (tolerances)
- Graceful degradation, Self-Diagnosis
- Cost, Computational speed
- Time delay to response, etc.

Modeling View of System

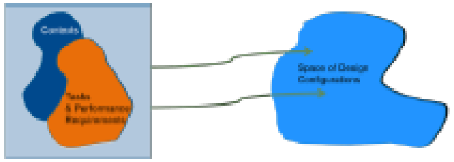
Formal Graphical Models of Context (of various degrees of approximation) along with Physics based Sensor models are linked to systems level analysis of sequential estimators.

Overview: A vision system is complex and composed of multiple stages of execution and is a non-linear estimator from Sensed Images → Semantics. Choice of estimator depends on application requirements and graphical models and involves approximations. Core ideas use in the methodology include:
- Statistical methods for describing population of inputs, quantifying performance and limitations, and to perform inference.
- Use of Probabilistic Contextual Models - Represent application constraints in terms of: priors on camera parameters and projection model, object parameters and dynamics, priors on illumination.
- Develop likelihood models L(features|image) taking into account sensor and feedforward computational process models correctly. Perform Bayesian inference by combining likelihood model and priors.
- Understanding of Vision application space → likelihood model choices and prior choices leads to the building of systems in a rapid and cost effective way.

Implementation View of System

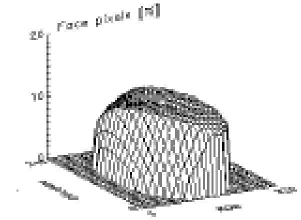


Cognitive Vision Platform for Rapid Prototyping & Validation



"Explore Model Space for Contexts, Task and Performance Requirements, Systematically Map them to Design configurations, Prototype Designs, Validate, Iterate" as in 'Design Thinking'. Extension to new design settings – via re-use of context elements and identification of gaps in models."

Evaluation & Validation of System

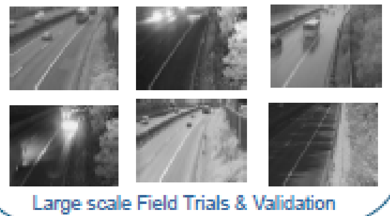


System Performance Model (Example)

Validation. First two lines show the predicted (hat), and experimental (tilde) variances for the tilt angle α. Next two lines correspond to pan angle β.

$\times 10^{-3}$	P1	P2	P3	P4	P5	P6	P7	P8
$\hat{\sigma}^2_{\tan\alpha}$	2.10	2.12	1.57	1.40	1.35	1.31	1.31	1.32
$\tilde{\sigma}^2_{\tan\alpha}$	2.05	2.04	1.60	1.34	1.36	1.32	1.40	1.31
$\hat{\sigma}^2_{\sin\beta}$	28.9	26.1	21.3	17.9	15.3	15.2	18.4	20.1
$\tilde{\sigma}^2_{\sin\beta}$	25.9	24.1	19.5	15.1	14.9	15.0	18.1	19.3

Model Validation (Theory vs Experiment)



Large scale Field Trials & Validation

"Visual Cognition is 'quasi-invariant Indexing' followed by detailed estimation (or deliberation, iteration) – component level research is quite mature, open research is on systems questions involving Cognitive Vision Platform with Continuous Learning and Self-Diagnostics".

Essence of Overall Design Framework: {Application contexts} x {sensor types + configurations} x {questions posed} x {perf specs/requirements} ----> {specific hypotheses generators} + {reasoning / optimization engine}

Past Lectures:

System engineering examples

- Greiffenhagen et al (2001)

Overall theme

(Context,Task,Performance) → Hw plus Sw configuration (hw + programs plus parameters)

- Context, Task, Performance
  - What is context -- Derek hoeim's Book and Slides today (2015)
  - Task - estimation of world state (or parts of it)
  - Performance - bias, variance , accuracy vs speed tradeoff
- What is a Program (Inference Engine)?
  - Program - filters and combinations (feedforward, deep, feedback and recurrent) (ML Literature, Bio-inspired vision literature 2016)
  - Program Design – Model based vs Data Driven, or Hybrid combinations
- What about Performance Characterization of Designs? (Ramesh, 1995)

Next Lectures:

- Recap – Greiffenhagen Thesis / Systems Engineering Methodology
- Model-Based Recognition Overview (Mann, 1996, Dissertation)
- What is Context ? (Slides based on Derek Hoeim)
- Link to Systems Engineering Methodology

Dual camera approach:

- constantly monitor large area of interest (Shree Nayar's OmniCamera)
- simultaneously: high resolution images of faces (e.g. face recognition)

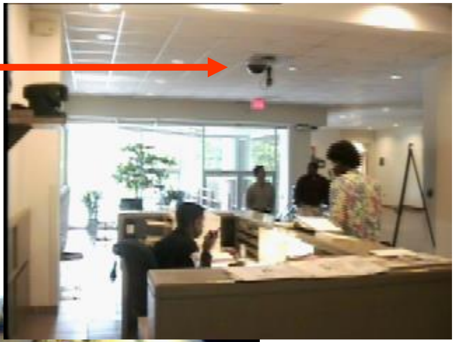
⇒ Monitoring / surveillance

⇒ Trigger alarms (sensitive areas)

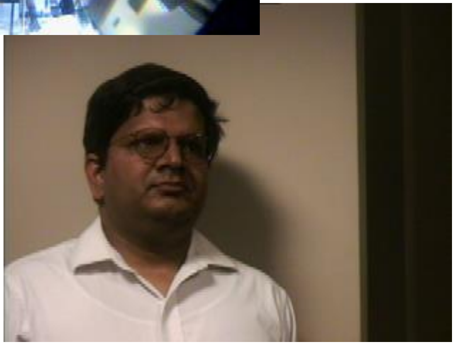
⇒ Log information (time, location, face)

⇒ Post-processing (recognition/data-base)

Lobby scene



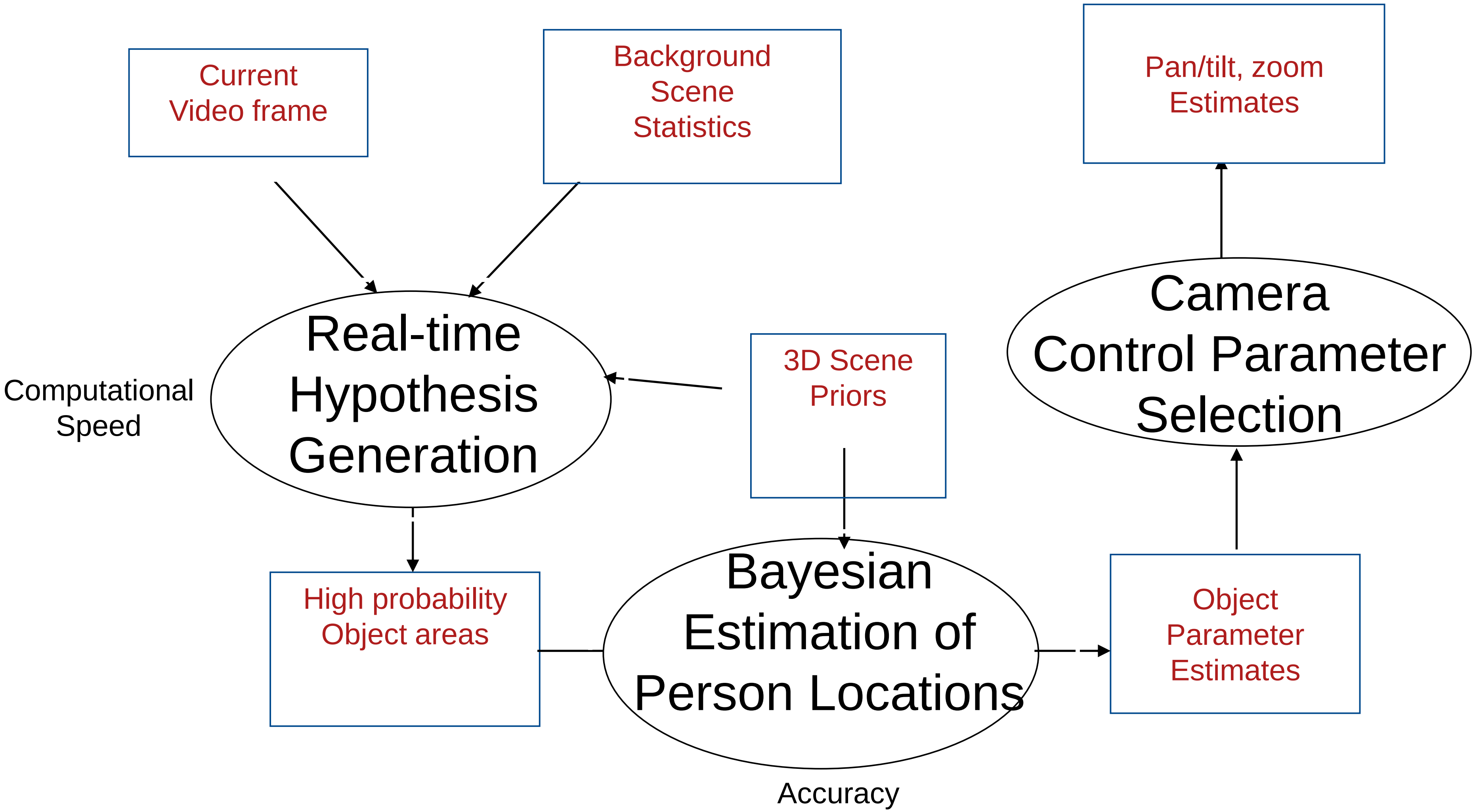
Omni-view



Foveal-view

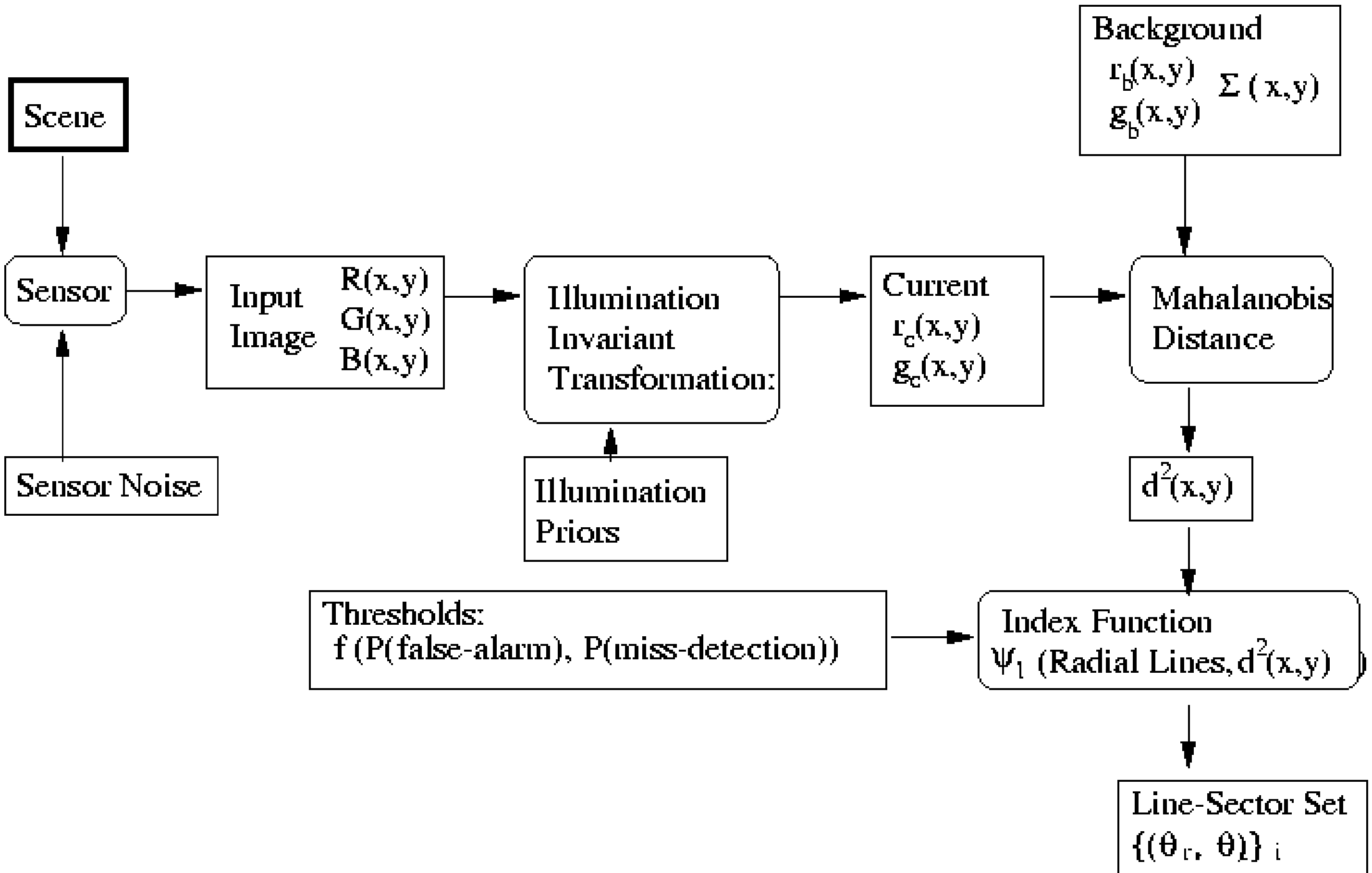
Bayesian Real-time System Example

5



(VR)

Indexing Step Modules (1)



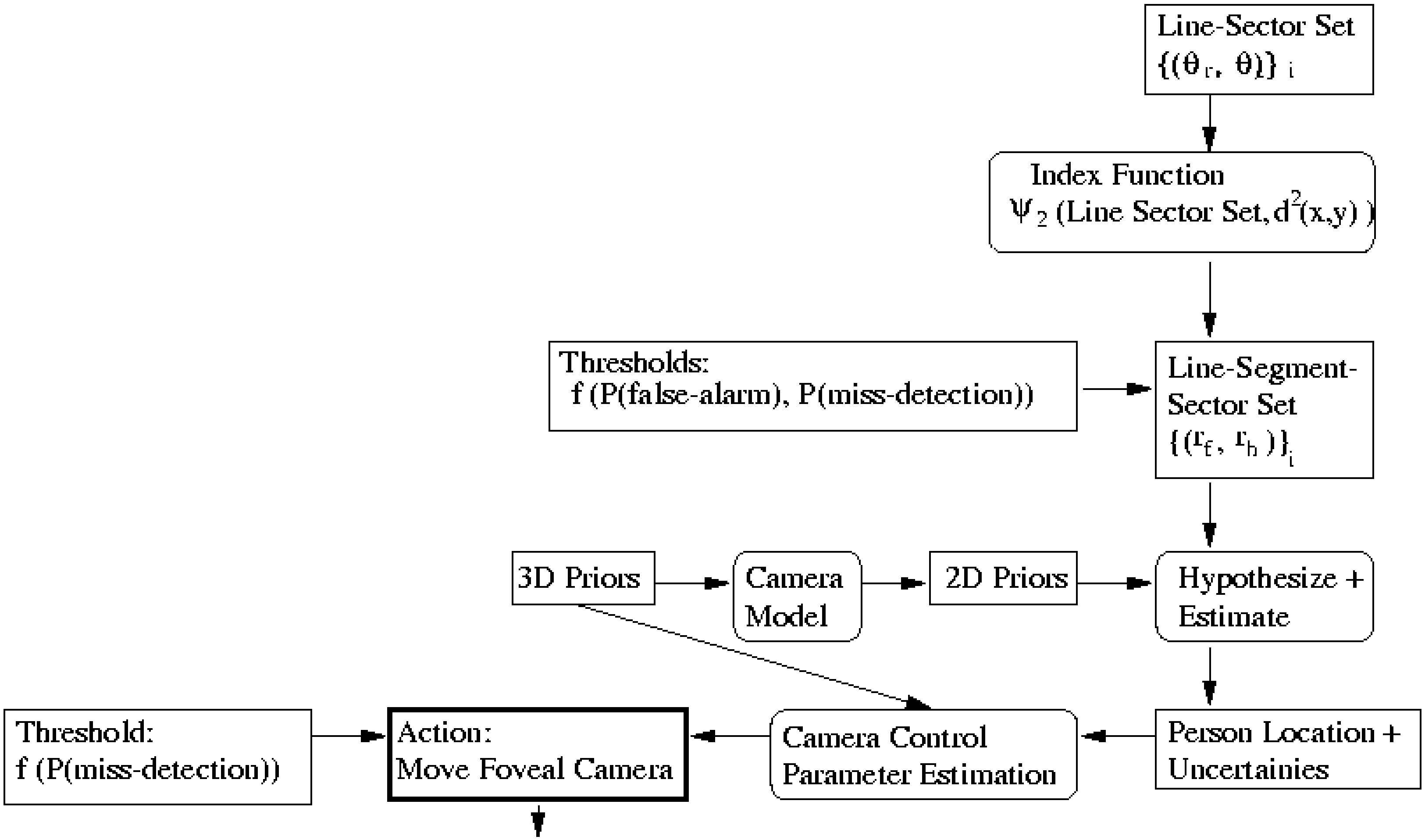
➢ models influence algorithm design in each module

(VR)

Location Estimate + camera control

FIAS Frankfurt Institute for Advanced Studies

GOETHE UNIVERSITÄT FRANKFURT AM MAIN



Models of Algorithm Steps

Abstract Model for Algorithm

Estimation	Transformation	Mapping
Illumination Invariance	T1	$\begin{pmatrix} \hat{R} \\ \hat{G} \\ \hat{B} \end{pmatrix} \longrightarrow \begin{pmatrix} \hat{r} \\ \hat{g} \end{pmatrix}$
Probability of Background	T2	$\left(\begin{pmatrix} \hat{r} \\ \hat{g} \end{pmatrix}_c \times \begin{pmatrix} \hat{r} \\ \hat{g} \end{pmatrix}_b, \mathbf{\Sigma_{r,g}}\right) \longrightarrow \hat{d}^2$
Indexing 1	T3	$\hat{d}^2(r,\theta) \longrightarrow \hat{M}_\theta \longrightarrow (\theta_l,\theta_r)$ with $\hat{M}_\theta = \sum_r \hat{d}^2(r,\theta), r \in \{0, r_m\}$
Feature Estimation	T4	$\hat{d}^2(x,y) \times (\theta_l,\theta_r) \longrightarrow \hat{M}_{\overline{r,\theta}}$ with $\hat{M}_{\overline{r,\theta}} = \sum_{(x,y)} \hat{d}^2(x,y)$ $\forall (x,y) \mid (x_c - x)\sin(\theta+\pi/2) - (y_c - y)\cos(\theta+\pi/2) = r,$ with $r \in \{r_f, r_h\}, \quad \theta \in \{\theta_l, \theta_r\}$
Location estimation 2D	T5	$\hat{M}_r \longrightarrow \begin{pmatrix} \hat{r}_p \\ \hat{\vartheta} \end{pmatrix}$
Location estimation 3D	T6	$\hat{r}_p \longrightarrow \hat{R}_p$
Pan/Tilt Estimation	T7	$\begin{pmatrix} \hat{R}_p \\ \hat{\vartheta} \end{pmatrix} \longrightarrow \begin{pmatrix} \tan(\hat{\alpha}) \\ \sin(\hat{\beta}) \end{pmatrix}$
Zoom Setting	T8	$\begin{pmatrix} \tan(\hat{\alpha}) \\ \sin(\hat{\beta}) \end{pmatrix} \longrightarrow$ zoom z

Systems Analysis

Statistical Analysis

Transformation	I/P distribution	O/P distribution	Type of Propagation
T1	$N\left(\begin{pmatrix}R\\G\\B\end{pmatrix}, \mathbf{diag}\begin{pmatrix}\sigma^2_{\bar{R}}\\\sigma^2_{\bar{G}}\\\sigma^2_{\bar{B}}\end{pmatrix}\right)$	$N\left(\begin{pmatrix}r\\g\end{pmatrix}, \mathbf{\Sigma}_{\mathbf{r},\mathbf{g}}\right)$	covariance propagation
T2	$\mathrm{N}\left(\begin{pmatrix}r\\g\end{pmatrix}, \mathbf{\Sigma}_{\mathbf{r},\mathbf{g}}\right)$	Background pixel: $\chi^2_2(0)$ Object pixel: $\chi^2_2(c), c \neq 0$	distribution propagation
T3	$\chi^2_2(c), c \in [0 \ldots \infty]$	$(r_m - k)\chi^2_{2(r_m - k)}(0) + k\chi^2_{2k}(c)$	distribution propagation
T4	$\chi^2_2(c), c \in [0 \ldots \infty]$	Background pixel: $\chi^2_{2s}(0)$ Object pixel: $\chi^2_{2s}(c), c \neq 0$	distribution propagation
T5	$n_b\chi^2_{2n_b}(0) + n_o\chi^2_{2n_o}(c)$	$N\left(\begin{pmatrix}r_p\\\vartheta\end{pmatrix}, \begin{pmatrix}\sigma^2_{\bar{r}_p}\\\sigma^2_{\bar{\vartheta}}\end{pmatrix}\right)$	bootstrap
T6	$N\left(r_p, \sigma^2_{\bar{r}_p}\right)$	$N\left(R_p, \sigma^2_{\bar{R}_p}\right)$	covariance propagation
T7	$N\left(\begin{pmatrix}R_p\\\vartheta\end{pmatrix}, \begin{pmatrix}\sigma^2_{\bar{R}_p}\\\sigma^2_{\bar{\vartheta}}\end{pmatrix}\right)$	$N\left(\begin{pmatrix}\tan(\alpha)\\\sin(\beta)\end{pmatrix}, \begin{pmatrix}\sigma^2_{\tan(\bar{\alpha})}\\\sigma^2_{\sin(\bar{\beta})}\end{pmatrix}\right)$	covariance propagation
T8	$N\left(\begin{pmatrix}\tan(\alpha)\\\sin(\beta)\end{pmatrix}, \begin{pmatrix}\sigma^2_{\tan(\bar{\alpha})}\\\sigma^2_{\sin(\bar{\beta})}\end{pmatrix}\right)$	$N\left(z, \sigma^2_z\right)$	covariance propagation

Parameter Optimization

10

CRITERION FUNCTIONS AND PRIORS INFLUENCING THE CHOICE OF TRANSFORMS

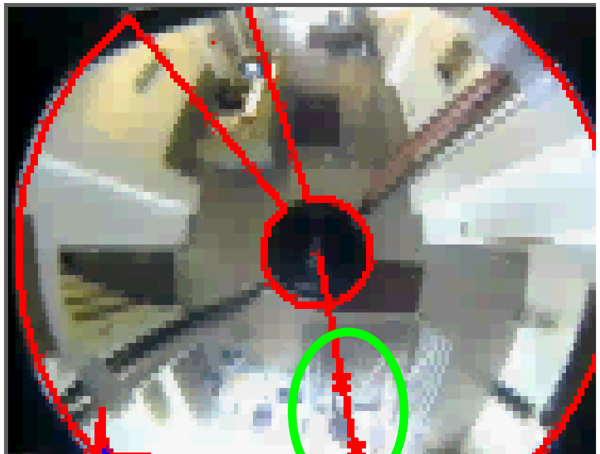
Transformation	Threshold/Criterion Fct.	Priors
T1	n/a	P(Illumination)
T2	n/a	n/a
T3	P(missing hypothesis), P(false hypothesis)($on\,object level$)	P(Proj. Geometry), P(Object Height, Location)
T4	n/a	P(Object Height)
T5	n/a	P(Object Radius)
T6	n/a	P(Geometry)
T7	n/a	P(Object Pose)
T8	P(head in foveal frame)	P(Object Head Size)

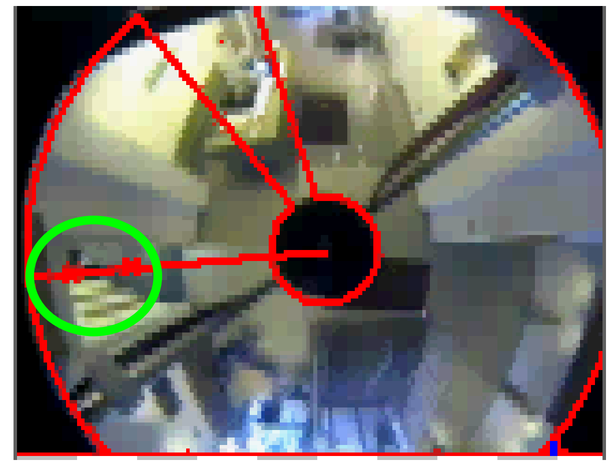
(VR)

System Performance: Over Continuous Changes over 24h

FIAS Frankfurt Institute for Advanced Studies

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

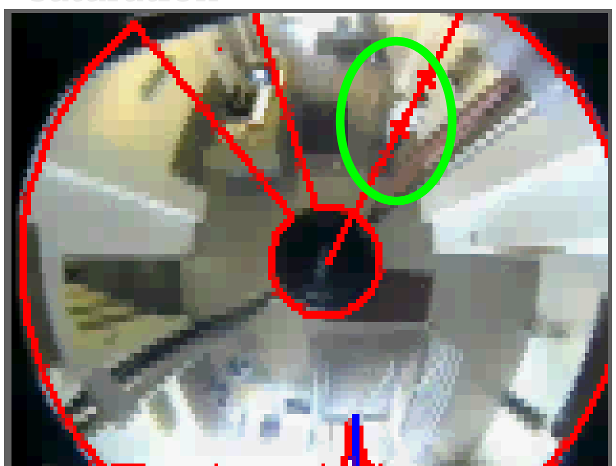


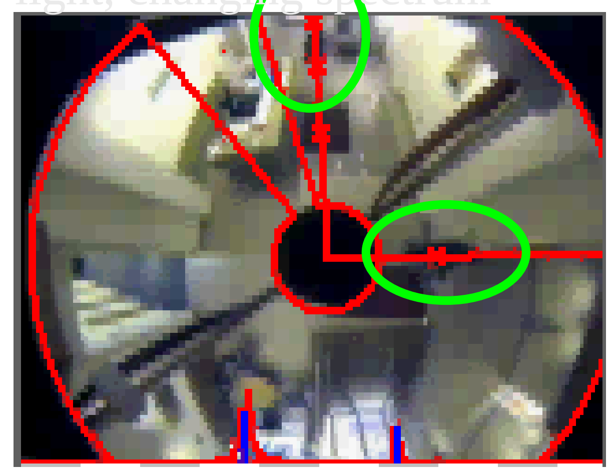


__Day__: natural+artificial light; saturation

__Night__: artificial light low contrast regions

__Morning/afternoon__: mixed light, changing spectrum





Summary

Showed how to systematically design a complete real-time vision system that is predictable, robust & guarantees performance within pre-defined bounds in a real-world setting

⇒ Decompose system into modules
⇒ Statistical modeling and analysis of each module
⇒ Propagation of uncertainties from input data to final output
⇒ Complete engineering cycle: design, analysis, validation, test
⇒ Optimize performance given data and task
⇒ Optimize setup / camera position

Demonstrated how to evolve an existing system incrementally to meet added requirements

⇒ without redesign of existing modules
⇒ fusion of existing and 3rd party module combining strength of both
⇒ maintaining analysis of prior system valid

⇒ **Built working system – *stable* and in lobby at SCR, Princeton, *in use***

Context - Overview

Visvanathan Ramesh

***Uses Sources from: Derek Hoeim, Mann & Binford, Systems engineering work**

Context and 3D Scenes

Context in Recognition

Objects usually are surrounded by a scene that can provide context in the form of nearby objects, surfaces, scene category, geometry, etc.

Context provides clues for functi

What is this?

These examples from Antonio Torralba

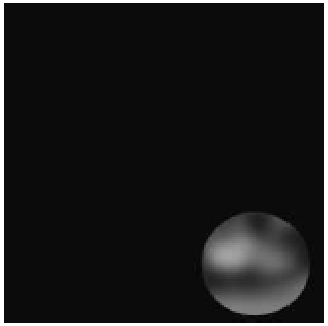
Context provides clues for function

What is this?

Now can you tell?

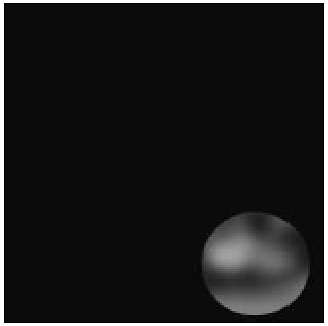
Sometimes context is *the* major component of recognition

What is this?



Sometimes context is *the* major component of recognition

What is this?



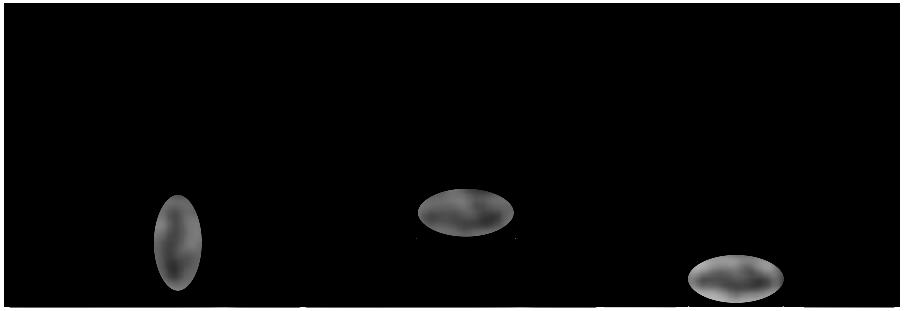
Now can you tell?

More Low-Res

What are these blobs?

More Low-Res

The same pixels! (a car)



There are many types of context

Local pixels

- window, surround, image neighborhood, object boundary/shape, global image statistics

2D Scene Gist

- global image statistics

3D Geometric

- 3D scene layout, support surface, surface orientations, occlusions, contact points, etc.

Semantic

- event/activity depicted, scene category, objects present in the scene and their spatial extents, keywords

Photogrammetric

- camera height orientation, focal length, lens distorition, radiometric, response function

Illumination

- sun direction, sky color, cloud cover, shadow contrast, etc.

Geographic

- GPS location, terrain type, land use category, elevation, population density, etc.

Temporal

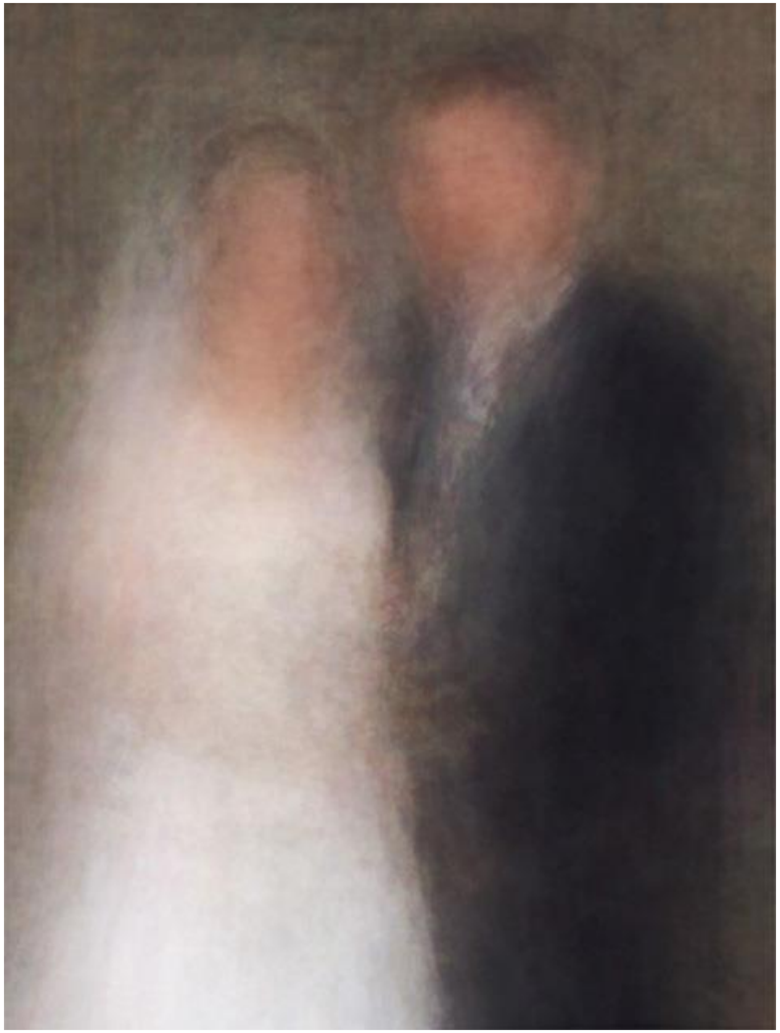
- nearby frames of video, photos taken at similar times, videos of similar scenes, time of capture

Cultural

- photographer bias, dataset selection bias, visual cliches, etc.

from Divvala et al. CVPR 2009

Cultural context



Jason Salavon: http://salavon.com/SpecialMoments/Newlyweds.shtml

Cultural context

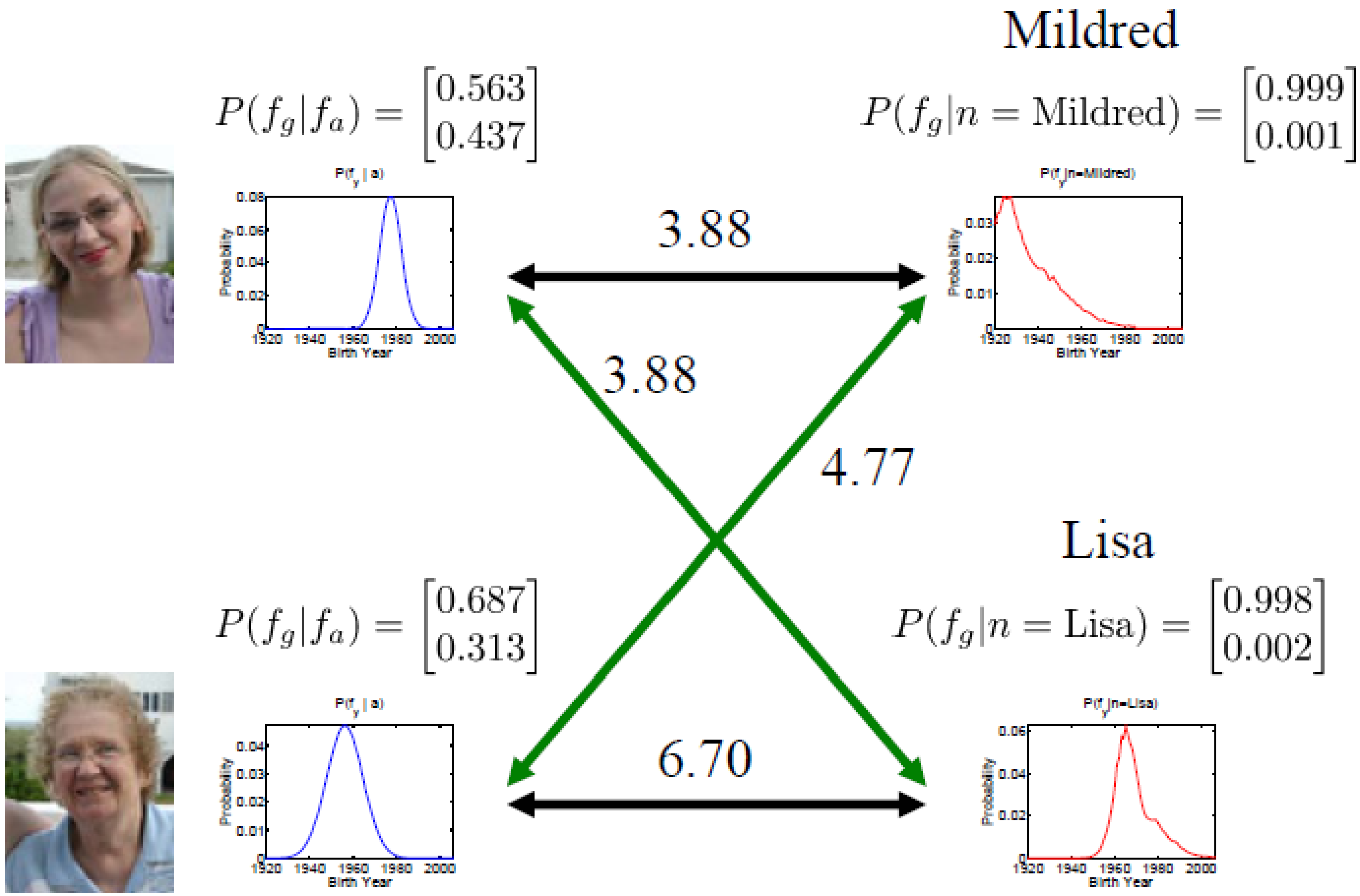
“Mildred and Lisa”: Who is Mildred? Who is Lisa?

Andrew Gallagher: http://chenlab.ece.cornell.edu/people/Andy/projectpage_names.html

Cultural context

Age given Appearance

Age given Name



Andrew Gallagher: http://chenlab.ece.cornell.edu/people/Andy/projectpage_names.html

Spatial layout is especially important

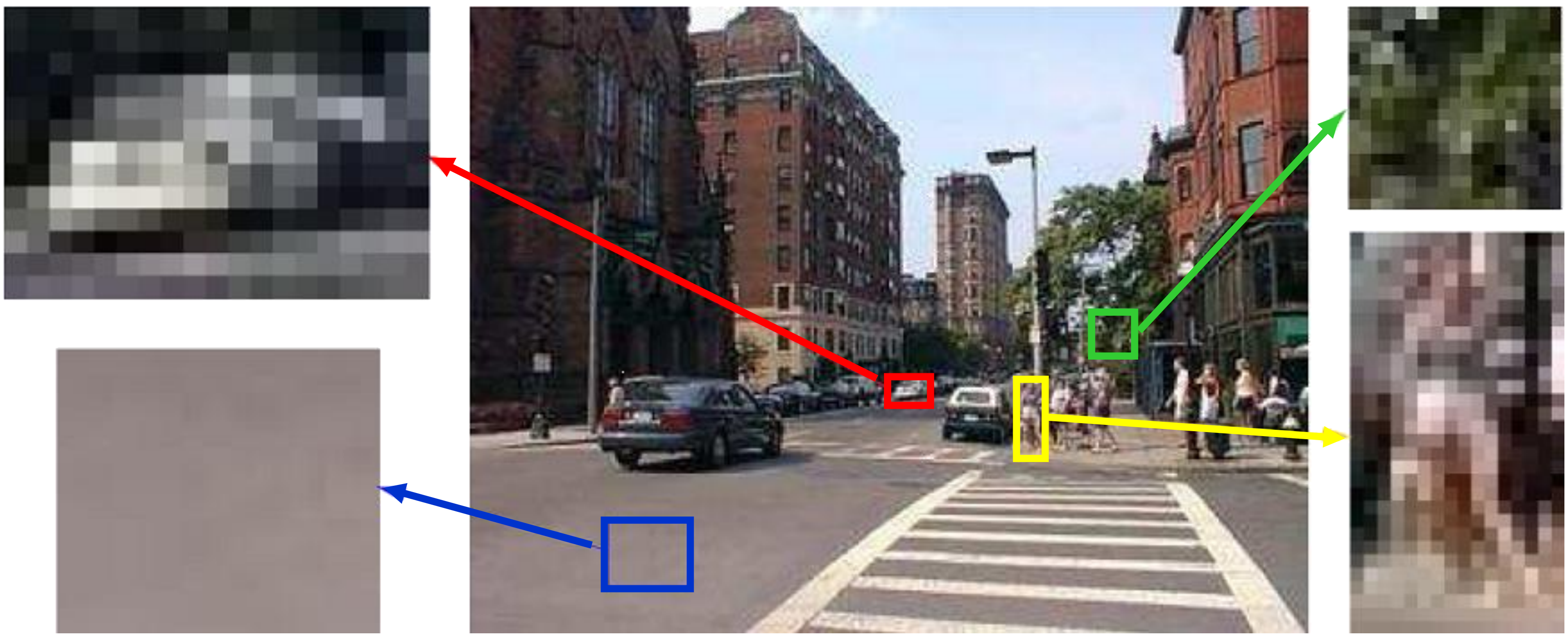
1. **Context for recognition**

Spatial layout is especially important

FIAS Frankfurt Institute for Advanced Studies

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

1. **Context for recognition**



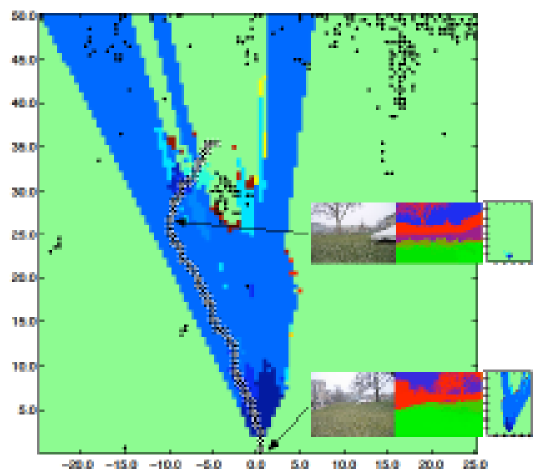
Spatial layout is especially important

1. **Context for recognition**
2. **Scene understanding**

Spatial layout is especially important

1. **Context for recognition**
2. **Scene understanding**
3. **Many direct applications**
    a) Assisted driving
    b) Robot navigation/interaction
    c) 2D to 3D conversion for 3D TV
    d) Object insertion

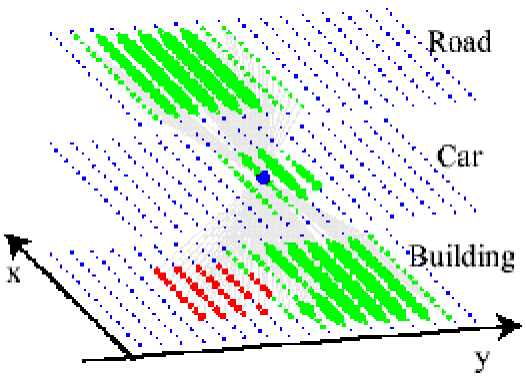
3D Reconstruction: Input, Mesh, Novel View



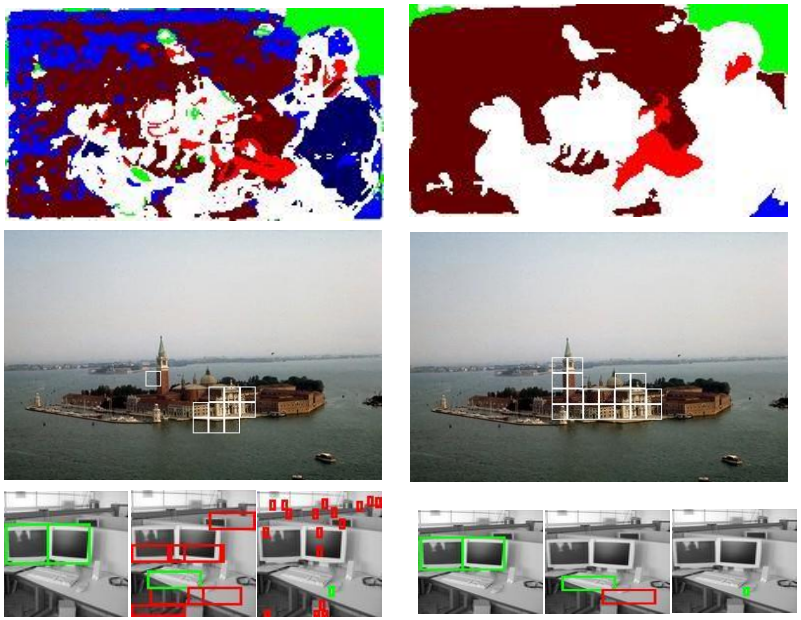
Robot Navigation: Path Planning

Spatial Layout: 2D vs. 3D?

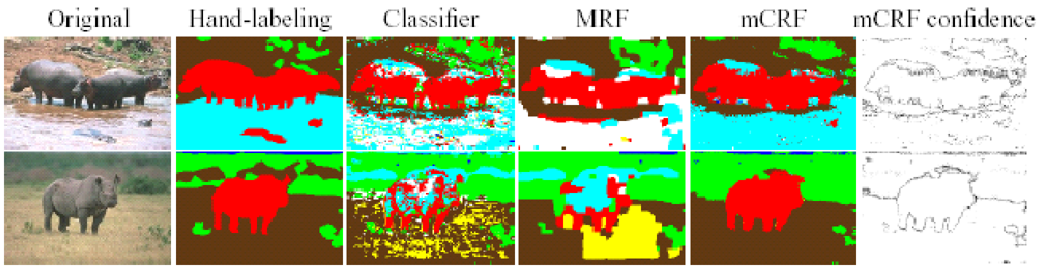
Context in Image Space



[Torralba Murphy Freeman 2004]



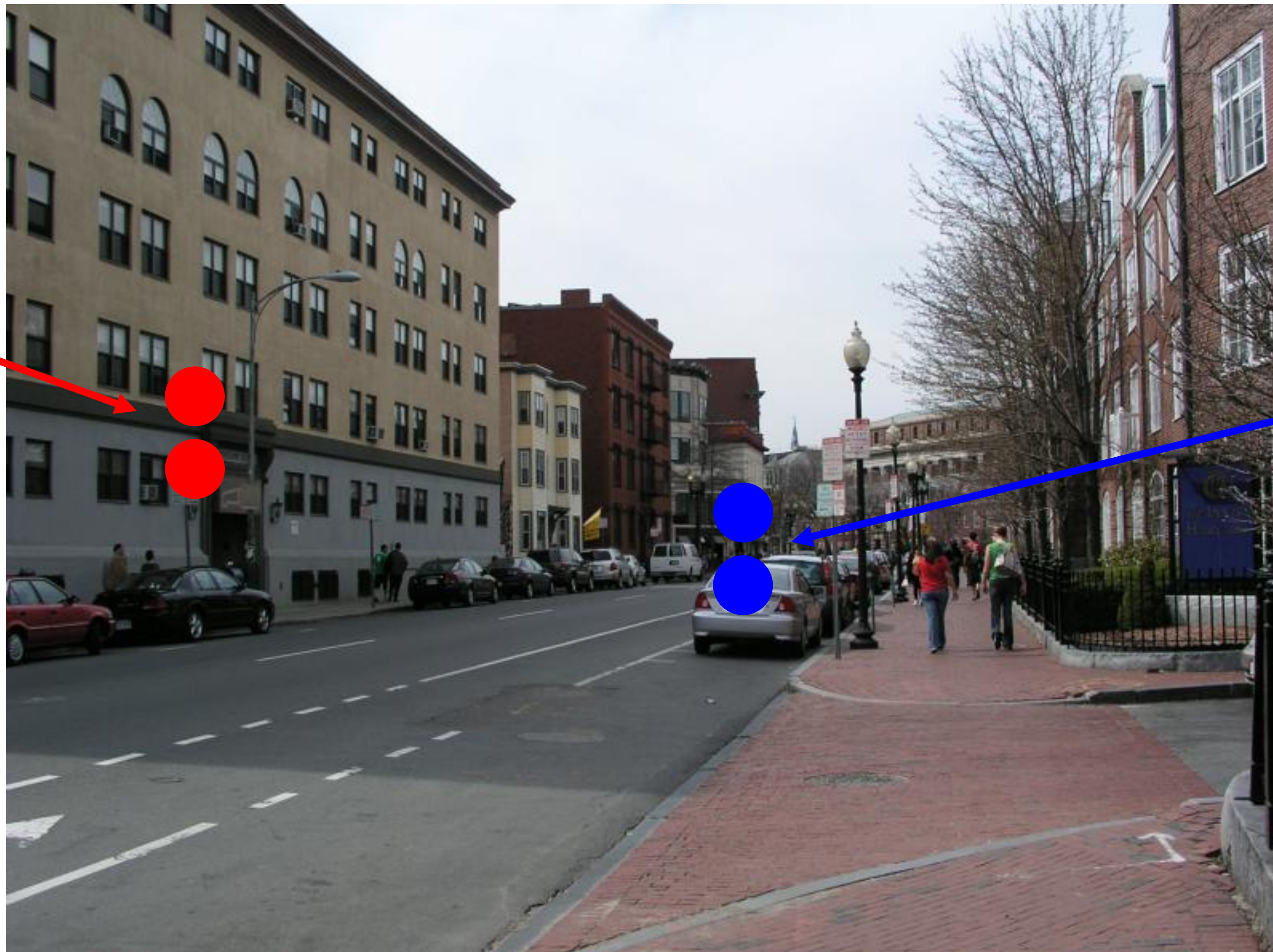
[Kumar Hebert 2005]



[He Zemel Cerreira-Perpiñán 2004]

But object relations are in 3D…

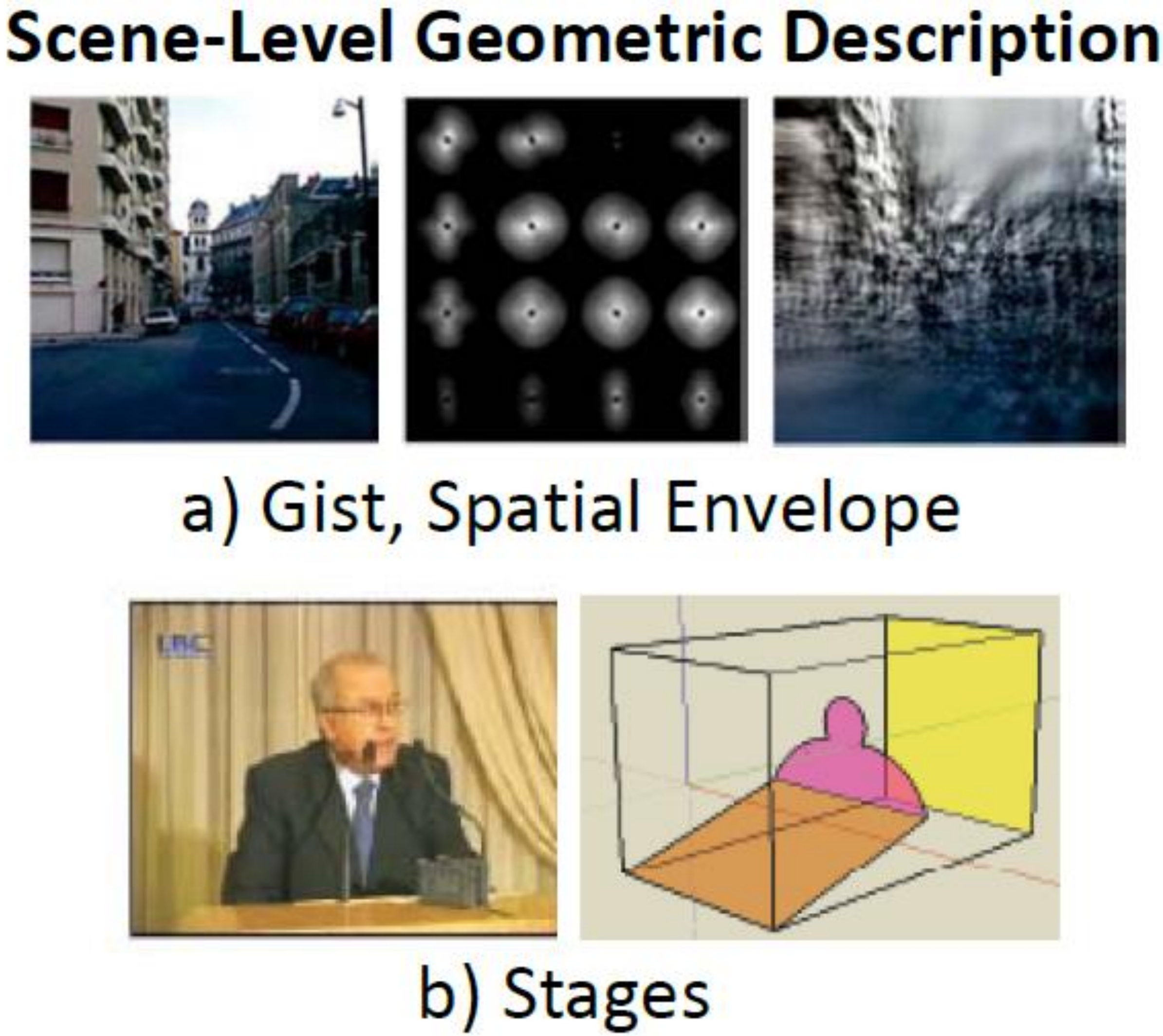
Close



Not Close

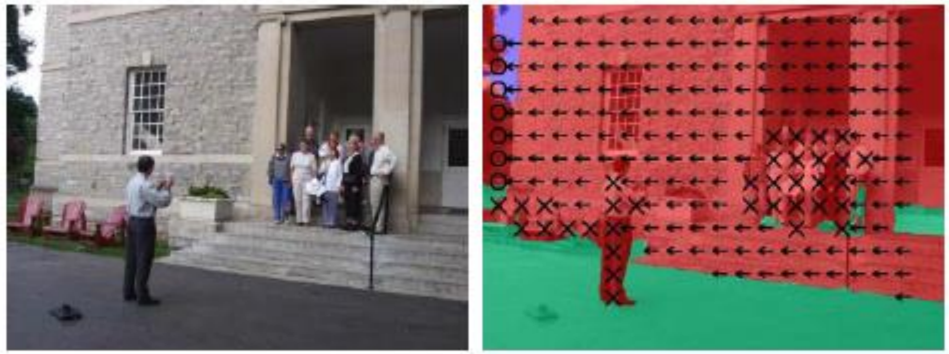
How to represent scene space?

Wide variety of possible representations



Figs from Hoiem - Savarese 2011 book

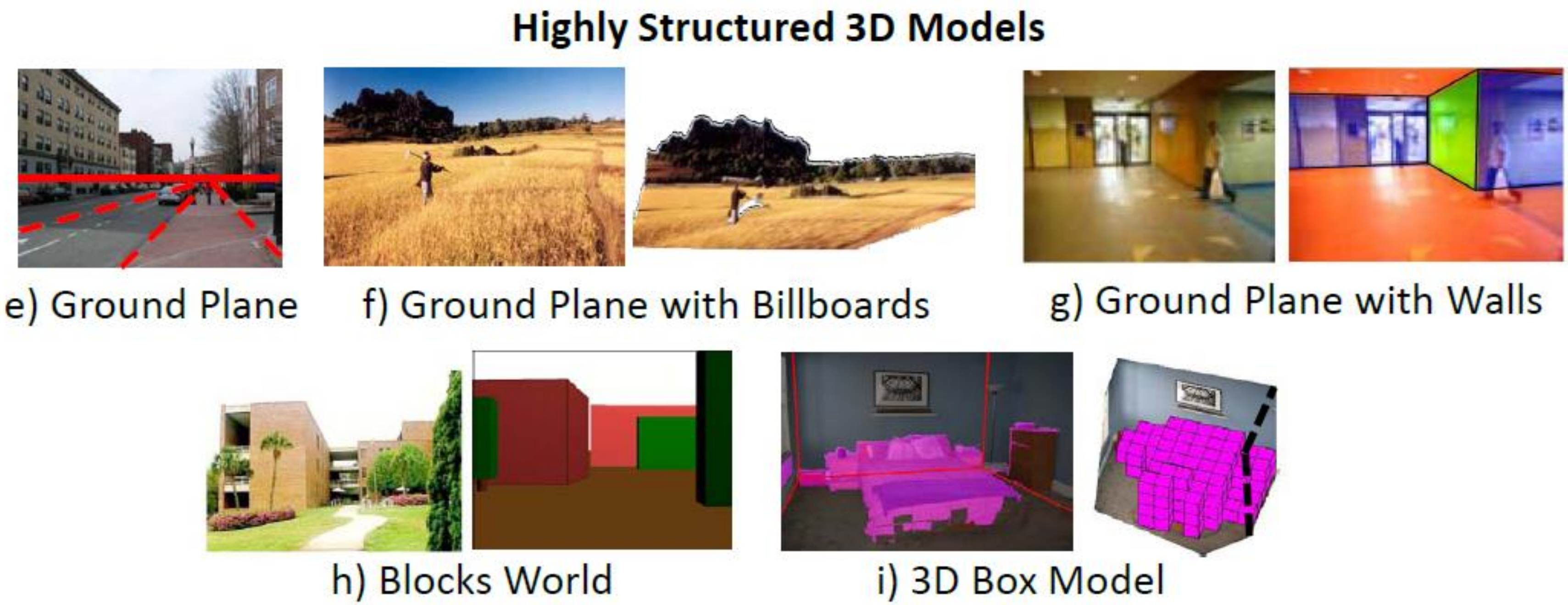
Retinotopic Maps



c) Geometric Context

d) Depth Maps

Figs from Hoiem - Savarese 2011 book



Figs from Hoiem - Savarese 2011 book

Key Trade-offs

Level of detail: rough “gist”, or detailed point cloud?

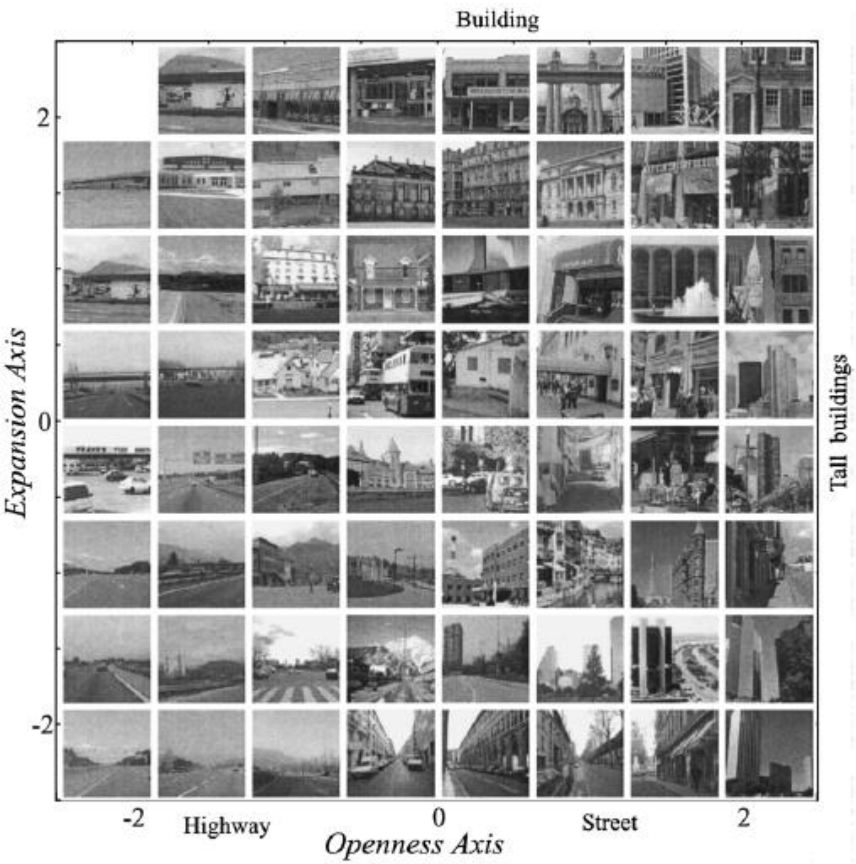
- Precision vs. accuracy
- Difficulty of inference

Abstraction: depth at each pixel, or ground planes and walls?

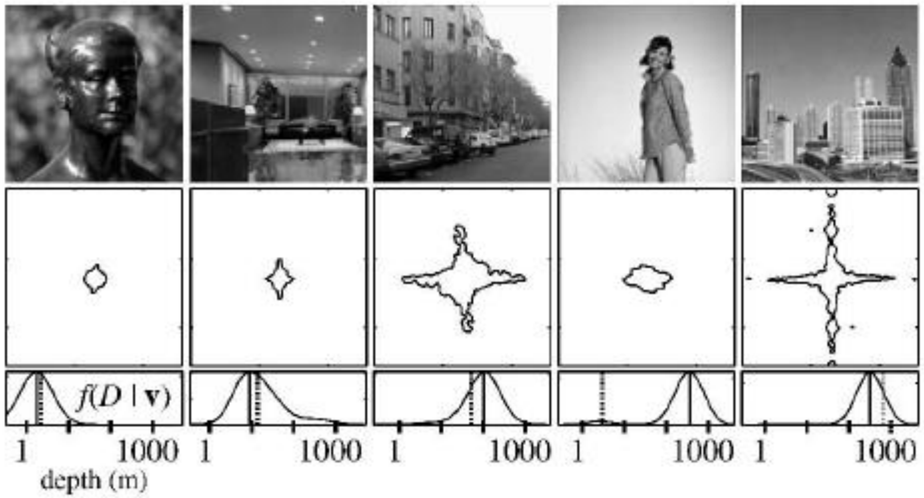
- What is it for: e.g., metric reconstruction vs. navigation

Low detail, Low/Med abstraction

Holistic Scene Space: “Gist”



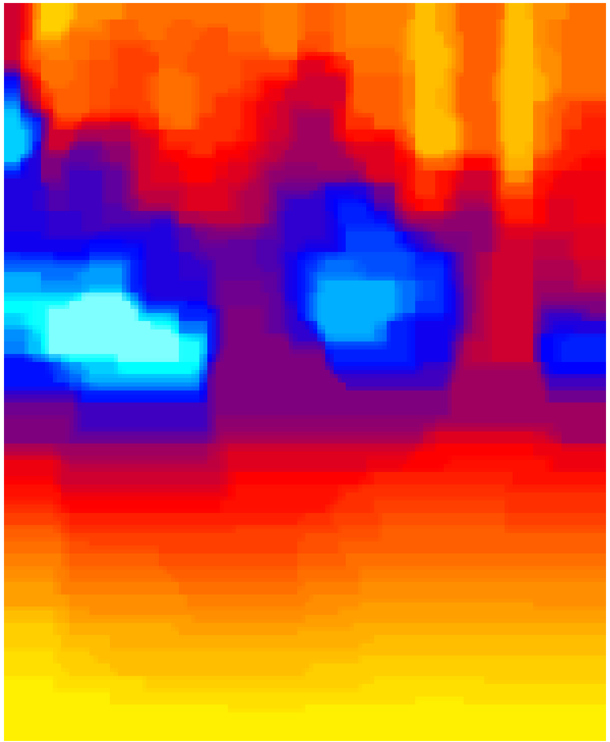
Oliva & Torralba 2001



Torralba & Oliva 2002

High detail, Low abstraction

Depth Map



Saxena, Chung & Ng 2005, 2007

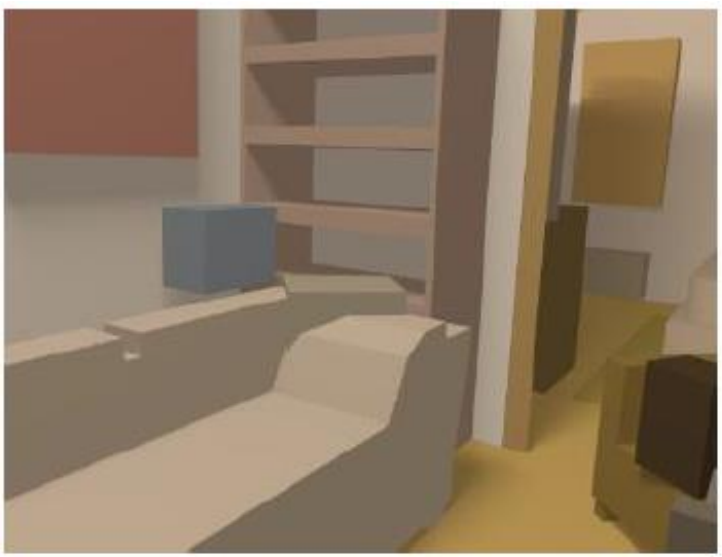
Medium detail, High abstraction

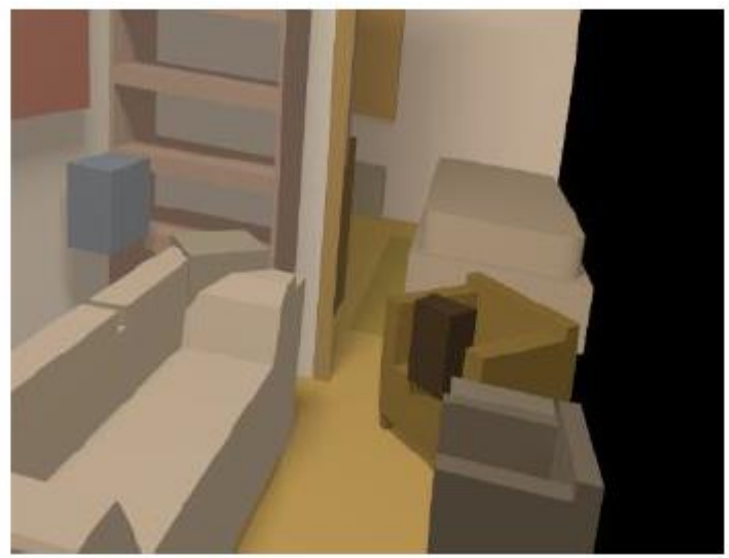
Room as a Box

Hedau Hoiem Forsyth 2009

Med-High detail, High abstraction

Complete 3D Layout





Guo Zou Hoiem 2015

How do we Specify Context?

Representation Choices ?

Probability Models for Patterns

- Graphical Models, Bayesian Networks

Role of Learning

Examples of spatial layout estimation

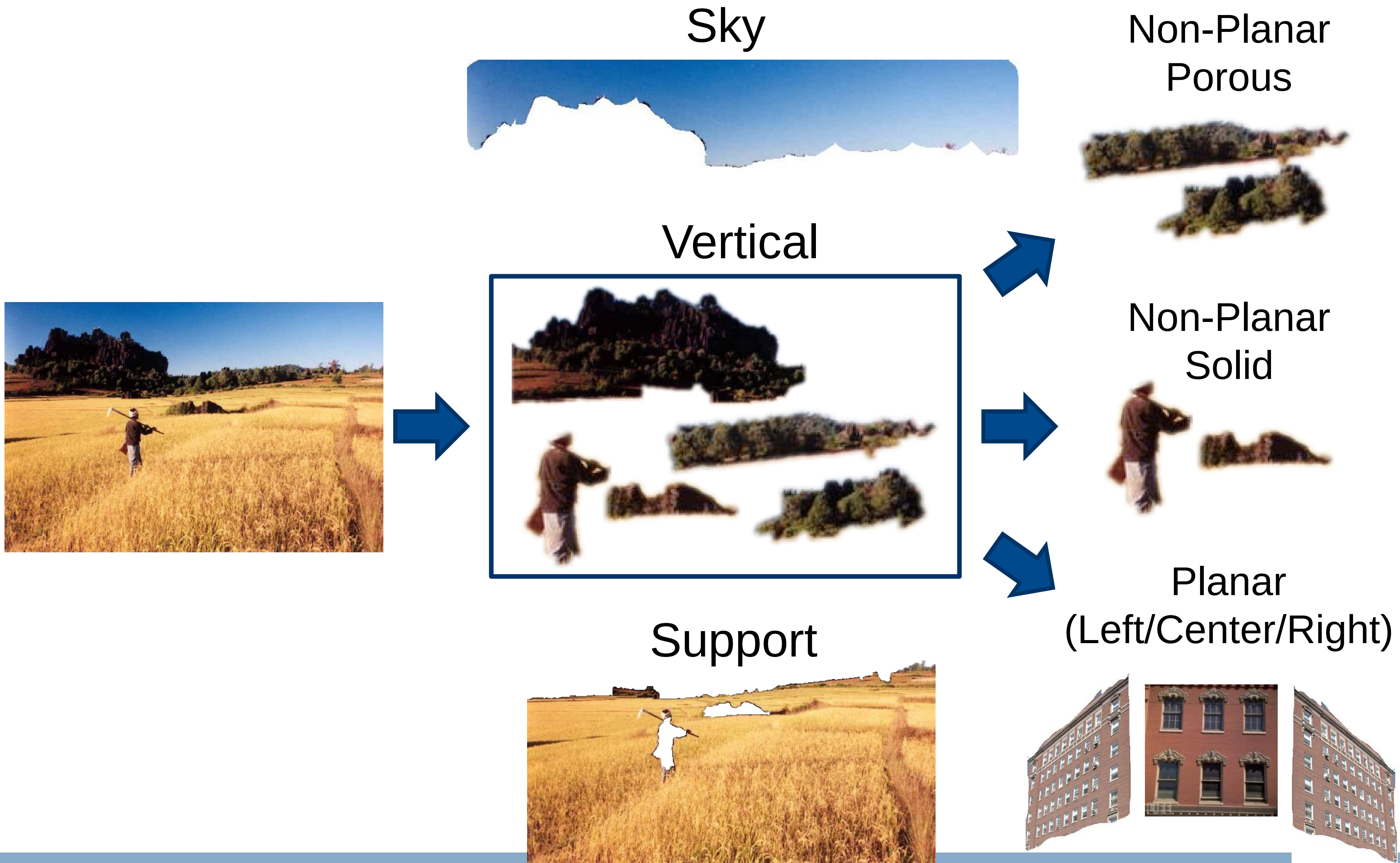
Surface layout

- Application to 3D reconstruction

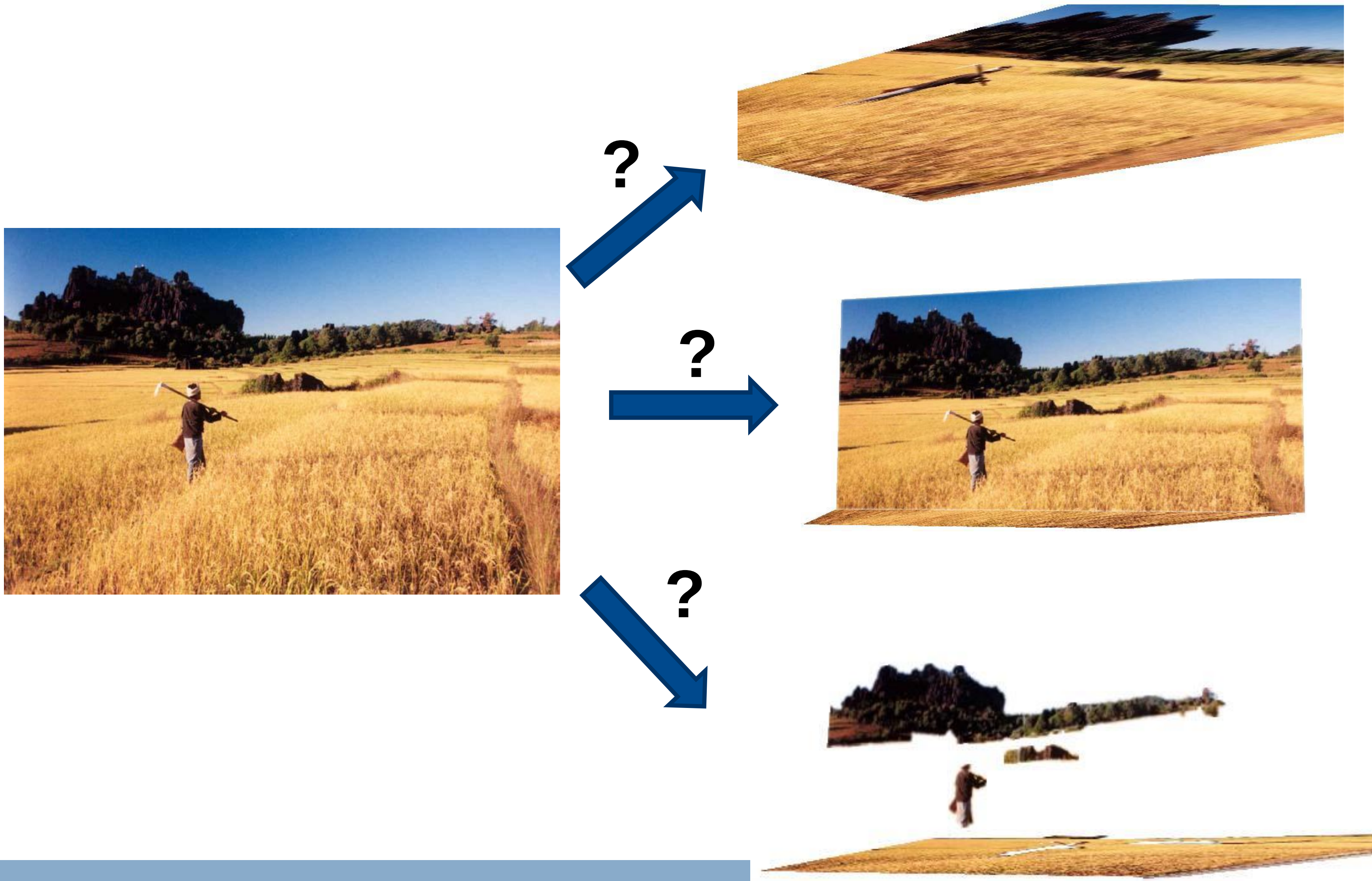
The room as a box

- Application to object recognition

Surface Layout: describe
3D surfaces with geometric classes

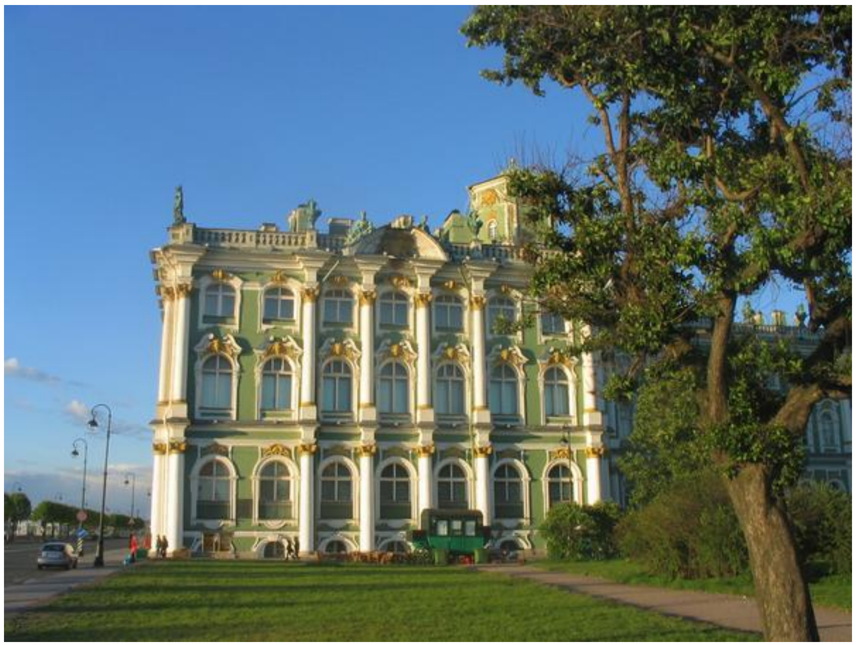


The challenge



Our World is Structured

Abstract World



Our World

Image Credit (left): F. Cunin
and M.J. Sailor, UCSD

Learn the Structure of the World

Training Images

Infer the most likely interpretation

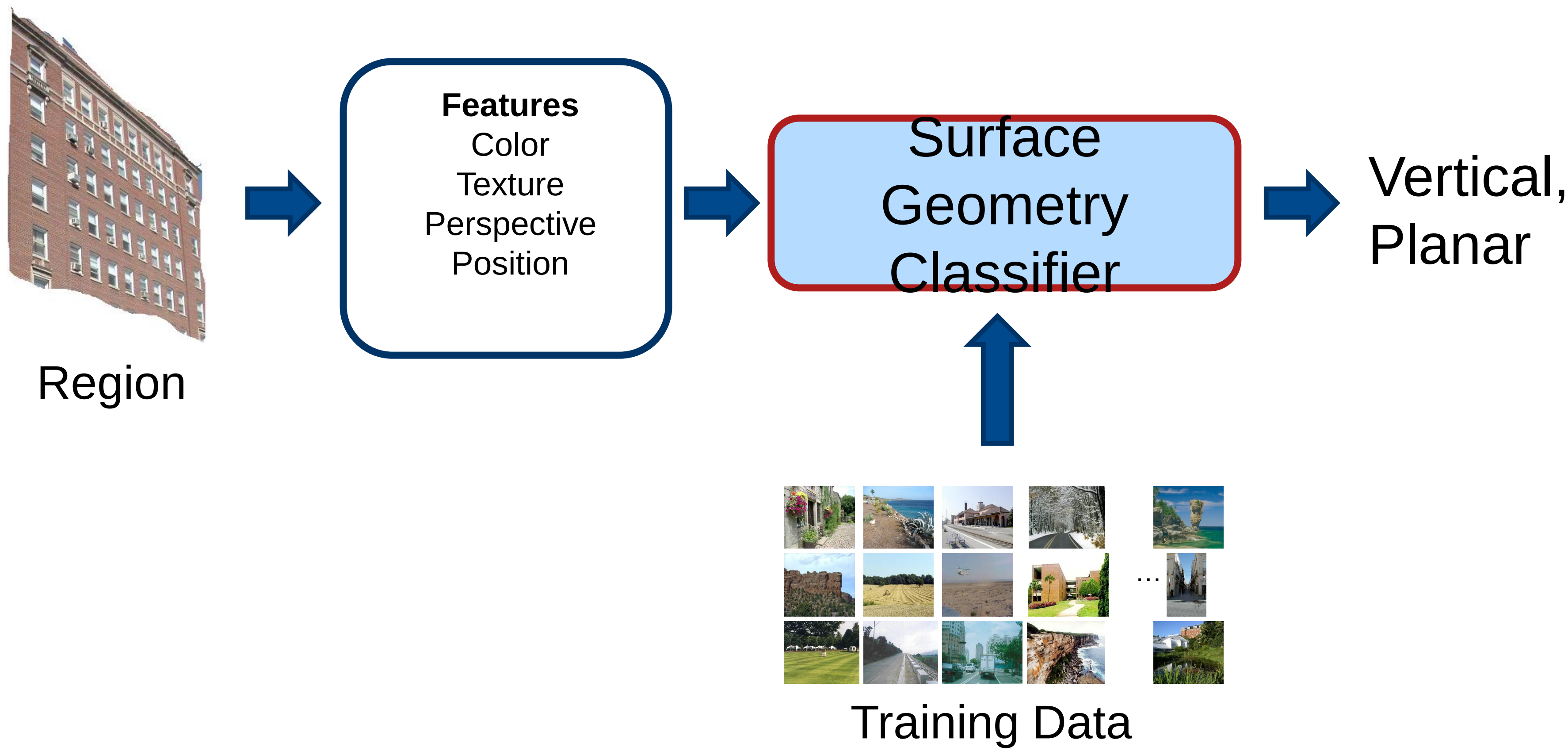
FIAS Frankfurt Institute for Advanced Studies

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Unlikely

Likely

Geometry estimation as recognition



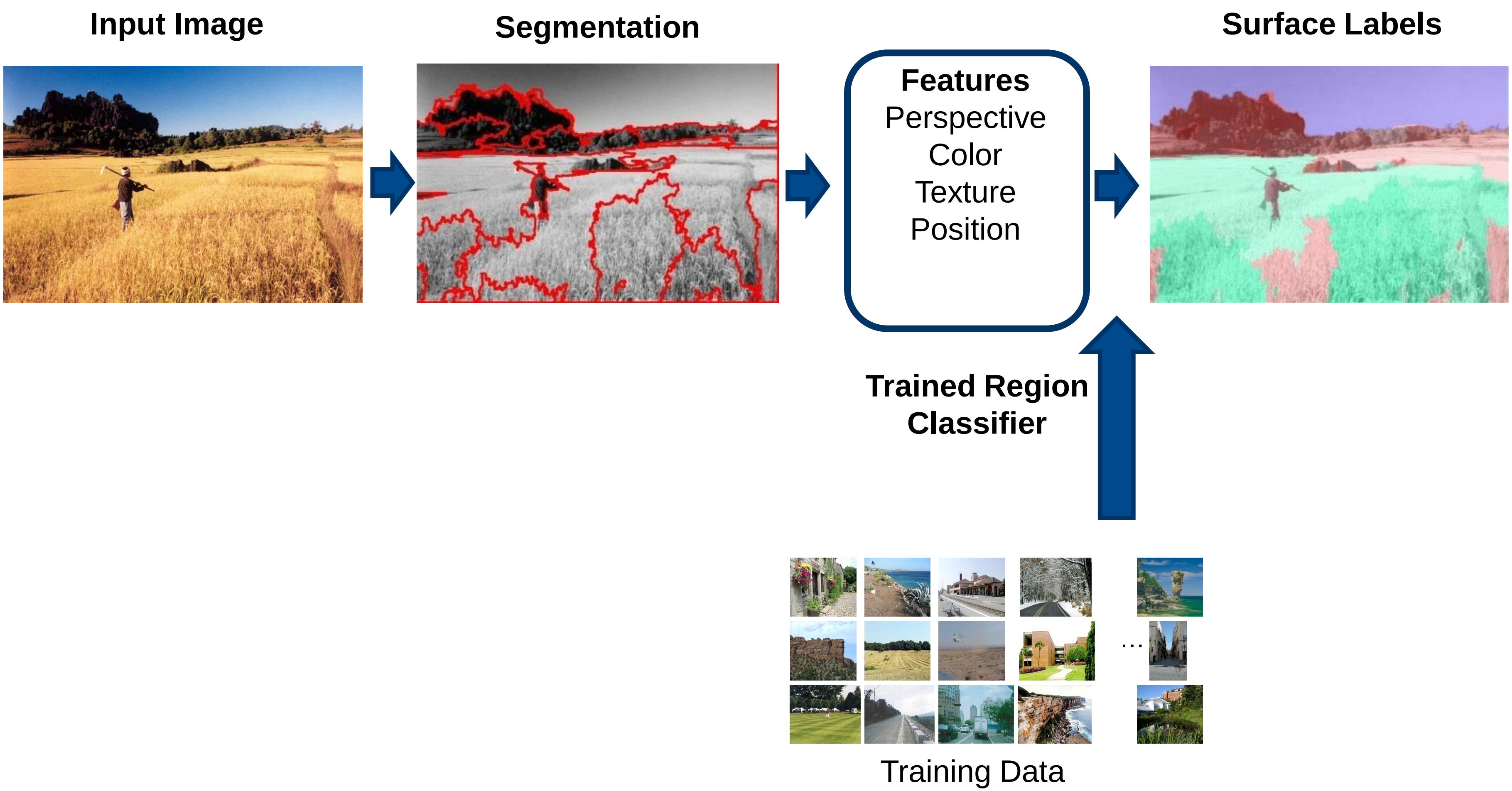
Use a variety of image cues

Vanishing points, lines

Color, texture, image location

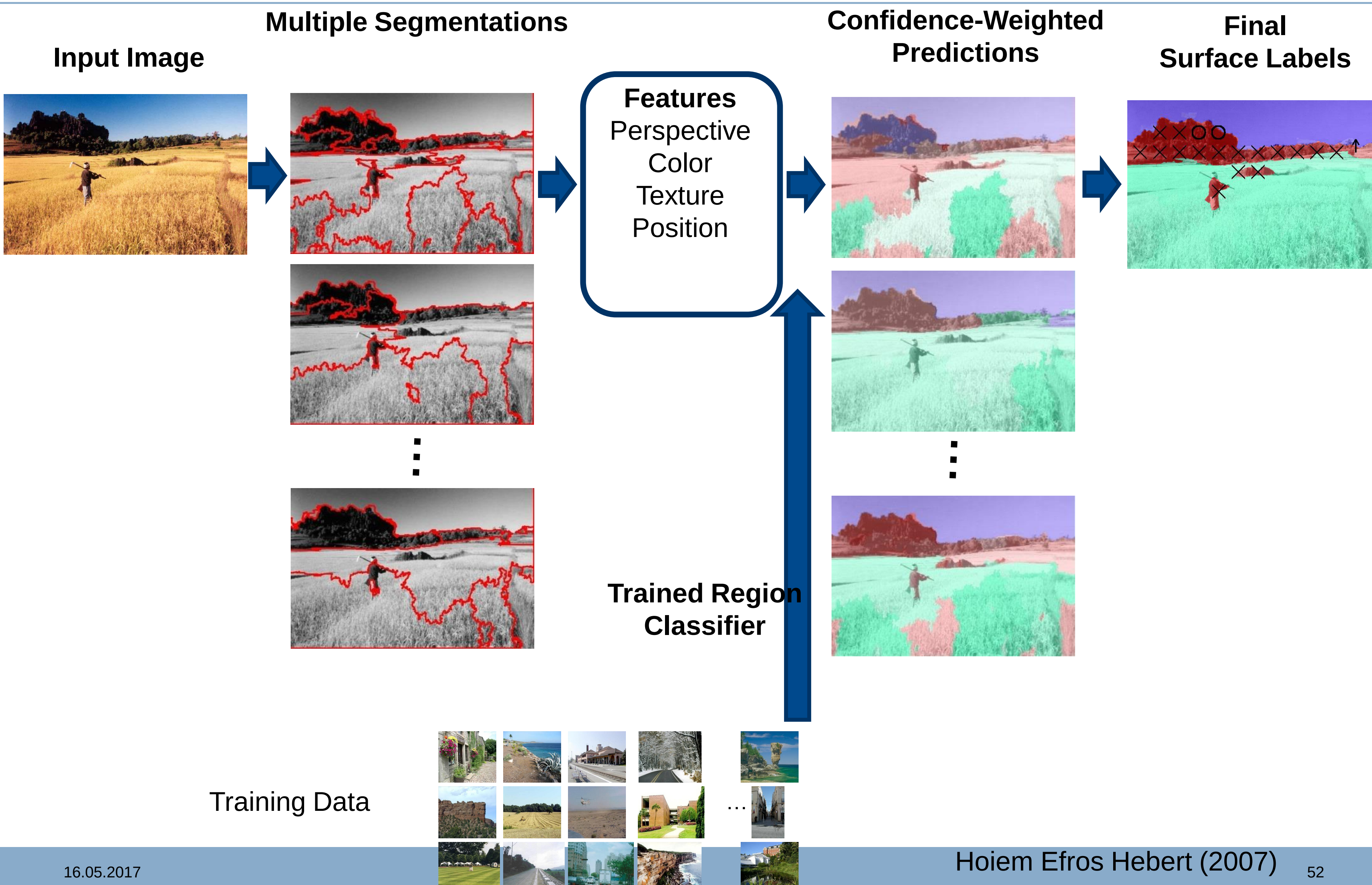
Texture gradient

Surface Layout Algorithm



Hoiem Efros Hebert (2007)

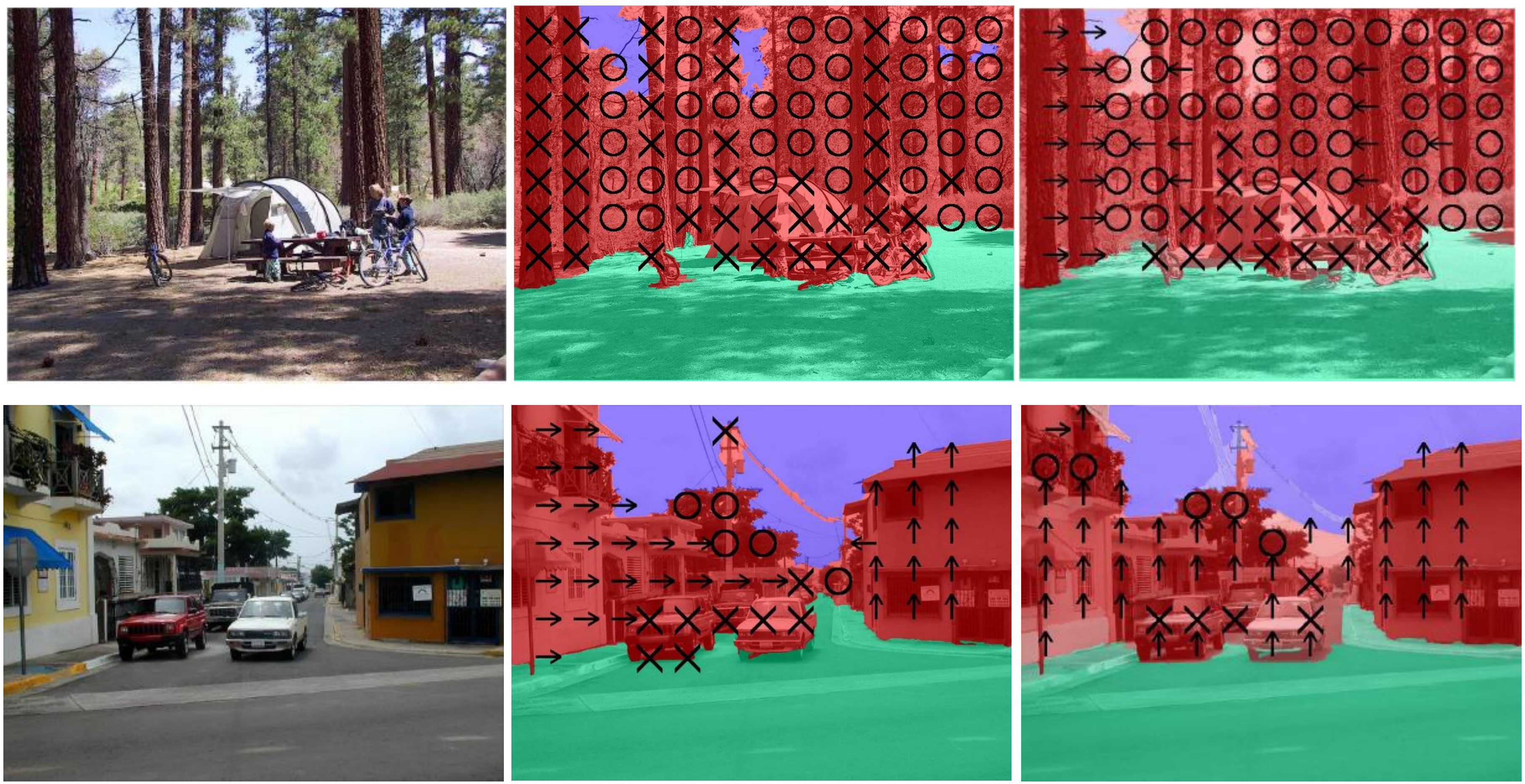
Surface Layout Algorithm



Hoiem Efros Hebert (2007)

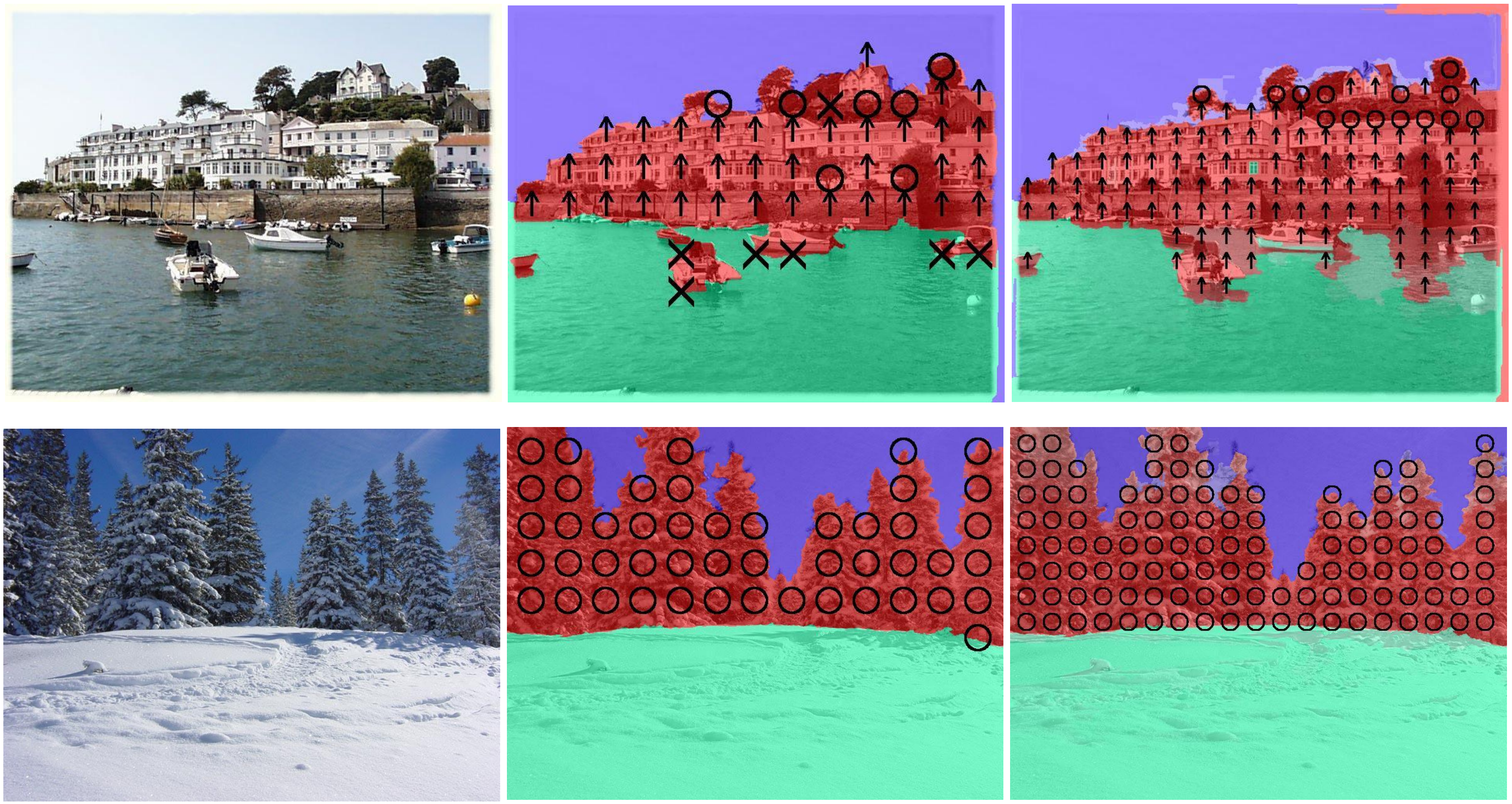
Surface Description Result

Results



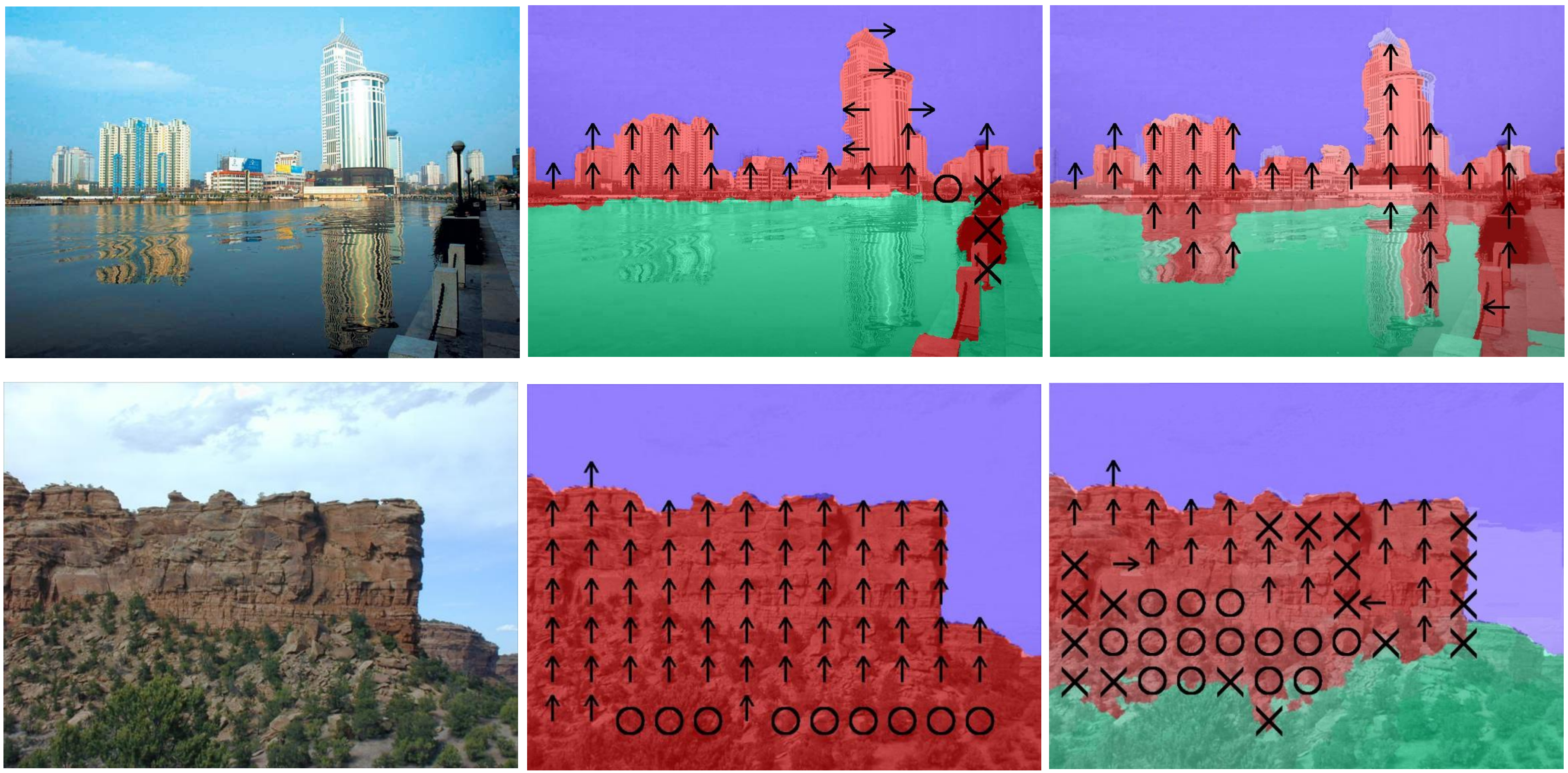
Input Image | Ground Truth | Hoeim et al Result

Results



Input Image | Ground Truth | Hoeim et al Result

Failures: Reflections, Rare Viewpoint



Input Image | Ground Truth | Hoeim et al Result

Average Accuracy (Hoeim et al)

Main Class: 88%

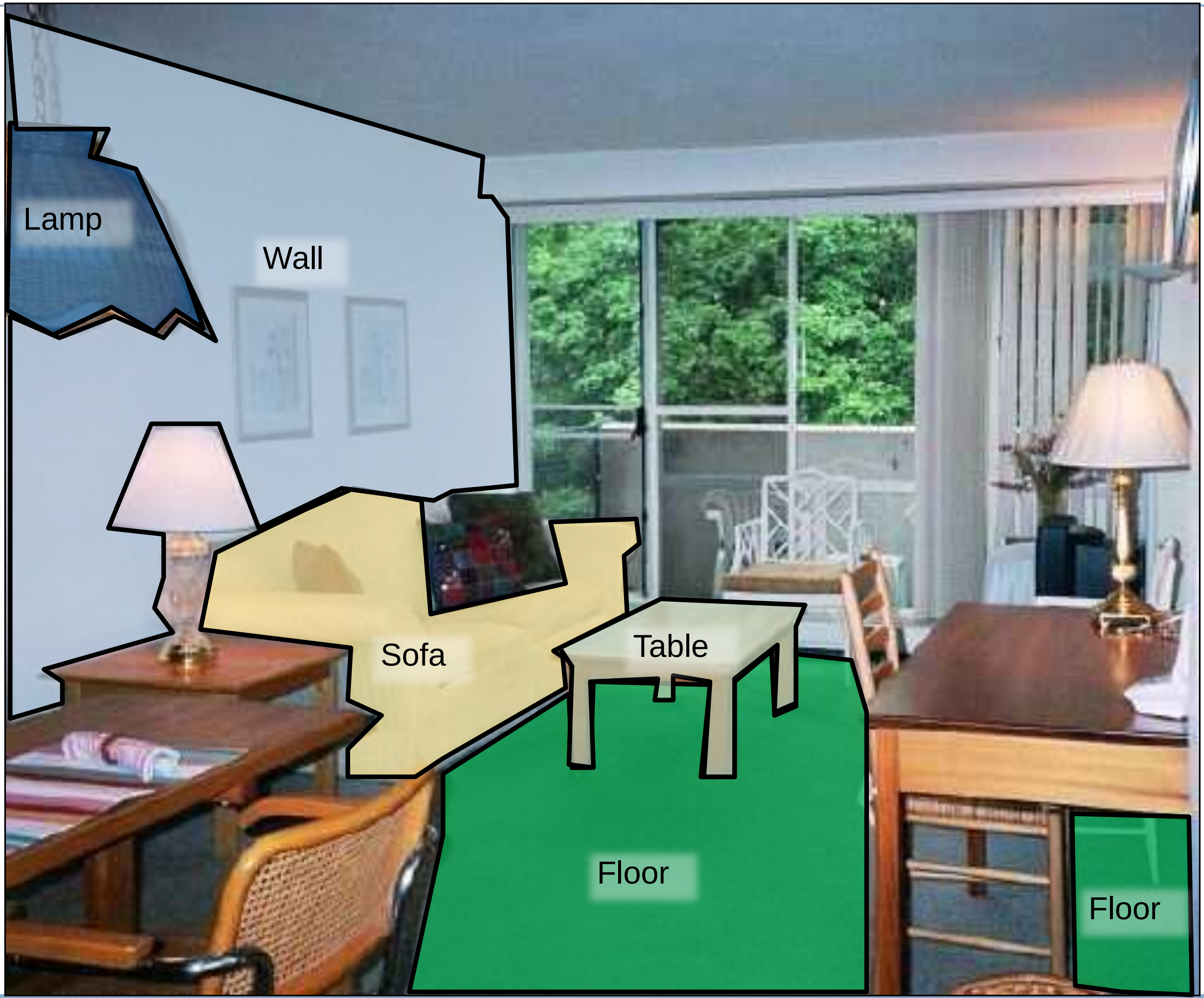
Subclasses: 61%

Main Class			
	Support	Vertical	Sky
Support	**0.84**	0.15	0.00
Vertical	0.09	**0.90**	0.02
Sky	0.00	0.10	**0.90**

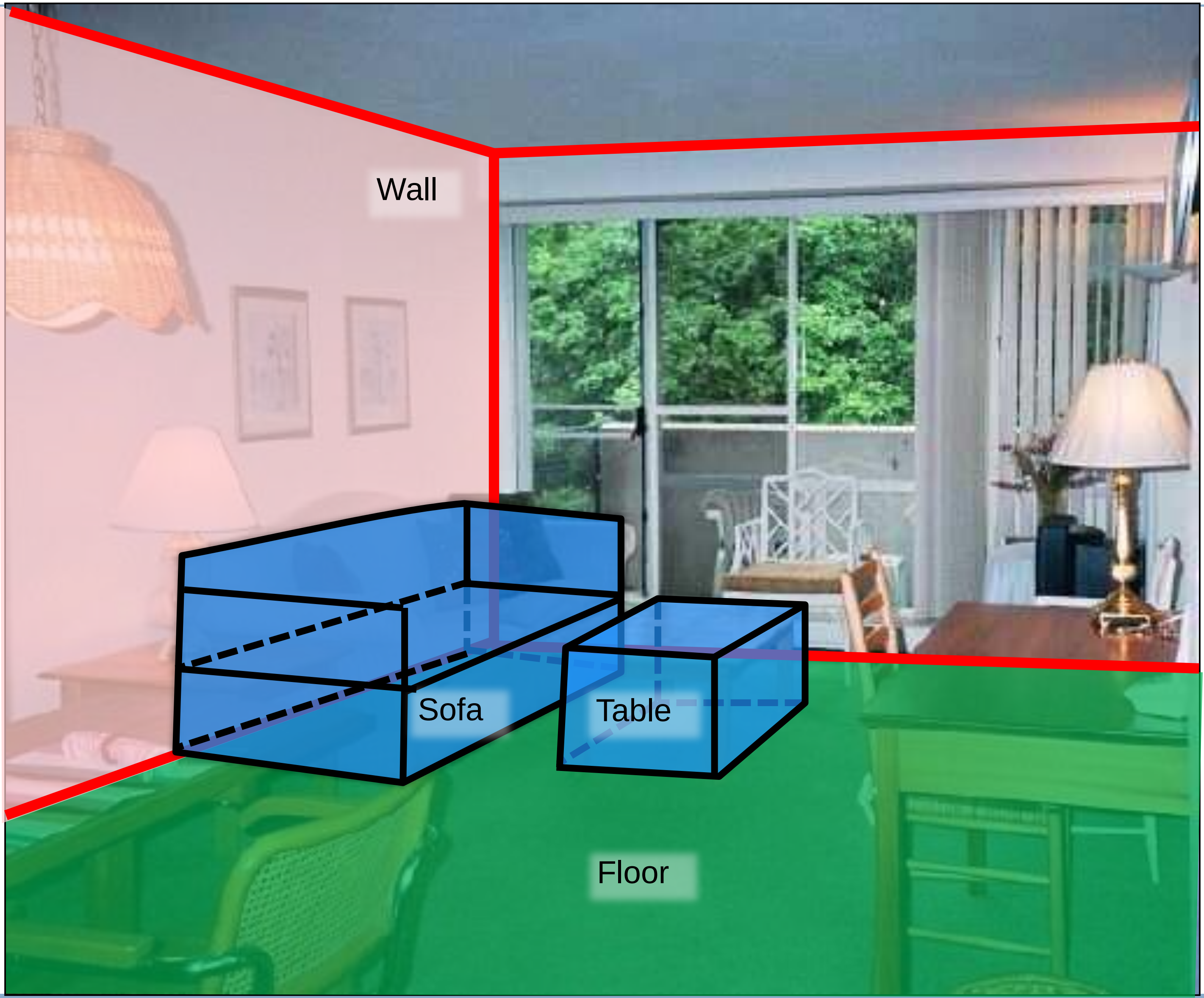
Vertical Subclass					
	Left	Center	Right	Porous	Solid
Left	**0.37**	0.32	0.08	0.09	0.13
Center	0.05	**0.56**	0.12	0.16	0.12
Right	0.02	0.28	**0.47**	0.13	0.10
Porous	0.01	0.07	0.03	**0.84**	0.06
Solid	0.04	0.20	0.04	0.17	**0.55**

Interpretation of indoor scenes

Vision = assigning labels to pixels?



Vision = interpreting within physical space



Physical space needed for affordance

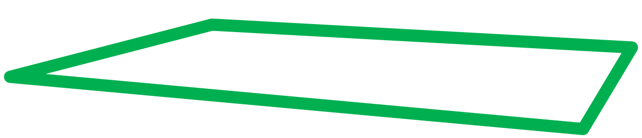
Is this a good place to sit?

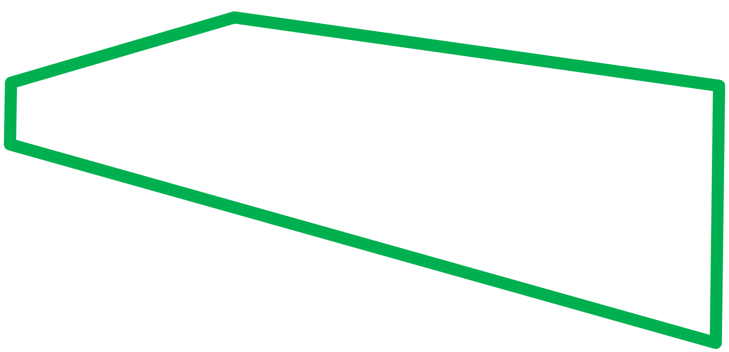
Could I stand over here?

Can I put my cup here?

Walkable path

Physical space needed for recognition



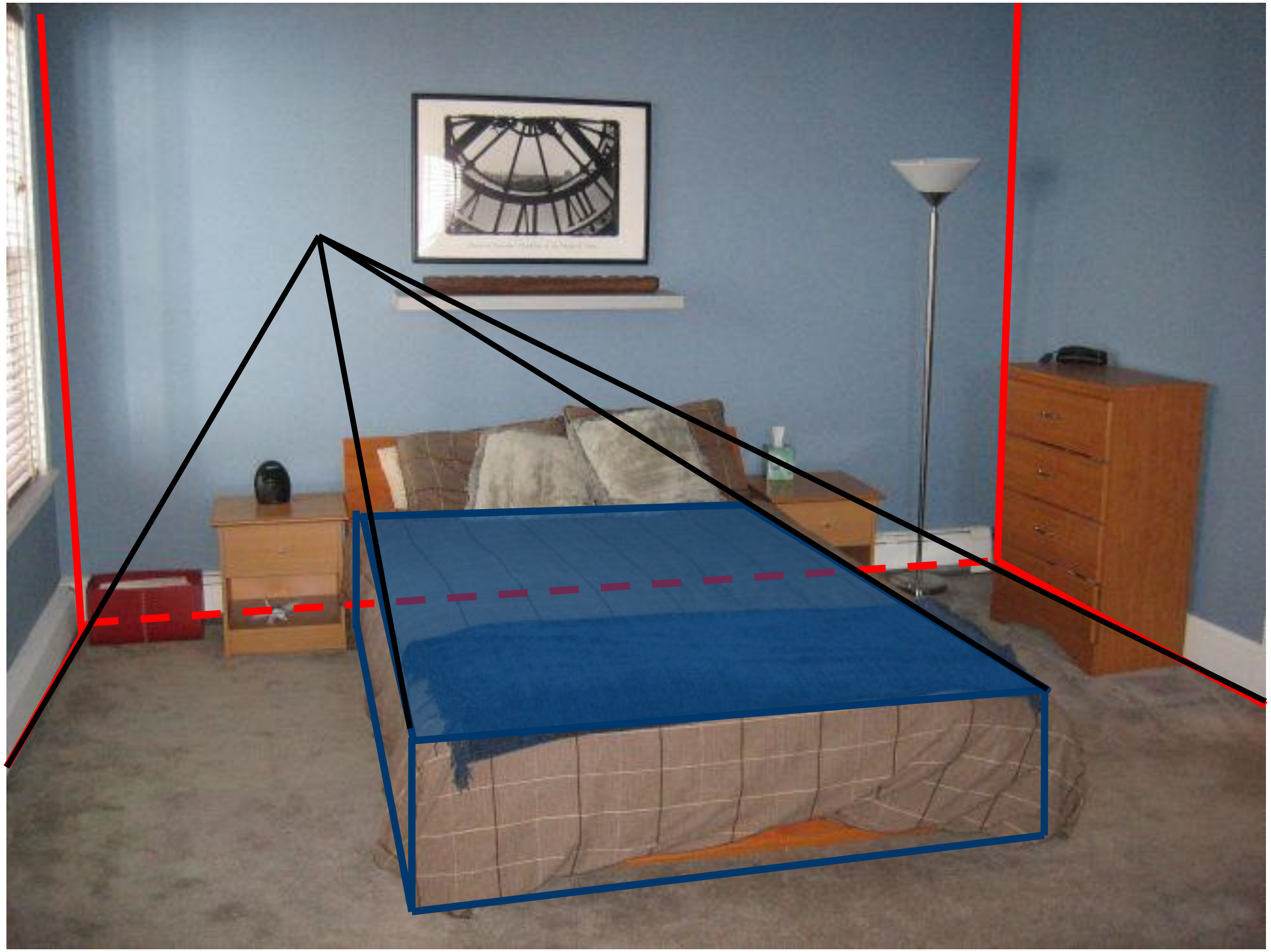


Apparent shape depends strongly on viewpoint

Physical space needed for recognition

FIAS Frankfurt Institute for Advanced Studies

GOETHE UNIVERSITÄT FRANKFURT AM MAIN



Key challenges

How to represent the physical space?

- *Requires seeing beyond the visible*

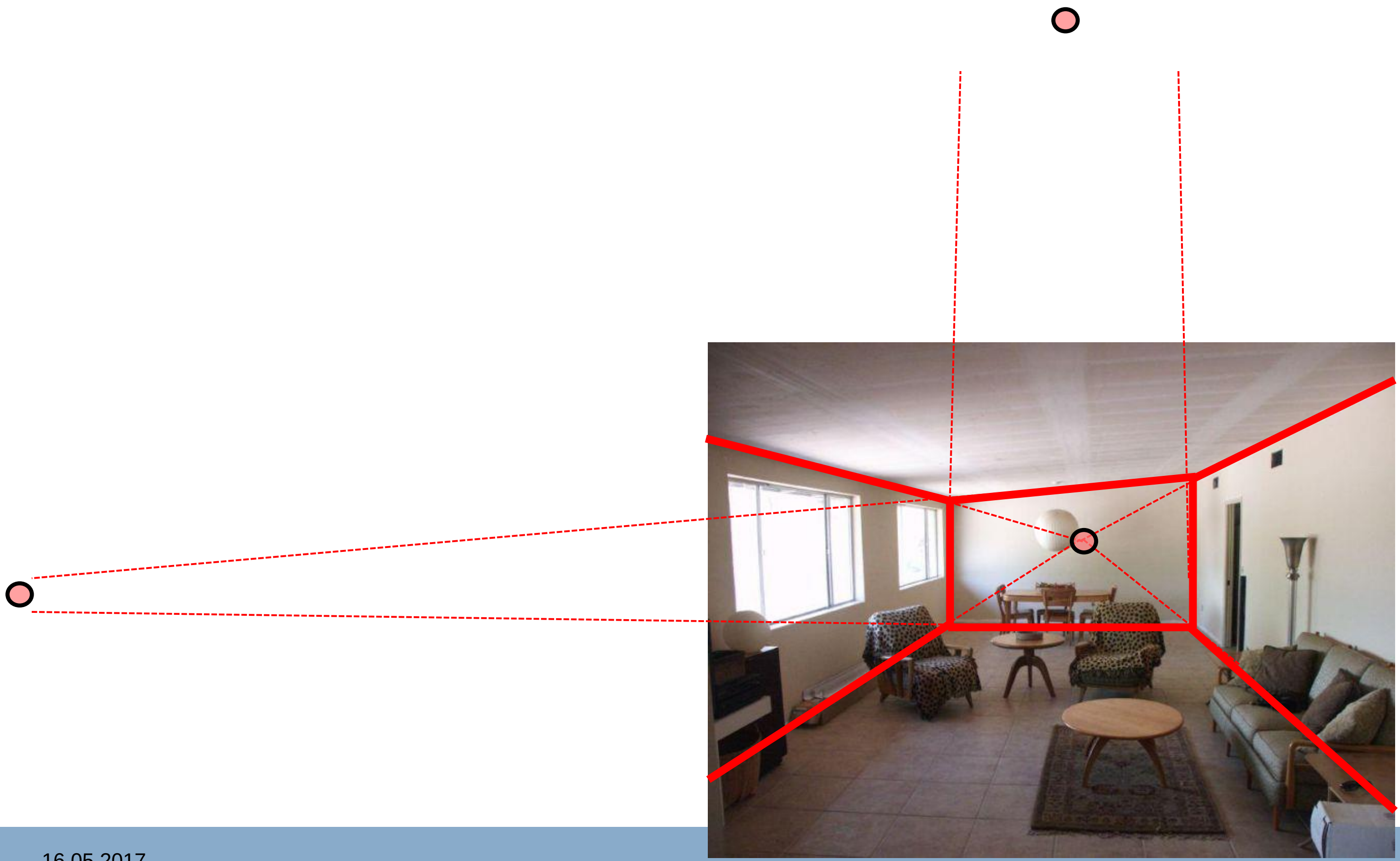
How to estimate the physical space?

- Requires simplified models
- Requires learning from examples

Example Box Layout

Room is an oriented 3D box

- Three vanishing points specify orientation
- Two pairs of sampled rays specify position/size



Box Layout

Room is an oriented 3D box

- Three vanishing points (VPs) specify orientation
- Two pairs of sampled rays specify position/size

Another box consistent with the same vanishing points

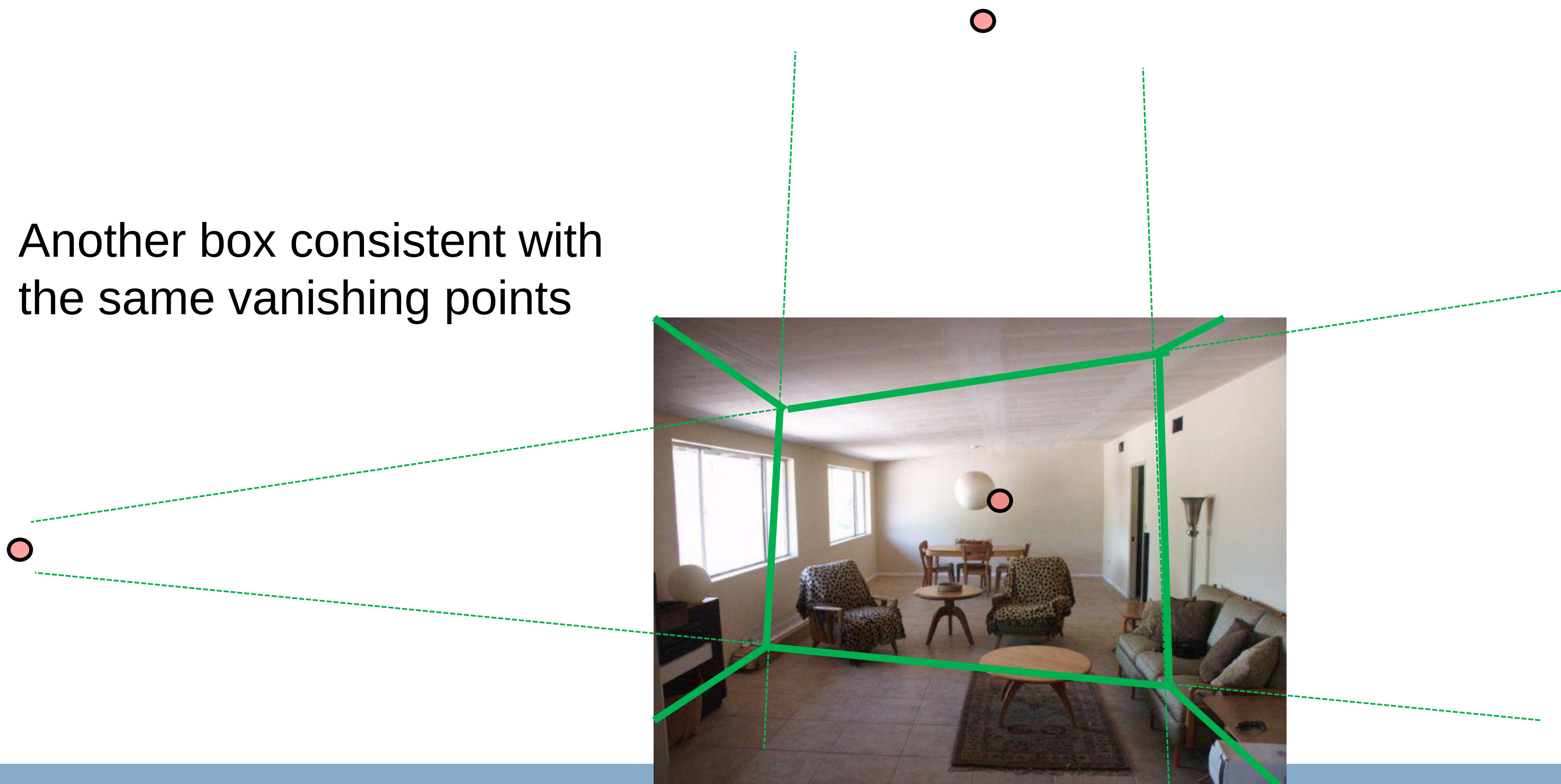
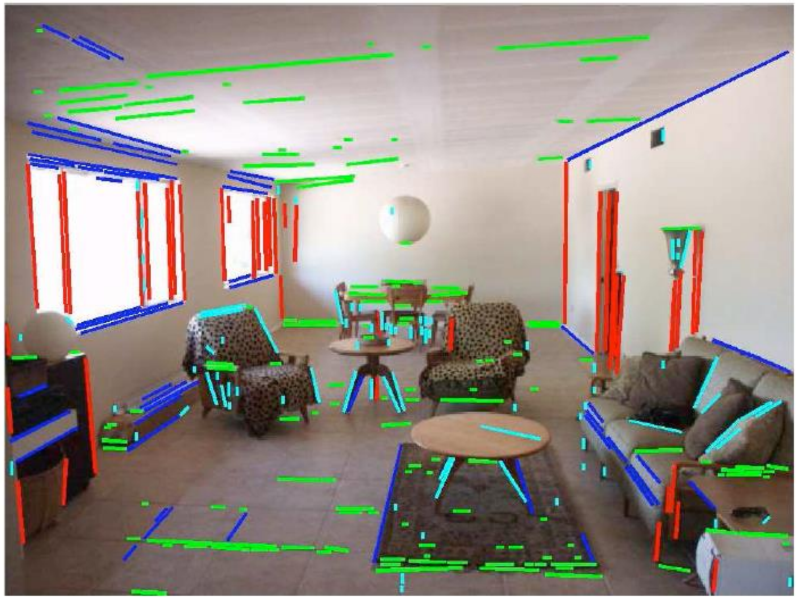


Image Cues for Box Layout

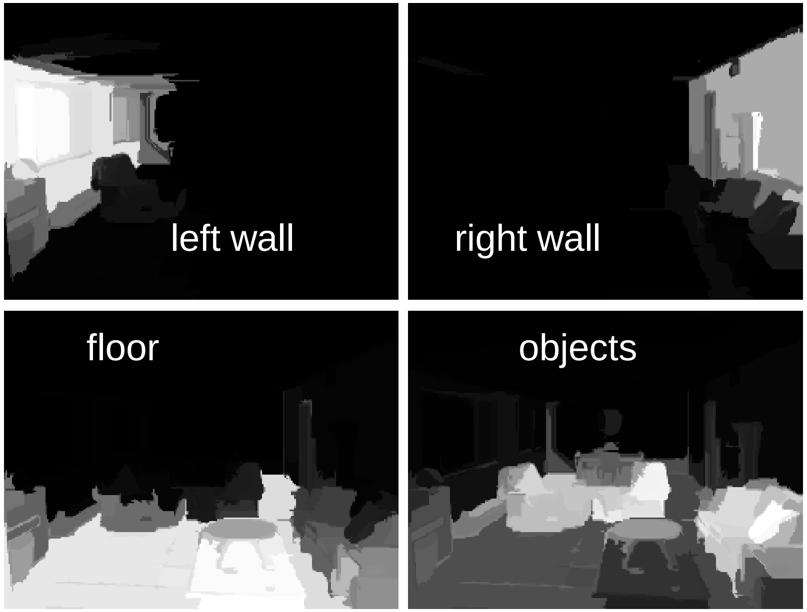
Straight edges

- Edges on floor/wall surfaces are usually oriented towards VPs
- Edges on objects might mislead

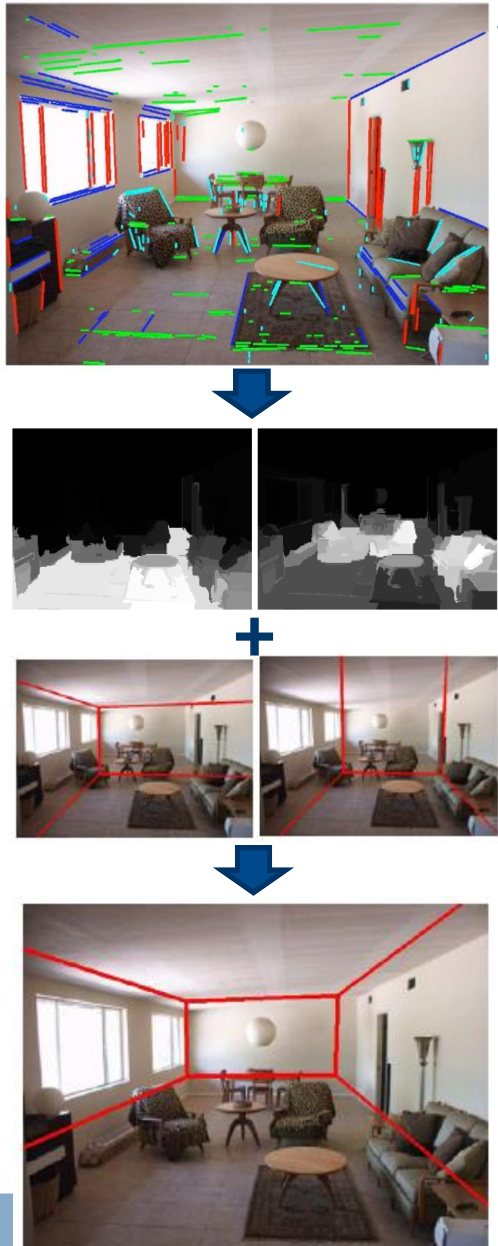


Appearance of visible surfaces

- Floor, wall, ceiling, object labels should be consistent with box



Box Layout Algorithm



1. **Detect edges**
2. **Estimate 3 orthogonal vanishing points**
3. **Apply region classifier to label pixels with visible surfaces**
   - Boosted decision trees on region based on color, texture, edges, position
4. **Generate box candidates by sampling pairs of rays from VPs**
5. **Score each box based on edges and pixel labels**
   - Learn score via structured learning
6. **Jointly refine box layout and pixel labels to get final estimate**

Evaluation

Dataset: 308 indoor images

- Train with 204 images, test with 104 images






Experimental results

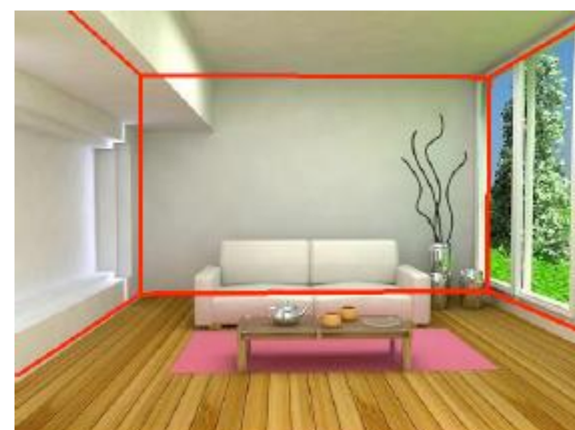
Detected Edges

Surface Labels

Box Layout

Detected Edges

Surface Labels



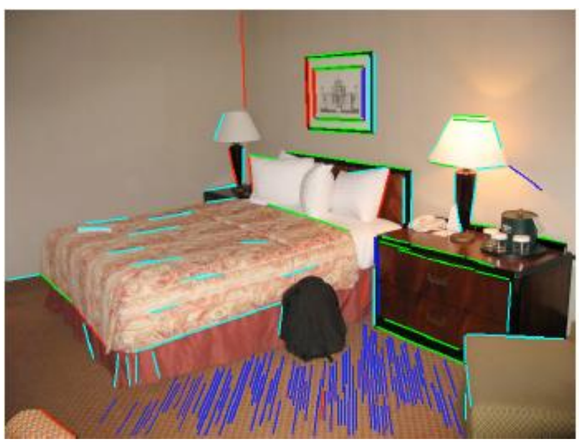
Box Layout

Experimental results

Detected Edges

Surface Labels

Box Layout



Detected Edges

Surface Labels

Box Layout

Experimental results

Joint reasoning of surface label / box layout helps

- Pixel error: 26.5% → 21.2%
- Corner error: 7.4% → 6.3%

Similar performance for cluttered and uncluttered rooms

Things to remember

Objects should be interpreted in the context of the surrounding scene

- Many types of context to consider

Spatial layout is an important part of scene interpretation, but many open problems

- How to represent space?
- How to learn and infer spatial models?
- Important to see beyond the visible

Consider trade-off of abstraction vs. precision

Relationship to Systems Engineering Example

"Our View" of Real-time Vision Systems Construction -- Summary

Overall Philosophy:

- Model-based Design using Graphical Models for Contexts
- (Context, Task, and Performance) requirements map to choice of (quasi-invariants and fusion, optimization) dual-system processing pipelines.
- Vision is Hierarchical Indexing (via use of quasi-invariants) followed by iterative estimation in a coarse to fine manner.

Key Observations:

- Bayesian formulation without proper design of priors and likelihood models is as "Adhoc" as any other approach
- "Real-time Bayesian Inference" is not possible with conventional approaches to Bayesian inference (e.g. Sampling techniques). Hence need a way to deal with the situation.

“Our View” of Real-time Vision Systems Construction -- Approach

FIAS Frankfurt Institute for Advanced Studies
GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Our Approach:

- Define appropriate statistical representations for application context (i.e. a graphical model, *G*) and learn them.

- Use these contextual priors, tasks, and performance requirements to select transforms (i.e. quasi-invariants) that devise hypothesis generator modules and configurations, *M*.
  - ❑ This process is non-trivial, currently structure of the module configuration is provided by expert while parameters can be learned.

- Perform systems identification of hypotheses generator pipeline, *M*.

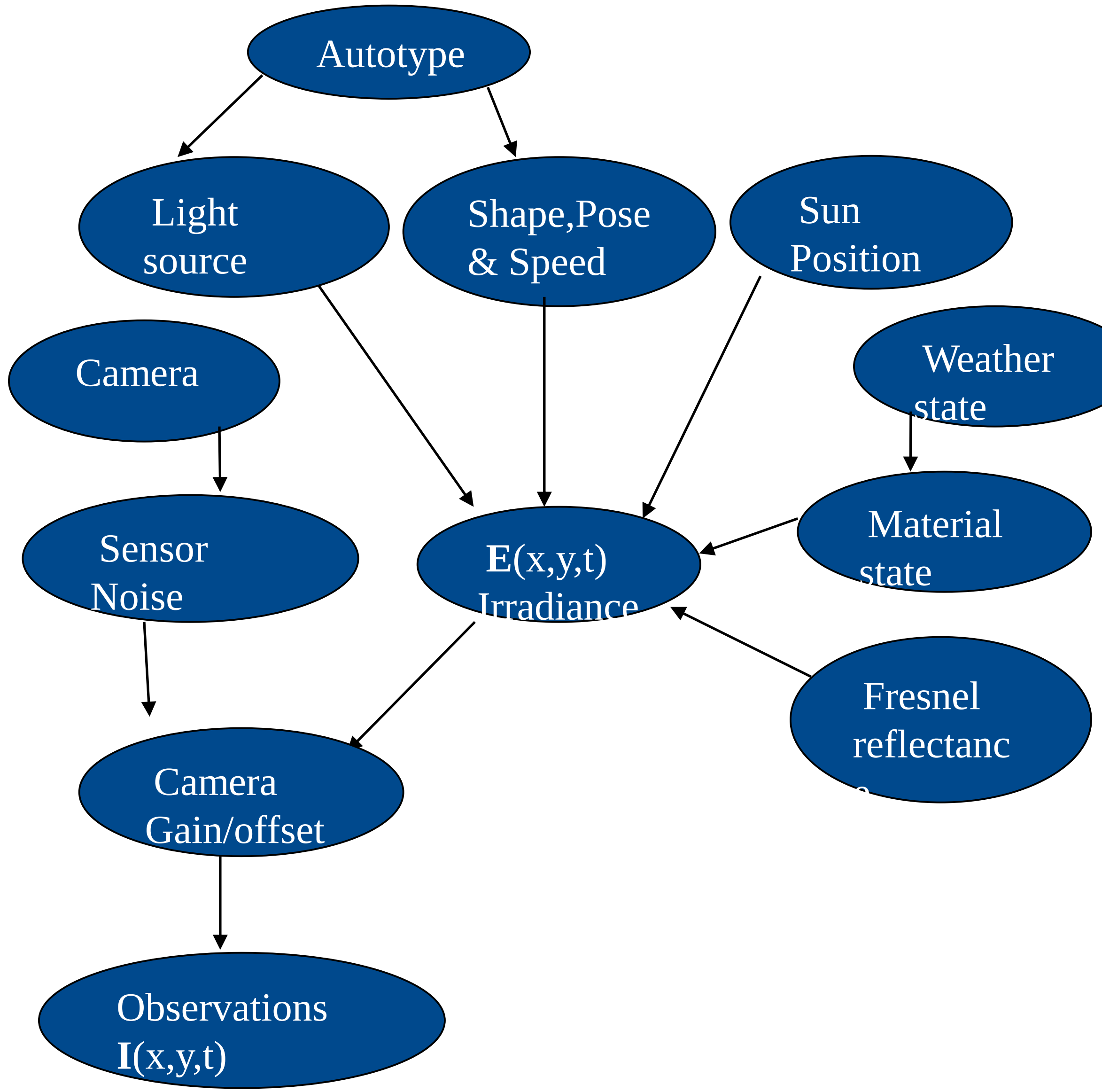
- Transform graphical model, *G*, to ‘G*’ corresponding to graphical model including random variables corresponding to hypotheses. Belief propagation or Markov-Chain Monte Carlo using ‘*G*’* provides the solution.

"Step 1: Vision System Specification Context Modeling + Task Requirements"

FIAS Frankfurt Institute for Advanced Studies
GOETHE UNIVERSITÄT FRANKFURT AM MAIN

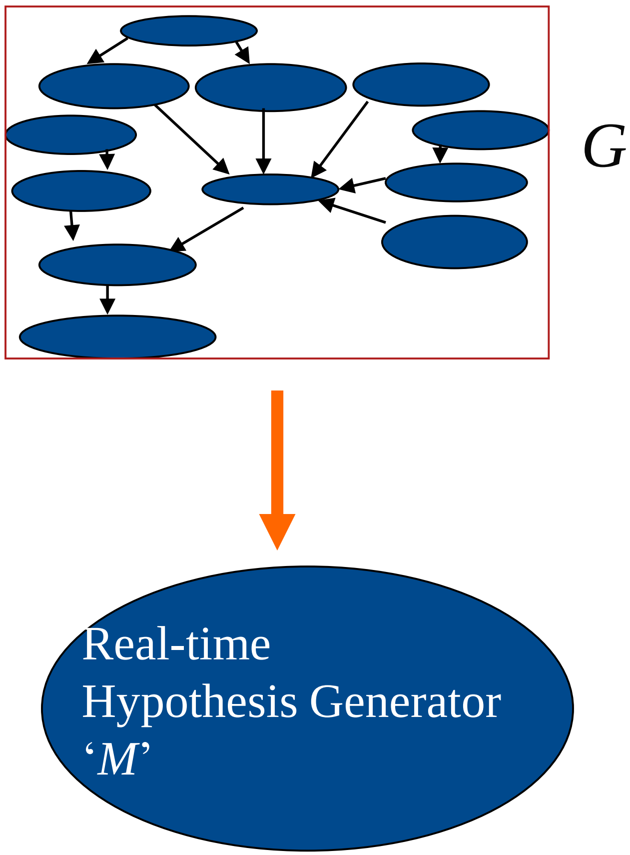
- Graphical Models as a specification of the generative model for observations
- Task specification including variables to be estimated, performance requirements (speed, accuracy)

Inference:
"Real-time
Inference intractable"



"Step 2: Design of Hypothesis Generation Module for specific context"

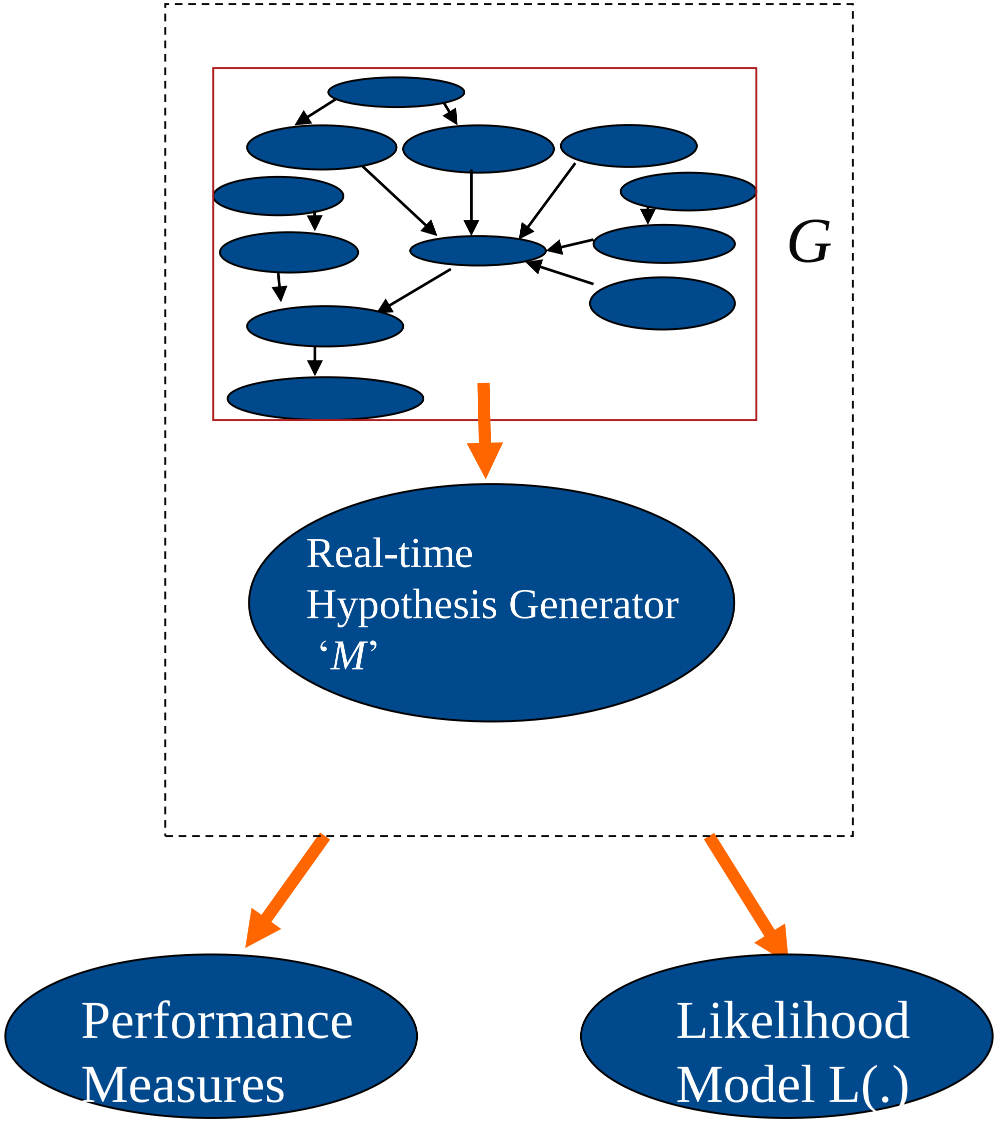
- ➢Translate Priors to a sequence of operator steps that generate feature measurements & hypotheses in real-time
- ➢ Key requirement: Hypothesis generator has to be quantifiable in performance via P(correct hypothesis generation), P(false hypothesis generation), and L(Features|Theta_Scene) is derivable.



"Step 3: Systems Identification of Hypothesis Generation Module"

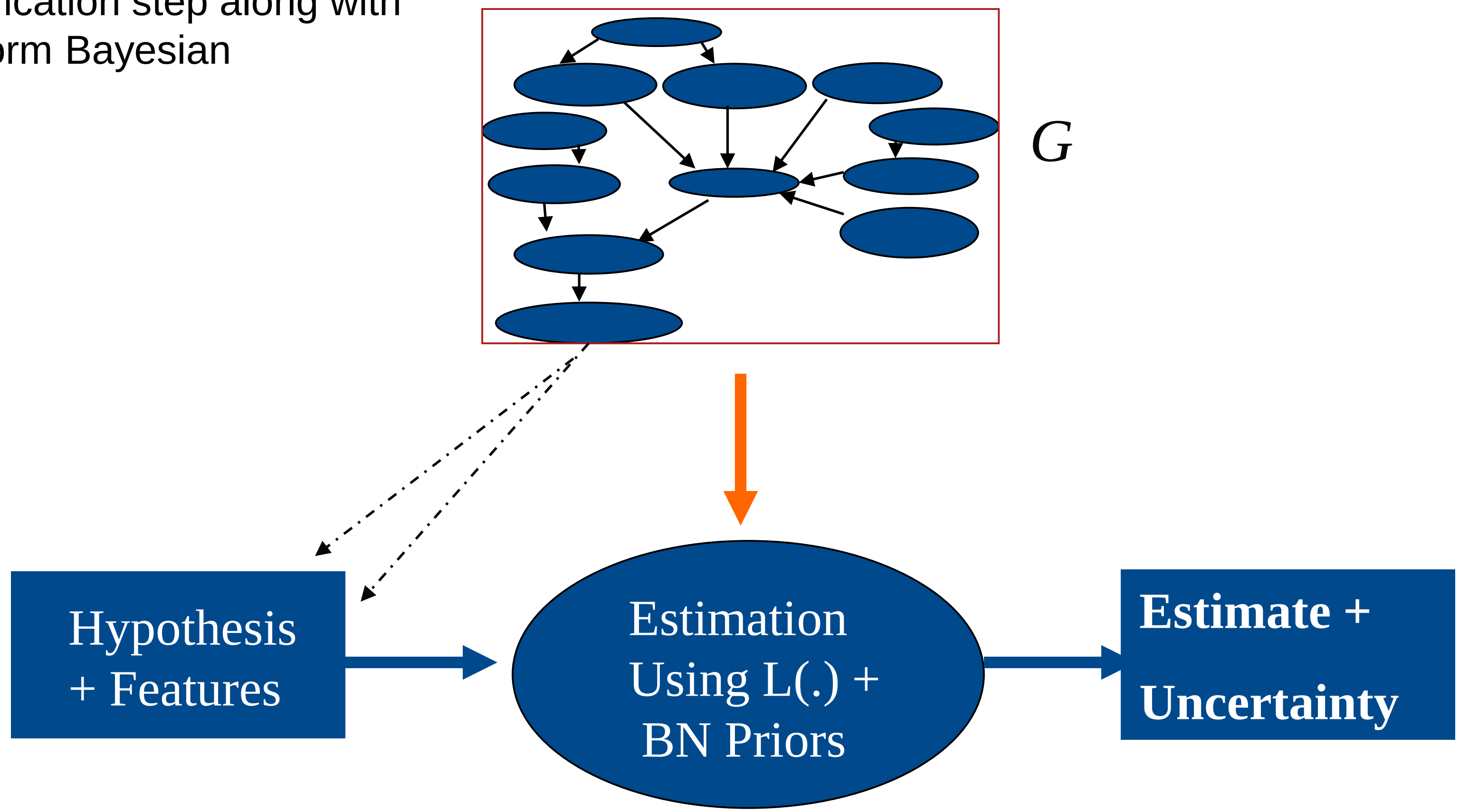
➢ Perform Systems analysis of hypothesis generation module: I.e. compute: P(correct hypothesis generation), P(false hypothesis generation)

➢Derive L(Features|Theta_Scene) as a function of *M* and *G*. (I.e. Treat *M* like a soft-sensor and develop the likelihood model for measurements).



“Step 4: Hypothesis Verification & Estimation Module”

- ➢Use Likelihood model derived in system identification step along with priors to perform Bayesian estimation



Real-time Vision Systems – Technical challenges

- Systems Analysis is complex, but critical to develop proper statistical likelihood models. “Analysis for Linear transforms solved, made progress on non-linear transform characterization but lots of fundamental issues remain.” (Ramesh and Haralick, 1992-97, Forstner 1994-2001, Parra, Lai and Ramesh 1998, Greiffenhagen and Ramesh, 1999-2001, Gao, Boult, Ramesh (2000, 2002), Tsin, Ramesh, Kanade (2001), Meer et al (1995-2001) etc.)
- Prior models and their choices for application space still an art. Empirical Bayesian methods can be used to learn these priors, but choice of representation for priors a difficult issue. ➔ “Priors at various levels interact” (Ramesh, Parra, Qian 1997, Greiffenhagen and Ramesh 2001).
- Priors in the 3D scene parameters are relatively easy to model. However, their counterparts in 2D have multiple plausible representations and one has difficulty to decide what representation to describe the prior in 2D. Our view is that there is a mapping from Applications → appropriate representations. (Coetzee and Ramesh 1999, Greiffenhagen and Ramesh 2001)

Technical challenges: Continued

- Stability of representation is dependent on variability and perturbations in data ➔ "Explore statistical distributions in various representation spaces: e.g. 2D geometry, Image pattern distributions, and intermediate data-structures, level sets" → Study various representations and understand their limits (e.g. Ramesh and Haralick (1994), Paragios (1998-2001), Zhu et al (2001), Gao et al (2000), Sun et al (1999-2001))
- Distributional representations cannot be simple ➔ " World is Non-Gaussian" → Our approach is to mix Non-parametric statistical representations with parametric ones (I.e) Semi-parametric representations. (e.g. Mean-shift, adaptive mean-shift, Comaniciu et al (1998-2001))
- Choice of Feature representations and Transforms ➔ Motivate from Physics and scene constraints (e.g. Greiffenhagen and Ramesh 2000, Tsin, Ramesh, Kanade 2001), Motivate through learning (Coetzee and Ramesh, 1999), Motivate through operator analysis (Gao, Ramesh, Boult 2002), Motivate from brain sciences. Automating this choice will be a major challenge!!!!

Contextual Models: Simulation from Generative Models

FIAS Frankfurt Institute for Advanced Studies

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

- Camera geometry -- projection model (orthographic, perspective), camera blur, lens distortion, intrinsic parameters, extrinsic parameters.
- Camera - gray level transformation model of camera pipeline
- Shape representation (surface/contour, volume)
- Material property (brdf)
- Appearance (texture map) dictionary
- Graphics pipeline parameters

Pattern Models

- Image models for textures
  - sparsity based
  - MRF model Pipelines
- Texture classification
  - Hand engineered features plus ML - deep learning
- Texture synthesis
  - Deep learning
  - Exemplar based with smoothness constraint (efros)
  - Bio-inspired (wavelets plus correlation, constrained sampling) - mrf models

Spatial Statistics - Simulation

Simulation Models –

- spatial point processes
  - poisson point process
    - in homogenous process
    - cluster processes (cox, matern hard core, etc)
    - Boolean germ-grain models
    - dead leaves model
- Simulation apparoach
  - rejection sampling
  - Markov Chain Monte Carlo
  - CFTP - Coupling from the past
  - Representations: Sparsity based reps, Texture models using MRF's

- Open questions:
  - Realism of models ?
  - Model validation against real data

Simulation for Systems Design, Analysis and Evaluation

FIAS Frankfurt Institute for Advanced Studies

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

- **Groundtruth collection seems to be an obstacle for Supervised learning based vision systems.**

- **Major advances in Computer Graphics (CG) field has spurred a renewed recent interest to utilize CG for CV.**

(a) 2001 (b) 2003 (c) 2005

(d) 2006 (e) 2013 (f) 2015

Figure 1-1: Evolution of Graphics in Video games from 2001 to 2016

Rendered Data - various scene conditions

Lambertian

Ray traced

Path traced (130 spp)

Noon

Night

Rain

Annotations are “free”

Simulation Research – Problem Statement

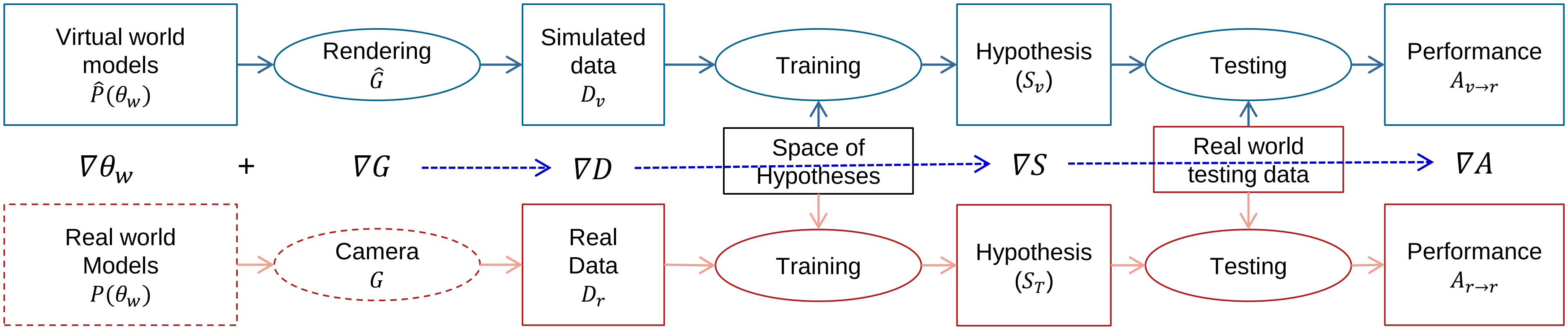
- **In 1990’s, CV community had been skeptic to use CG to train.**
  - CG may use some mathematical simplifications and approximations that CV models based on. Hence, they might generate ideal or near ideal to CV models.
  - That was good question in the days of model-driven designs.

- **Recent video games may also use approximated models for realistic effects for the sake of interactive real time display.**
  - Now, in the days of data driven designs, how these approximations effects the CV ?

- **Deviations of Scene (parameters) distributions plus Physical accuracy of rendering processes contributes towards Domain-shift issue b/w virtual and real world data.**

- **Especially, the impact of modeling errors and computational rendering approximations, due to choices in the rendering and generation pipeline, on trained CV systems generalization performance is still not clear.**

Transfer and Domain shift



- No free lunch in the selection of $\hat{P}$ and $\hat{G}$ for data simulation processes.
- In principle, $\nabla\theta_w$ and ∇G impact the magnitudes of ∇D, ∇S, and ∇A.

- ***What is the impact of $\hat{G}$ on ΔA?***
  - Real time Photo-realism vs Expensive physics-realism?
- ***What is the impact of parameters of $\hat{P}(\boldsymbol{\theta_w})$ on ΔA?***
  - How far can we go with an arbitrary scene generative model?
  - Can unsupervised generative learning from target real data help?

- **However, one can bypass these issues by simply adding some real samples to simulated data.**

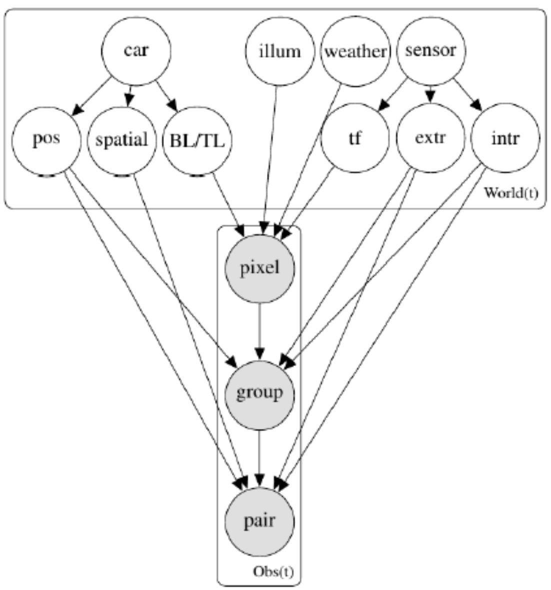
Anomaly Detection – Case Study

FIAS Frankfurt Institute for Advanced Studies

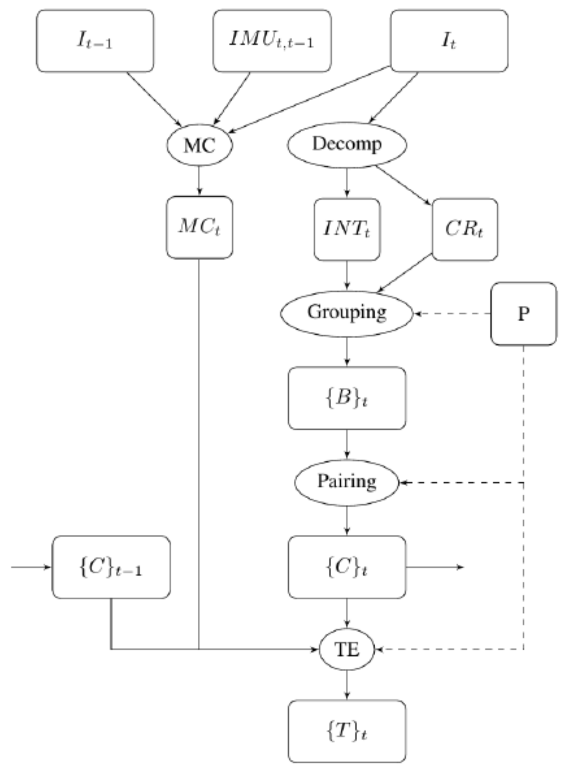
GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Thesis / case study – Brakelight transition detection

GOETHE UNIVERSITÄT FRANKFURT AM MAIN



car-pos	6d-pose of observed car
car-spatial	spatial layout of observed car's parts
car-BL/TL	Brakelight/Taillight characteristics (color, emission-strength)
illum	illumination
sensor-tf	transfer function
sensor-extr	extrinsic parameters
sensor-intr	intrinsic parameters
pixel	observed pixel values
group	spatial poolings of pixels
pairs	configurations of groups



Worldmodel:

- Latent variables
- Causal relations

Populate by

- Specifications
- Physics
- Simulations

Mapping to algorithms & pipeline:

- Identification of relations of latent contextual variables and observations -> submodalities
- C,T,P -> Modules + Parameters

Model-based Sys-engy approach:

- Loose coupling, single module evaluation
- Uncertainty propagation

Next Classes

- Introduction to Probabilistic Graphical Models
- Pattern Grammars and Inference

‘Thank you’ -- The End

Backup