BROWN

Hierarchy and Reusability in Image Analysis

Stuart Geman

Eran Borenstein, Ya Jin, Wei Zhang

I. Remarks on Computer Vision
II. Approaches
III. Bayesian Image Analysis
IV. Probability Models
V. Demonstration System: Reading License Plates
VI. Generalization: Face Detection

I. Remarks on Computer Vision
- Vision is hard
- Why is vision hard?

II. Approaches
III. Bayesian Image Analysis
IV. Probability Models
V. Demonstration System: Reading License Plates
VI. Generalization: Face Detection

License plate images from Logan Airport

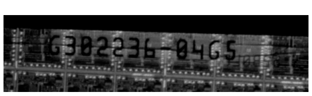




Machines *still* can't reliably read license plates

Wafer ID's



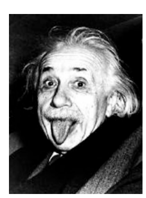
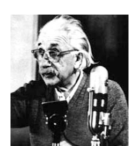
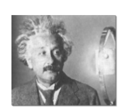


Machines can't read fixed-font fixed-scale characters as well as humans

---

Super Bowl



Machines can't find the bad guys at the Super Bowl

Instantiation

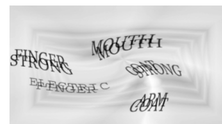
same

Empire style table

twins

Vision is *content sensitive*

"Clutter"



Human Interactive Proofs

Background is *structured,* and made of the *same stuff!*

I. Remarks on Computer Vision

II. Approaches

- **Pure learning**
- **Fodor & Pylyshyun, 1988, and the critique of neural networks**
- **Observations from the cognitive, neural, and mathematical sciences**

III. Bayesian Image Analysis

IV. Probability Models

V. Demonstration System: Reading License Plates

VI. Generalization: Face Detection

Pure learning

Google:

- 2 billion images
- train classifier: pornographic/not pornographic
- good classification
- not nearly as good as human performance

Pure learning

Human learning:

- 1 sample/10 seconds
- 16 hours/day
- 80 years
- < 170 million samples/lifetime

Enough examples?

Did *evolution* have enough examples?

D. Geman: "The interesting limit is N goes to zero,
not N goes to infinity"

Fodor & Pylyshyun, 1988, and the critique of neural networks

Properties of human cognition:

* <u>compositionality</u>

  roughly: representation through syntactically constrained hierarchy of reusable parts

* <u>productivity</u>

  roughly: capable of an infinite number of well-formed actions, thoughts, sentences …

* <u>systematicity</u>

  roughly: invariance

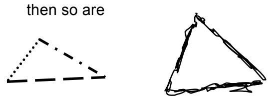
Fodor & Pylyshyun, 1988, and the critique of neural networks

systematicity

if 'the boy ran home'
makes sense, then so does
'the girl ran home'
'john ran home'
'john ran to school'

If  is a triangle,

then so are



4 4 4 4 4 4 4
4

Observations from the cognitive, neural, and mathematical sciences

Human brains utilize *strong representations*

- **Damassio (simulation=perception)**
- **Kosslyn ("resolution of the imaging debate")**
- **Lakeoff, Fauconnier (the role of mental imagery in language understanding)**

Observations from the cognitive, neural, and mathematical sciences

Consider the <u>ventral visual pathway</u>:

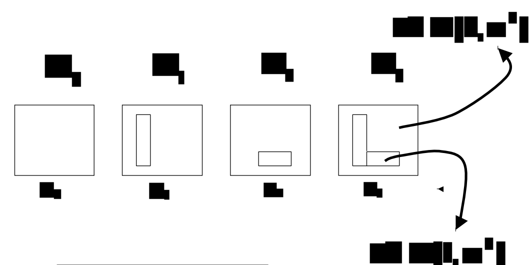
Structure:	**retina ↔ LGN ↔ V1 ↔ V2 ↔ V4 ↔ IT**	
Topography: topology	**highly retinotopic** ↔	**little**
Receptive Field:	**small** ↔	**large**
Specificity:	**low** ↔	**high**
Invariance:	**low** ↔	**high**

SUMMARY: A hierarchy of less-to-more invariant representations

Observations from the cognitive, neural, and mathematical sciences

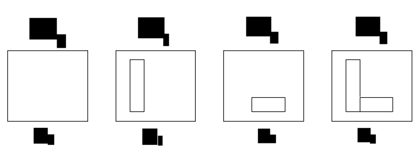
Thought experiment:

what if <u>the world</u> is a hierarchy of reusable parts?



Problem: test for 'L'

Observations from the cognitive, neural, and mathematical sciences

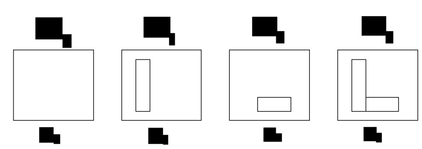


God's ROC curve (Neyman-Pearson Lemma):

[illegible]

impossible in practice…

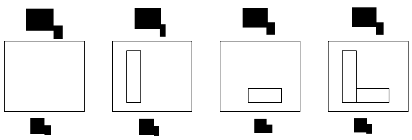
Observations from the cognitive, neural, and mathematical sciences



Testing against a universal null (pragmatic):

[illegible]

Observations from the cognitive, neural, and mathematical sciences

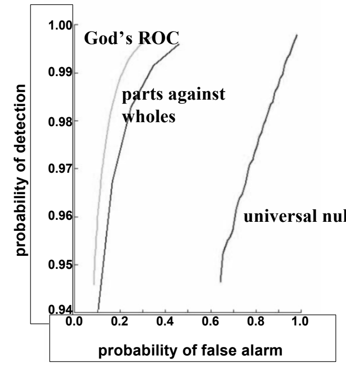


Testing "parts" against "wholes":

[redacted]

Observations from the cognitive, neural, and mathematical sciences

ROCs



Observations from the cognitive, neural, and mathematical sciences

Theorem (S. Geman, Y. Jin, W. Zhang) At any fixed probability of detection (`type I error'):

[redacted]

& various generalizations…

Take Home Message: in a compositional world

- background confusions occur at subsets of objects
- objects come equipped with their own background models

Bayesian image analysis - overview

$\mathcal{I}$ abstract space of interpretations

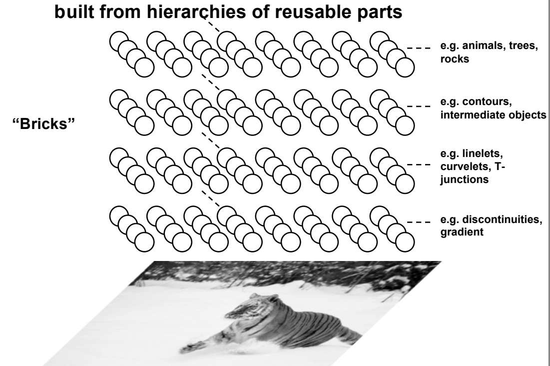
$P(I)$ prior probability, $I \in \mathcal{I}$

Y image – vector of pixel intensities

$P(Y|I)$ data model

Given an image Y:
find a 'good' interpretation I using P(I|Y)

Interpretations:
built from hierarchies of reusable parts



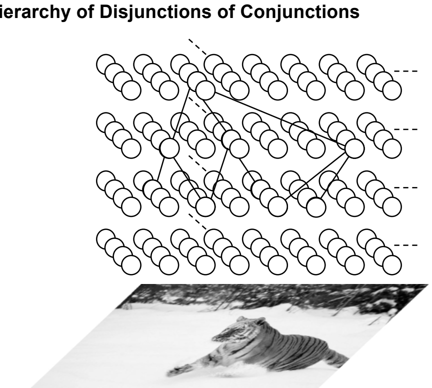
Hierarchy of Disjunctions of Conjunctions

Hierarchy of Disjunctions of Conjunctions

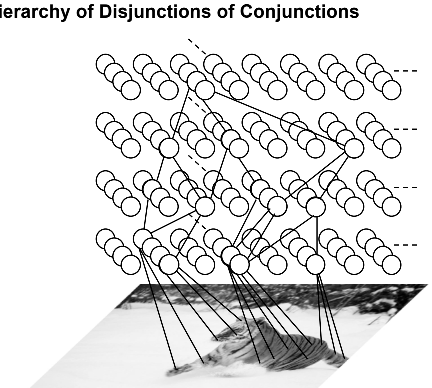
Hierarchy of Disjunctions of Conjunctions

Hierarchy of Disjunctions of Conjunctions

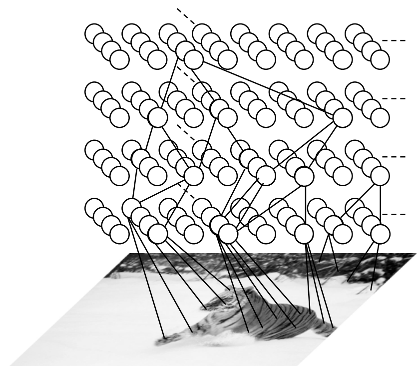
Hierarchy of Disjunctions of Conjunctions



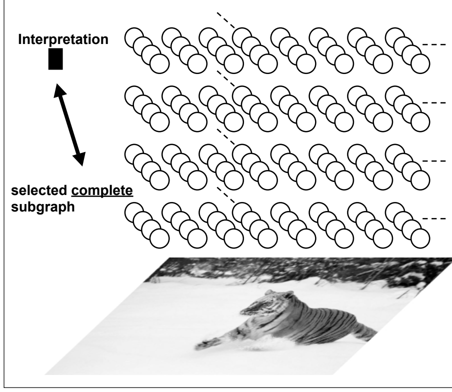
Hierarchy of Disjunctions of Conjunctions



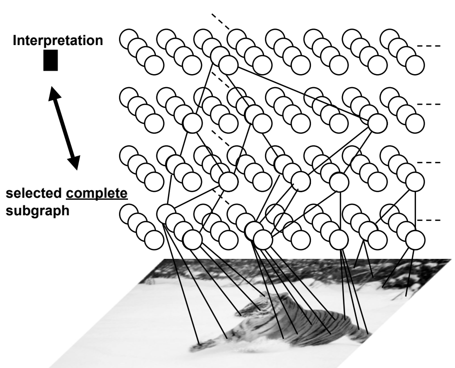
Hierarchy of Disjunctions of Conjunctions



Interpretations

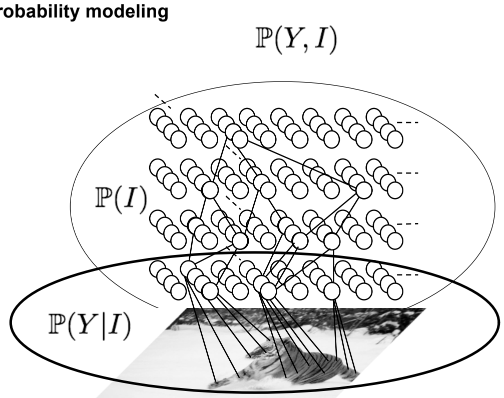


Interpretations



I. **Remarks on Computer Vision**
II. **Approaches**
III. **Bayesian Image Analysis**
IV. **Probability Models**
  - **P(I)**
  - **P(Y|I)**
V. **Demonstration System: Reading License Plates**
VI. **Generalization: Face Detection**

Probability modeling



Architecture

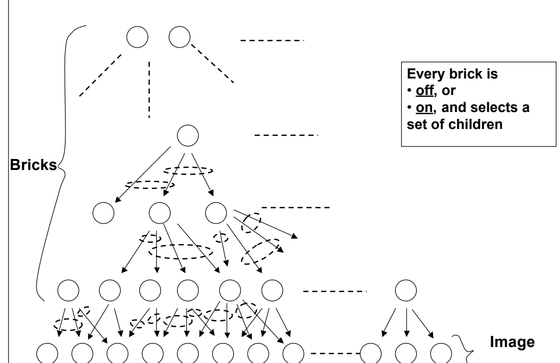
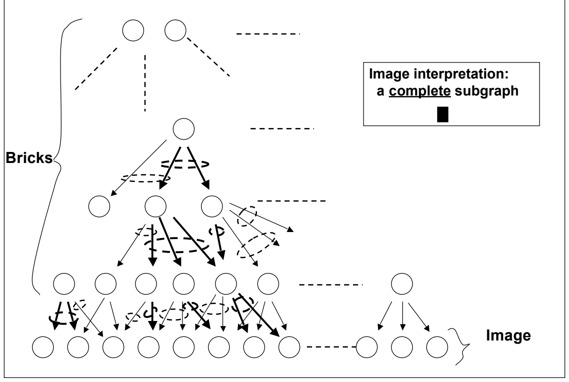
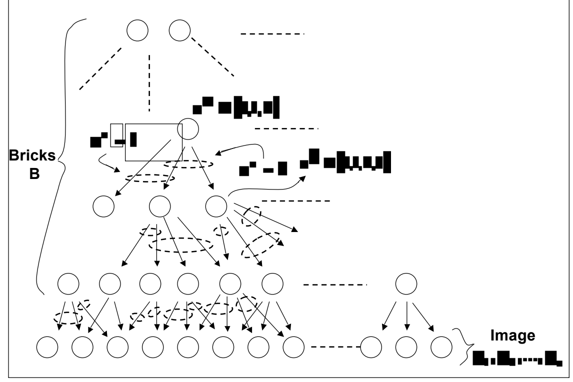


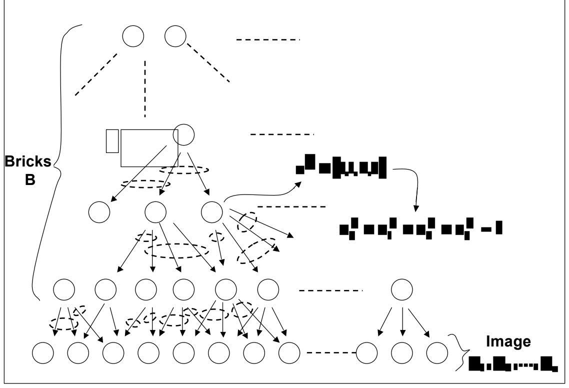
Image interpretation



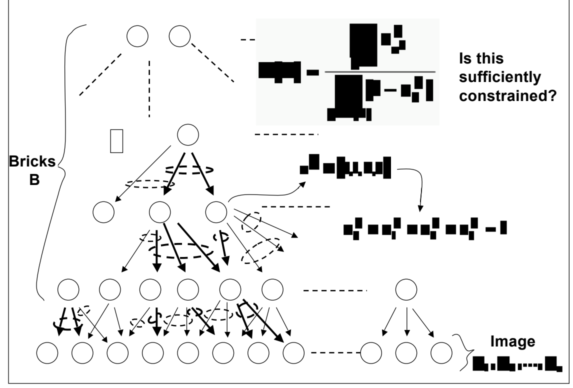
Notation



Markov backbone

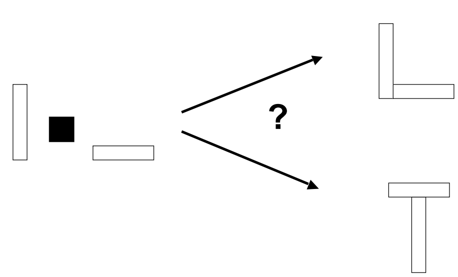


Markov backbone

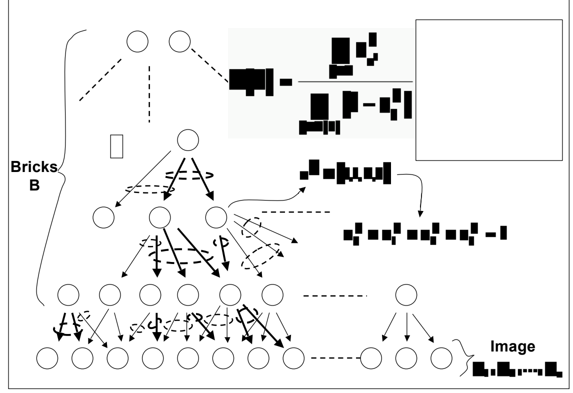


Beyond Markovian distributions

Compositions depend on instantiations…e.g. the positioning of parts



Compositional distribution – a perturbed Markov model



Content-Sensitive Perturbation

I	$\{x^\beta\}_{\beta\in B}$, state of each brick
β	arbitrary, fixed, brick
$a^\beta(I)$	arbitrary function of the progeny of β
$p^c(a^\beta)$	target cond. $(x^\beta>0)$ prob. on a^β
$p^o(a^\beta)$	unpertrubed cond. $(x^\beta>0)$ prob. on a^β

Content-Sensitive Perturbation

$$\begin{aligned}p(I) &= p(I|x^\beta=0)p(x^\beta=0)+p(I|x^\beta>0)p(x^\beta>0)\\ &= p(I|x^\beta=0)p(x^\beta=0)+p(I,a^\beta|x^\beta>0)p(x^\beta>0)\\ &= p(I|x^\beta=0)p(x^\beta=0)+p(I|a^\beta,x^\beta>0)p^o(a^\beta)p(x^\beta>0)\end{aligned}$$

$$\begin{aligned}&\rightarrow p(I|x^\beta=0)p(x^\beta=0)+p(I|a^\beta,x^\beta>0)p^c(a^\beta)p(x^\beta>0)\\ &= p(I|x^\beta=0)p(x^\beta=0)+p(I|a^\beta,x^\beta>0)p^o(a^\beta)p(x^\beta>0)\frac{p^c(a^\beta)}{p^o(a^\beta)}\\ &= p(I)\left(\frac{p^c(a^\beta)}{p^o(a^\beta)}\right)^{1_{x^\beta>0}}\end{aligned}$$

Content-Sensitive Perturbation

…. but perturbations interact!

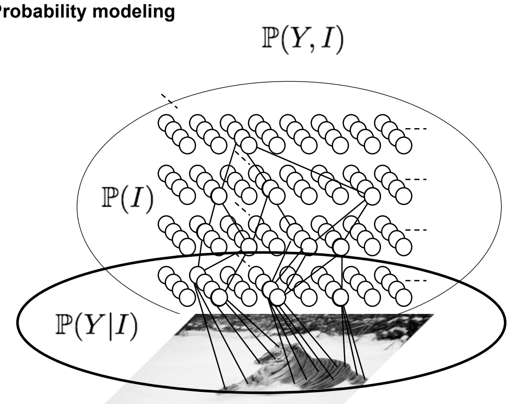
…. nevertheless …

THEOREM (Wei Zhang):

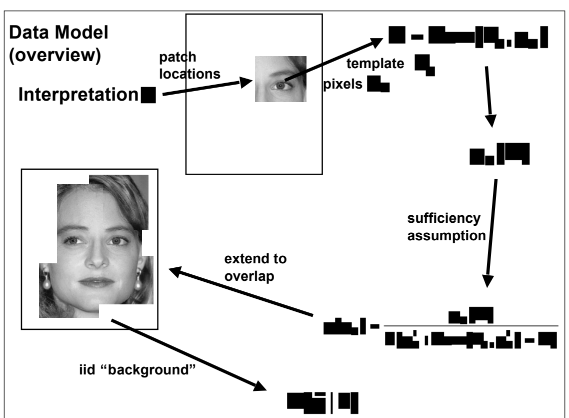
(1) if $p^c(a^\beta(I))>0\ \forall\ \beta\in B\ \&\ I\in\mathcal{I}$, then $\exists\ p(I)$ consistent with $\{p^c(a^\beta(I))\}_{\beta\in B}$

(2) iterative perturbation, over all $\beta\in B$, generates such a distribution.

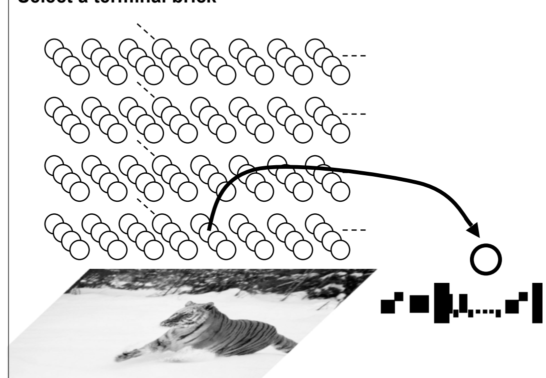
Probability modeling



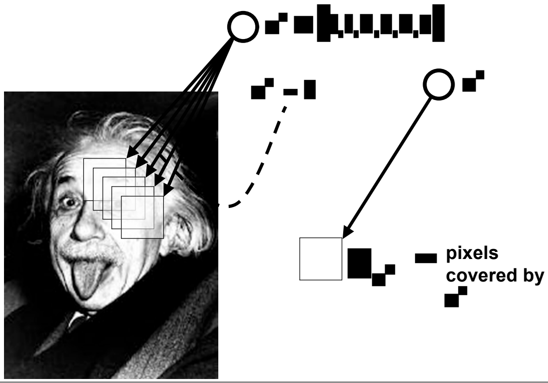
Data Model (overview)



Select a terminal brick



Select a state (e.g. left-eye brick)



Model

Template – one for each state of brick

Image data

Assume C = Corr(,) <u>sufficient:</u>

N(C) and the Central Limit Theorem

…intractable?

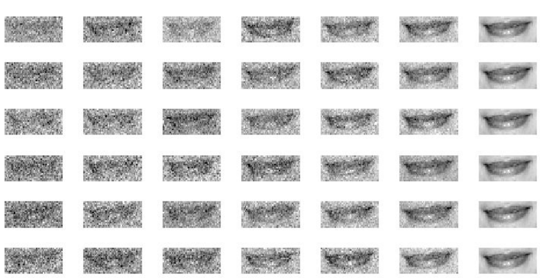
Then, *<u>exactly</u>*, if

CLT

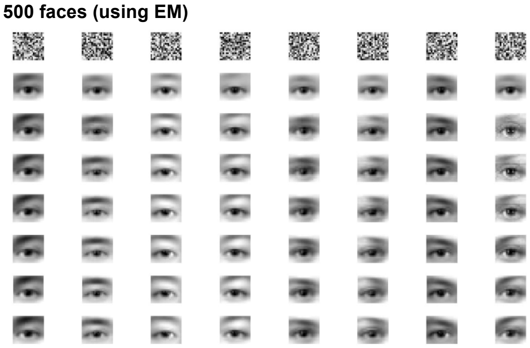
Hence

Example: sampling when = eye patch

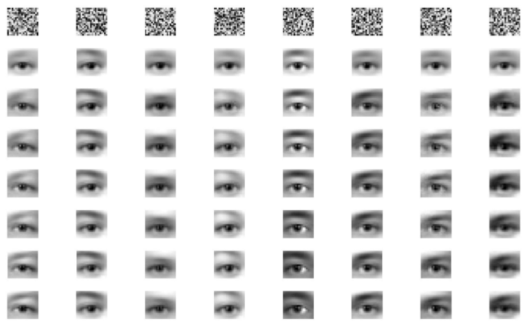
Example: sampling when = mouth patch



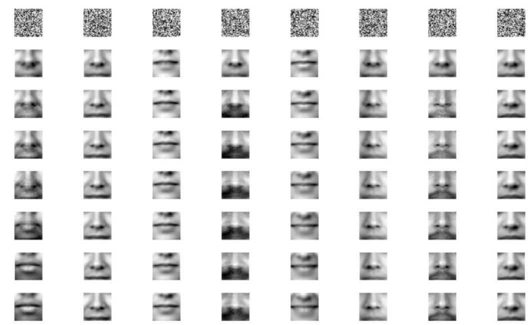
Example: learning a mixture of 16 eye patches from 500 faces (using EM)



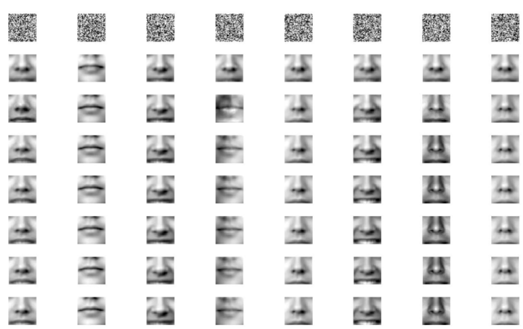
Example: learning a mixture of 16 eye patches from 500 faces (using EM)



Example: learning a mixture of 16 nose and mouth patches from 500 faces (using EM)



Example: learning a mixture of 16 nose and mouth patches from 500 faces (using EM)



- **I. Remarks on Computer Vision**
- **II. Approaches**
- **III. Bayesian Image Analysis**
- **IV. Probability Models**
- **V. Demonstration System: Reading License Plates**
- **VI. Generalization: Face Detection**

Test set: 385 images, mostly from Logan Airport

Courtesy of Visics Corporation

Architecture

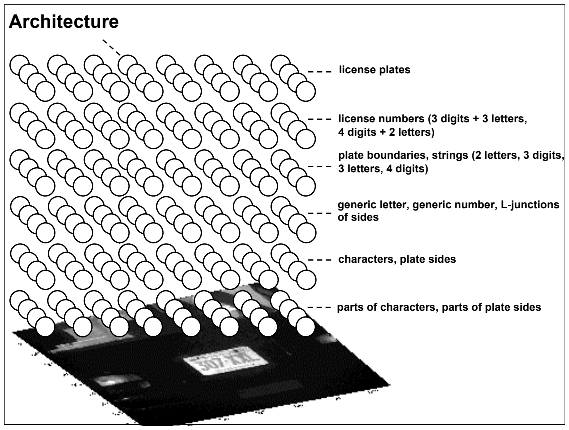


Image interpretation

Original Image

Top object

Top 10 objects


Top 25 objects

Image interpretation

Test image

Top objects

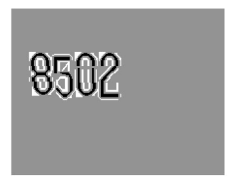
Performance

- 385 images
- Six plates read with mistakes (>98%)
- Approx. 99.5% characters read correctly
- Zero false positives

Efficient discrimination: Markov versus Content-Sensitive dist.

Sample 4-digit string


Markov backbone



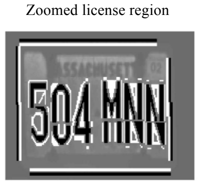
Perturbed distribution

Efficient discrimination: Markov versus Content-Sensitive dist.


Original image

Zoomed license region

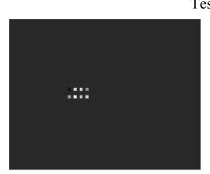

Top object under Markov distribution

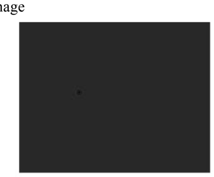


Top object under content-sensitive distribution

Efficient discrimination: testing objects against their parts

Test image


9 active “8” bricks under whole model


1 active “8” brick under parts model

Summary

Vision is Content Sensitive

Non-Markovian probability models

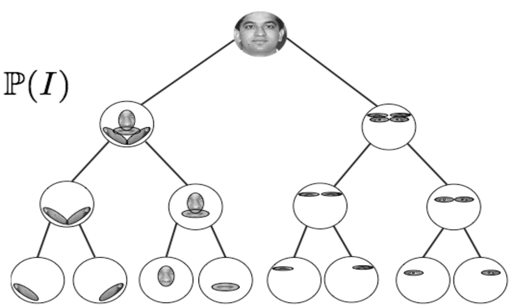
Background is Structured, and Made of the Same Stuff

Objects come equipped with their own background models

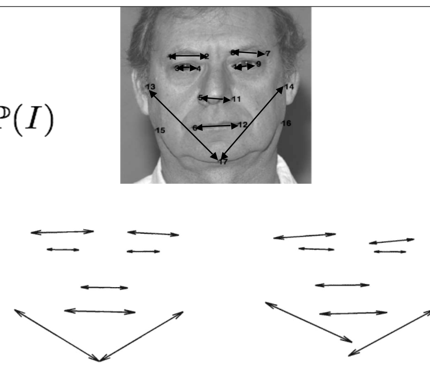
I. Remarks on Computer Vision
II. Approaches
III. Bayesian Image Analysis
IV. Probability Models
V. Demonstration System: Reading License Plates
VI. Generalization: Face Detection

Face Hierarchy

$\mathbb{P}(I)$

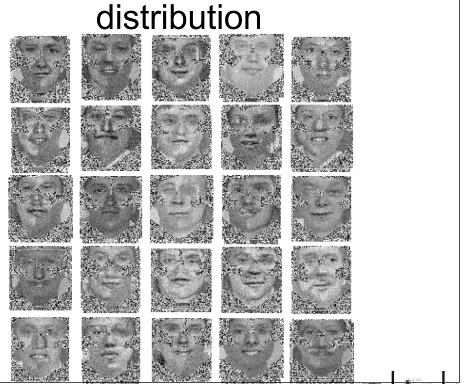


$\mathbb{P}(I)$



Sampling faces from the distribution

$\mathbb{P}(Y, I)$

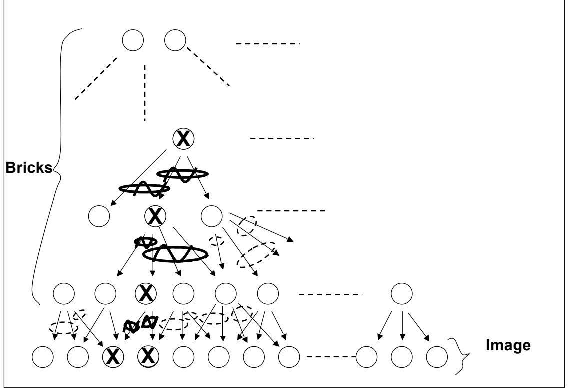


PATTERN SYNTHESIS

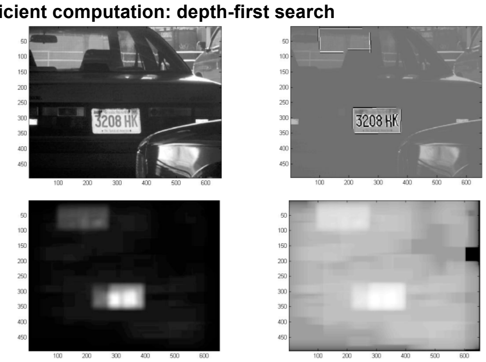
= PATTERN ANALYSIS

Ulf Grenander

Efficient computation: depth-first search

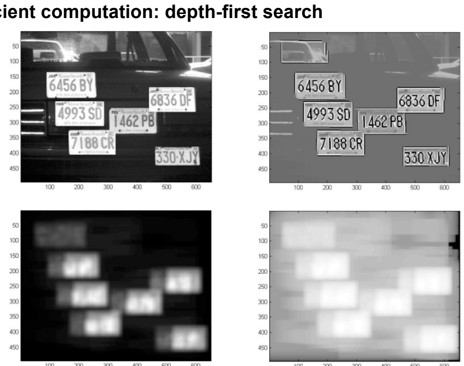


Efficient computation: depth-first search



Number of visits to each pixel. Left: linear scale Right: log scale

Efficient computation: depth-first search



Number of visits to each pixel. Left: linear scale Right: log scale