Computational Vision: Principles of Perceptual Inference

Daniel Kersten
Psychology, University of Minnesota

NIPS*98

http://vision.psych.umn.edu/www/kersten-lab/papers/NIPS98.pdf

Announcements

NIPS*98 Workshop on Statistical Theories of Cortical Function (Friday, December 4, 1998 : 7:30 am , Breckenridge)

IEEE Workshop on Statistical and Computational Theories of Vision: Modeling, Learning, Computing, and Sampling
June 22, 1999, Fort Collins, CO. (Yellow handout)

Yuille, A.L., Coughlan, J. M., and Kersten, D. Computational Vision: Principles of Perceptual Inference.
http://vision.psych.umn.edu/www/kersten-lab/papers/yuicouker98.pdf

Outline

Introduction: Computational Vision

- Context
- Working definition of Computational Vision
- History: Perception as inference

Theoretical framework

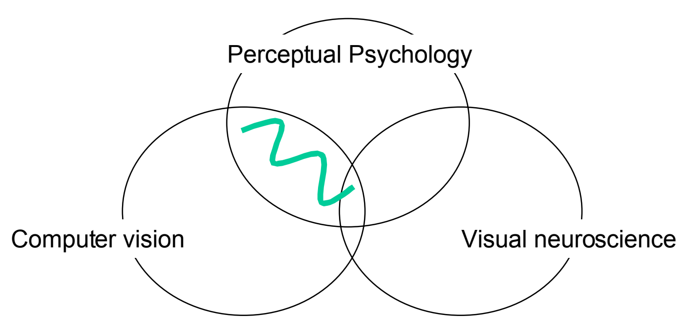
- Pattern theory
- Bayesian decision theory

Vision overview & examples

- Early: local measurements, local integration
- Intemediate-level: global organizational processes
- High-level: functional tasks

Computational Vision

Relation to Psychology, Computer Science, Neuroscience



Textbook References: (Ballard, & Brown, 1982; Bruce, Green, & Georgeson, 1996; Horn, 1986; Goldstein, 1995; Spillman, & Werner, 1990; Wandell, 1995)

Vision as image decryption

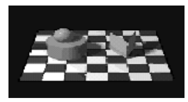
Visual tasks

Object recognition
Basic-level (shape)
Subordinate-level (shape, articulation, material)

Spatial layout
Viewer-object relations
Reach (viewpoint)
Grasp (shape)
Object-object relations
Events (relative position)
Scene recognition (relative position, material,lighting)

The (inverse) Inference problem

Image data, I



The image rendering problem I = F(S)

Viewpoint

Relative position

Illumination

Object shape, articulation, material

Challenges

Theoretical challenge

- Complexity of natural images, Inference for functional tasks

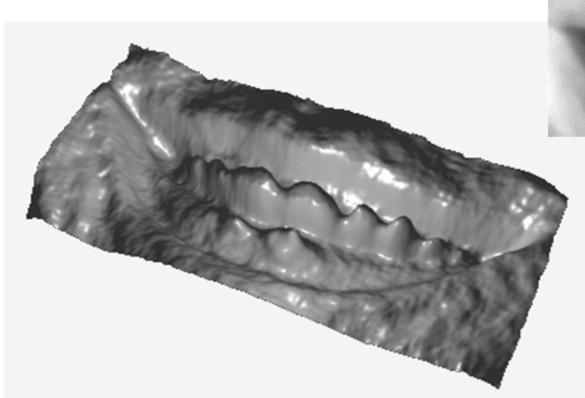
Empirical challenge

- Testing quantitative theories of visual behavior

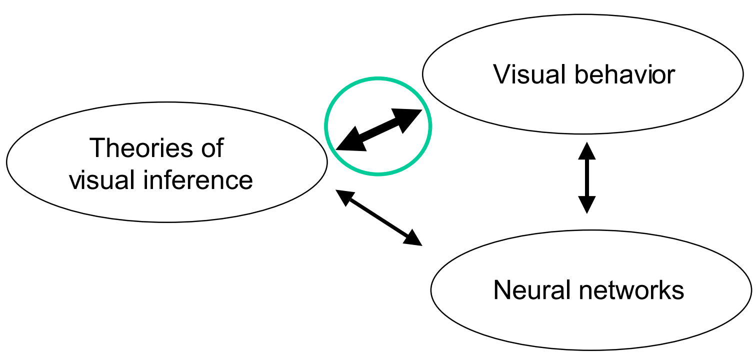
Proposed solution:

- Statistical theories of visual inference bridge perception and neural theories

Complexity of Natural Images



Computational Vision: Theories of inference and behavior



History of perception as statistical inference

Perception as inference

- Helmholtz (1867), Craik (1942), Brunswick (1952), Gregory (1980), Rock (1983)

1950’s & ‘60’s : Signal Detection Theory (SDT)

1970’s & ‘80’s : Vision is harder than expected

1950’s & ‘60’s : Signal Detection Theory (SDT)

External/physical limits to reliable decisions

Models of internal processes of perceptual decisions

Ideal observer analysis brings the two together

Limited to simple images, tasks

Ideal observer analysis

Brief history in visual psychophysics
- Quantum efficiency of light detection
  - Hecht et al. (1942), Barlow (1962)
- Pattern detection efficiency & simple cell receptive fields
  - Burgess et al (1981)., Watson et al. (1983), Kersten (1984)
- Perceptual organization, symmetry
  - Barlow & Reeves (1979)
- 3D object recognition efficiency.
  The informativeness of shading, edges, and silhouettes
  - Tjan et al. (1995), Braje et al. (1995)
- 3D object recognition and the problem of viewpoint
  - Liu et al., 1995

1970's & '80's : Computer Vision

Computer vision: Vision is harder than expected
- Marr program
  - Bottom-up
  - Levels of analysis (Marr)
    - **Qualitative** computational/functional theories
    - Algorithmic theories
    - Neural implementation theories

1970’s & ‘80’s : Computer Vision

Problems with Marr program:

– Bottom-up difficulties
  - Segmentation, edge detection difficult
  - Early commitment, uncertainty

– Levels of analysis
  - Still debating

1970’s & ‘80’s : Computer Vision

Solutions

– **Confidence-driven** processing

– **Quantitative** computational theory of statistical inference--- in the spirit of SDT
  - Extend SDT, “ideal observer” to handle natural image patterns, tasks

Extending SDT

Signals are not simple functions image intensities

Useful information is confounded by more than noise.

Natural images are not linear combinations of relevant signals

Extending SDT

Variables of interest are rarely Gaussian

Perception involves more than classification

Most of the interesting perceptual knowledge on priors and utility is implicit

Have SDT:	$I = P + \text{noise}$
Need:	$I = f(P_1, P_2, \ldots; S_1, S_1, \ldots)$

Pattern theory

Emphasis on decryption: Analysis by synthesis

Generative modeling

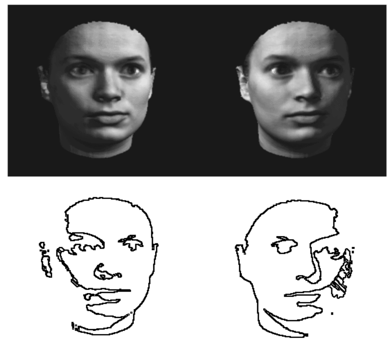
– References: (Cavanagh, 1991; Dayan, Hinton, Neal, & Zemel, 1995); Grenander, 1993; Grenander, 1996; Grossberg, Mingolla, & Ross, 1997; Hinton, & Ghahramani, 1997; Jones, Sinha, Vetter, & Poggio, 1997; Kersten, 1997; Mumford, 1995; Ullman, 1991)

Pattern theory

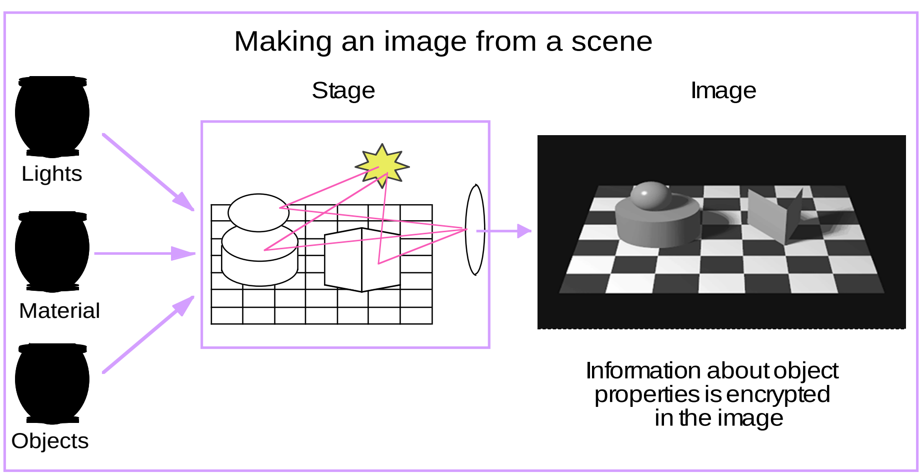
Synthesis/generative modeling

- Example illustrating need: Mooney pictures and edge classification
- Modeling underlying causes
  - Computer vision: Inverse graphics & computer graphics
  - Pattern theory approach, learning

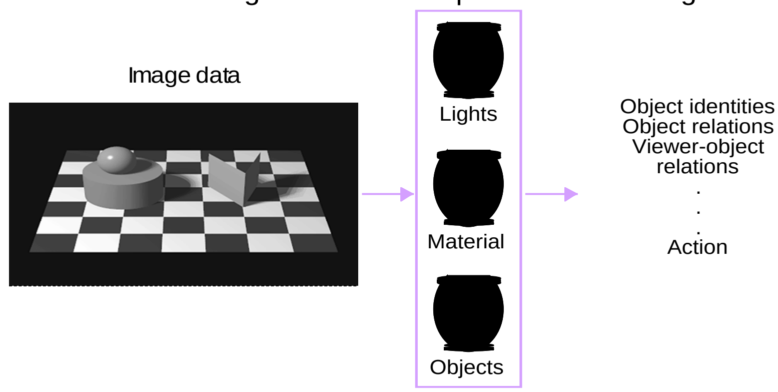
“Mooney face”



Edge ambiguity



Inverse computer graphics:
Constructing a scene description from an image

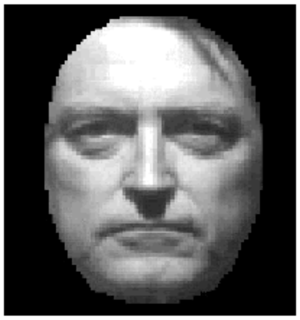


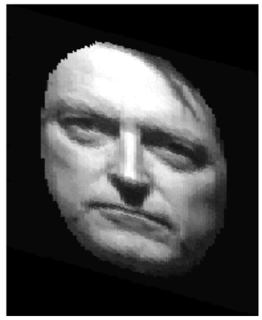
Pattern Theory

Types of transformation in natural patterns (Grenander, Mumford)

- Blur & noise
- Processes occur over multiple scales
- Domain interruption, occlusion
- Domain warps

Types of transformations

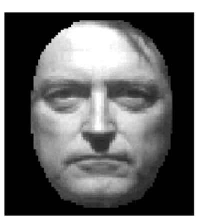


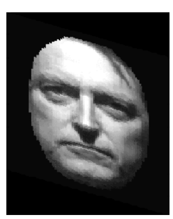


Domain
warping

Warping
+
occlusion

Types of transformations



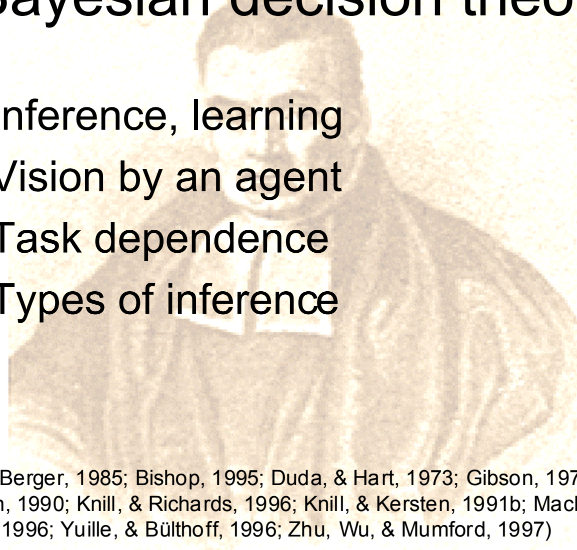


Domain
warping

Warping
+
occlusion

Superposition,
warping
+
occlusion

Bayesian decision theory



Inference, learning

Vision by an agent

Task dependence

Types of inference

References: (Berger, 1985; Bishop, 1995; Duda, & Hart, 1973; Gibson, 1979; Jordan, & Bishop, 1996; Kersten, 1990; Knill, & Richards, 1996; Knill, & Kersten, 1991b; MacKay, 1992; Rissanen, 1989; Ripley, 1996; Yuille, & Bülthoff, 1996; Zhu, Wu, & Mumford, 1997)

Bayes: Analysis & synthesis

Information for inference
- Prior
- Likelihood

Learning & sampling
- Density estimation
  - P(S,I)
    - => P(S|I), through marginalization & conditioning
- Bayes nets, MRFs

Bayesian Analysis

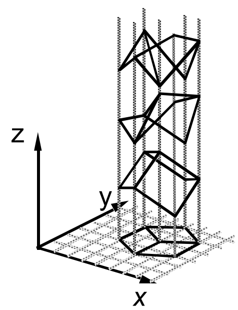
Characterize posterior probability, *P(S | I,C),* using Bayes' rule:

$$p(S \mid I, C) = \frac{p(I \mid S, C)\,p(S \mid C)}{p(I \mid C)} = \frac{p(I \mid S, C)\,p(S \mid C)}{\sum_{S'} p(I \mid S', C)\,p(S' \mid C)}$$

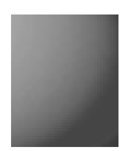
P(S|C) **prior probability for *S***
P(I| S,C) **likelihood from model of image formation**
P(I|C) **"evidence" for category or "model"**

Problems of ambiguity

Many 3D shapes can map to the same 2D image



The scene causes of local image intensity change are confounded in the image data



Scene causes of intensity change

 ?

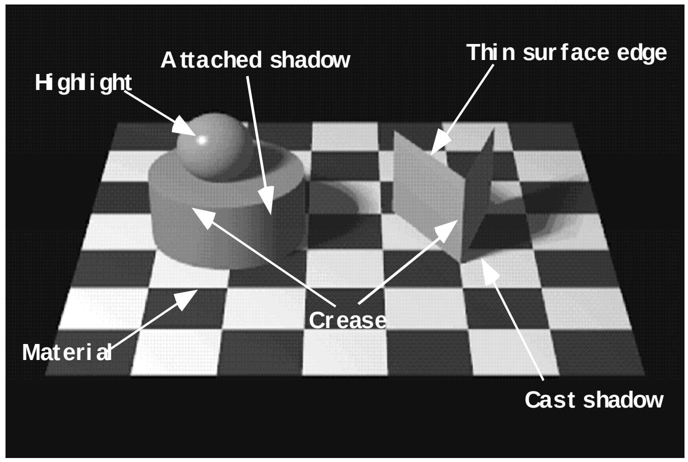
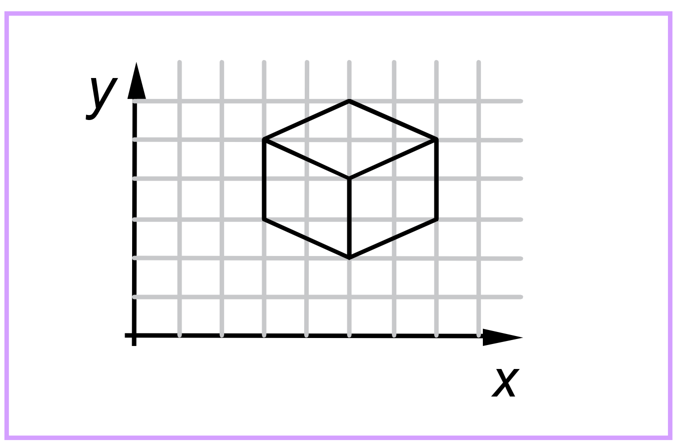
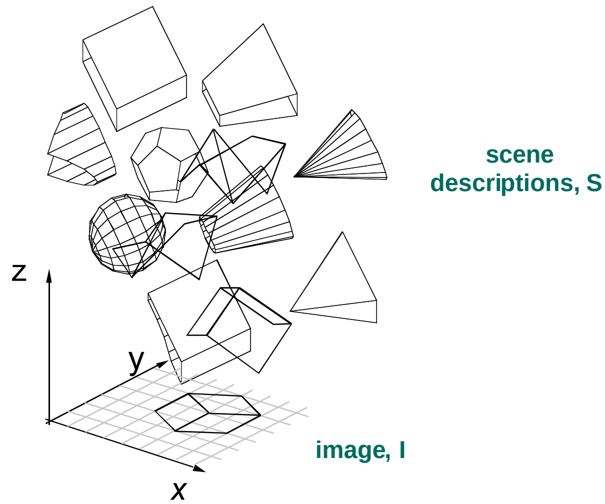


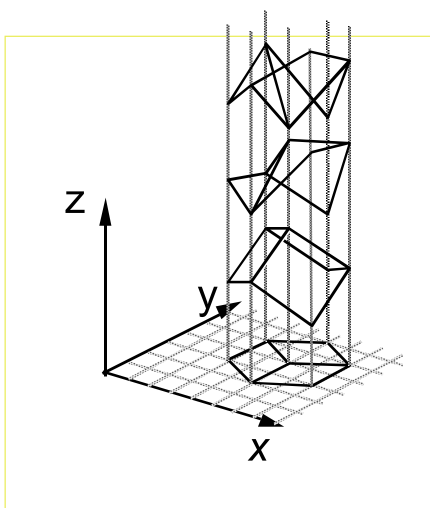
Image data



Which scene descriptions are likely to give rise to the image data?



Likelihood selects subset of scene descriptions consistent with image data



Likelihood $p(image|scene)$

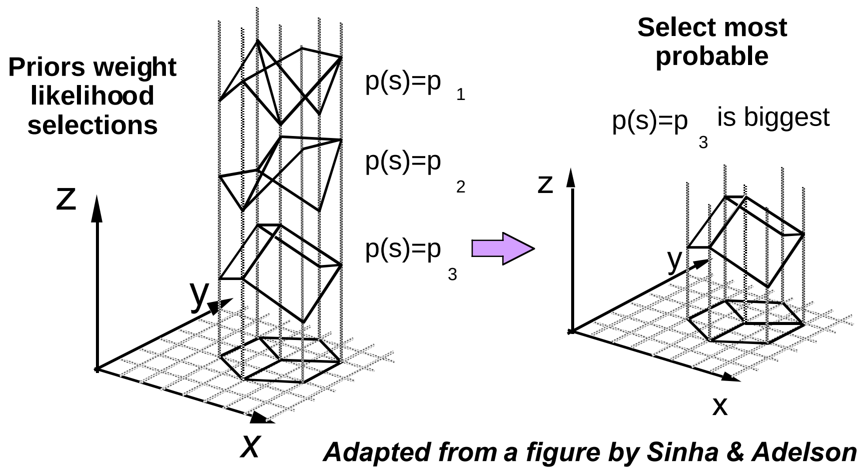
e.g. for additive noise

$p(image|scene) = p(image - f(scene))$

for no noise

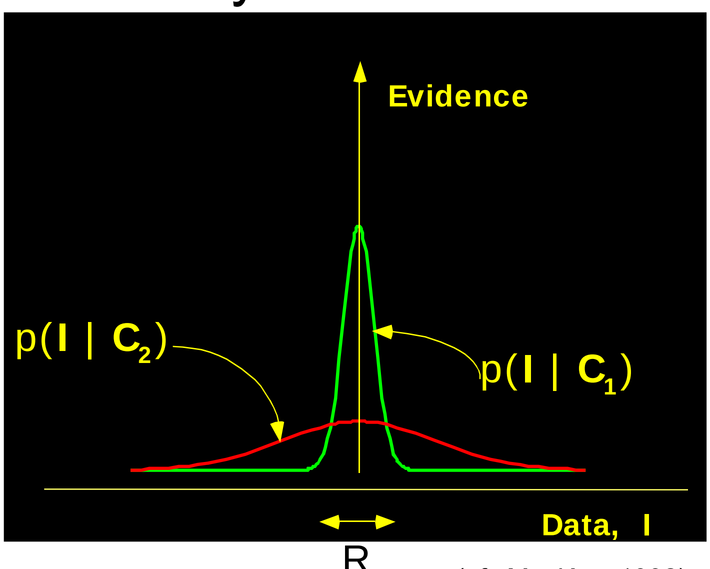
$p(image|scene) = \delta(image - f(scene))$

Prior further narrows selection



Adapted from a figure by Sinha & Adelson

Bayes & Occam



(cf. MacKay, 1992)

Bayes, Shannon & MDL

length(code(I)) = $-\log_2 p(I)$

length(code(I, s)) = length(code(s)) + length(code(I using s))

Bayes: Decision theory

Vision by an agent
- Loss functions, risk
- Marginalization

Task dependence for visual tasks
- Sample taxonomy: recognition, navigation, etc..

Visual inference tasks
- Inference: classification, regression
- Learning: density estimation

Vision by an agent

Loss functions, risk

$$R(A;I) = \sum_{S} L(A,S)P(S \mid I)$$

Special case: Maximum a posteriori estimation (MAP)

$$L(A,S) = \begin{cases} -1 & \text{if } A = S \\ 0 & \text{otherwise} \end{cases}$$

$$R(A;I) = -P(A \mid I)$$

Find **A** to maximize: *P*(**A|I**)

Marginalize over generic scene parameters

Two types of scene parameters

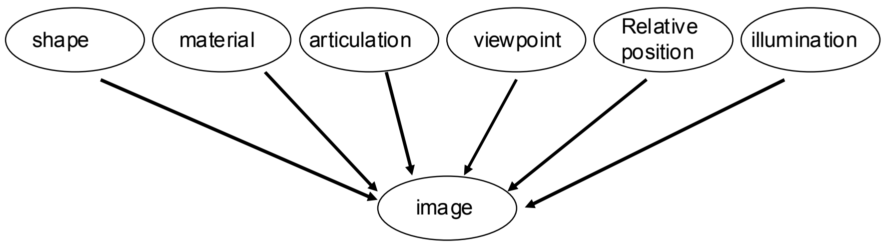
Scene variables that are important to know, S_m

Generic variables that contribute to the image, but do not need to be explicitly estimated, S_g

$$p(S_m \mid I,C) = \int p(S_m, S_g \mid I,C)\,dS_g$$

$$p(S_m \mid I,C) \propto \int p(I \mid S_m, S_g, C)\,dS_g$$

Perception's model of the image should be robust over variations in generic variables



Task dependency: explicit and generic variables

I=f(shape, material, articulation,viewpoint,relative position, illumination)

	Object perception		**Spatial layout**		
	Object-centered (object recognition)		World-centered	Observer-centered (hand action)	
	Entry-level	*Subordinate-level*	*Planning*	*Reach*	*Grasp*
Shape	E	E	G	G	G
Material	G	E	G	G	G
Articulation	G	E	G	G	E
Viewpoint	G	G	G	E	G
Relative position	G	G	E	G	G
Illumination	G	G	G	G	G

Explicit (E) = Primary

Generic (G) = Secondary

Learning & sampling

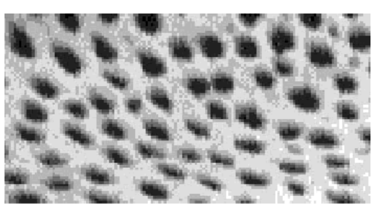
Density estimation

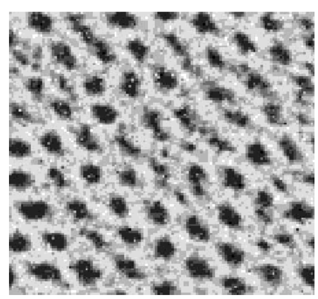
P(S,I) => P(S|I), through marginalization & conditioning

Bayes nets, Markov Random Fields

– Image coding

Learning & sampling





Zhu, Wu, Mumford, 1997

Working definition of perception

Revisionist Helmholtz

"Perception is (largely) unconscious statistical inference involving unconscious priors and unconscious loss functions"

Working definition of perception

Revisionist Marr

- Levels of analysis
  - Qualitative computational/functional theories
  - **Quantitative theories of statistical inference**
  - Algorithmic theories
  - Neural implementation theories
- Confidence-driven perceptual decisions

Early vision

Two Definitions:

- Local image measurements, especially those related to surface properties
  - Utility--task assumptions
- Efficient image coding
  - Utility neutral--information preserving

Local measurements, local integration

Change detection

- Types
  - » Intensity edges
  - » Color
  - » Motion
  - » Stereo
  - » Texture
- Adelson & Bergen's (1991) plenoptic function
  - » $P(x,y,t,\lambda,V_{x,} V_{y,} V_z)$, derivatives

References: (Adelson, & Bergen, 1991; Belhumeur, & Mumford, 1992; Blake, Bulthoff, & Sheinberg, 1992a; Bulthoff, 1991; Buelthoff, 1991; Freeman, 1994; Geman, & Geman, 1984; Heeger, Simoncelli, & Movshon, 1996; Julesz, 1984; Knill, 1998b) (Knill, 1998a; Malik, & Perona, 1990; Poggio, Torre, & Koch, 1985; Schrater, Knill & Simoncelli, submitted; Simoncelli, & Heeger, 1998; Szeliski, 1989; Yuille, & Grzywacz, 1988; Yuille, Geiger, & Bülthoff, 1991)

Surface perception—local constraints on smoothing

Ill-posed problems & regularization theory

(Poggio, Torre & Koch, 1985)

$$I = \mathbf{A}S$$

$$E = (I - \mathbf{A}S)^T (I - \mathbf{A}S) + \lambda S^T \mathbf{B} S$$

$$p(I \mid S) = k \times \exp\left[-\frac{1}{2\sigma_n^2}(I - \mathbf{A}S)^T (I - \mathbf{A}S)\right] \qquad p(S) = k' \times \exp\left[-\frac{1}{2\sigma_s^2} S^T \mathbf{B} S\right]$$

MRFs (Geman & Geman, 1984)

Measurements for segmentation, depth, orientation, shape

- Stereo (e.g. Belhumeur & Mumford, 1992)
- Shape-from-X (e.g. Bülthoff, 1991; Mamassian & Landy, 1998; Freeman, 1994; Blake, Buelthoff, Sheinberg, 1996)
  - Contours, shading, texture
- Orientation from texture (e.g.; Knill, 1998a, 1998b)
- Motion (e.g. Yuille & Grzywacz,1988; Schrater, Knill, & Simoncelli, submitted; Simoncelli, Heeger & Movshon, 1998)
  - **Motion, aperture problem**
    - **Weiss & Adelson (1998) ; Heeger & Simoncelli (1991)**

Motion

Slow & smooth: A Bayesian theory for the combination of local motion signals in human vision, Weiss & Adelson (1998)

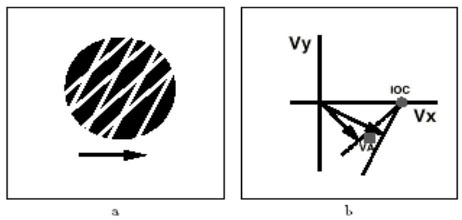
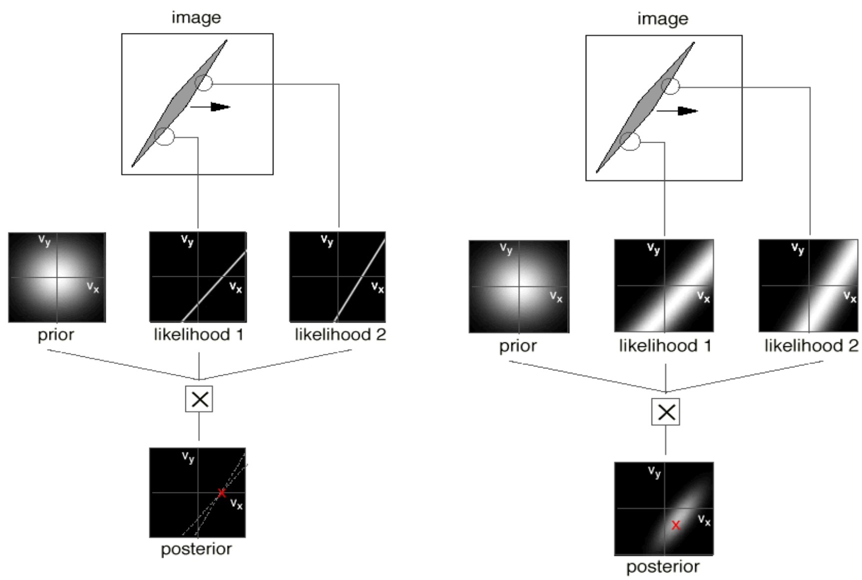


Figure from: Weiss & Adelson, 1998

Weiss: Friday morning NIPS*98 Workshop

Show Weiss & Adelson video



Figures from: Weiss & Adelson, 1998

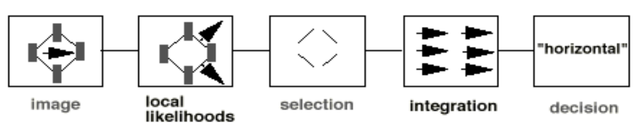


Figure from: Weiss & Adelson, 1998

Extensions (Weiss & Adelson, 1998)

Base likelihoods on actual image data

- spatiotemporal measurements

Include “2D” features

- E.g. corners

Rigid rotations, non-rigid deformations

Local likelihood: $L(v) \propto e^{-\sum_r w(r)(I_x v_x + I_y v_y + I_t)^2 / 2\sigma^2}$

Global likelihood: $L_r(v) \to p(I \mid \theta) \propto \prod_r L_r(\theta)$

Prior: $P(V) \propto e^{-\sum_r (Dv)^t (r)(Dv)(r)/2}$

$P(V) \to P(\theta)$

Posterior: $P(\theta \mid I) \propto P(I \mid \theta) P(\theta)$

From: Weiss & Adelson, 1998

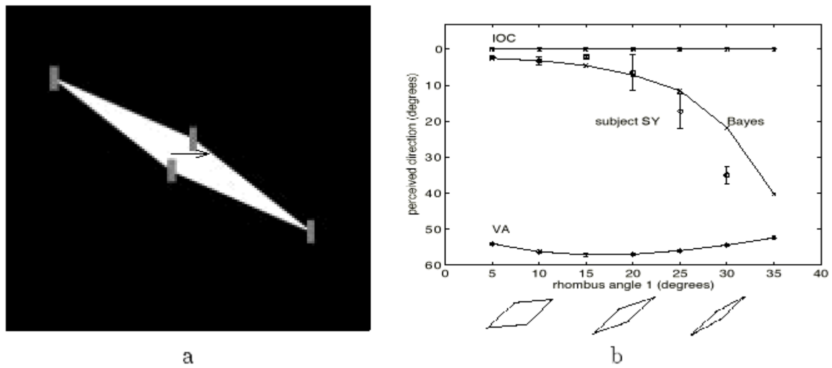


Figure from: Weiss & Adelson, 1998

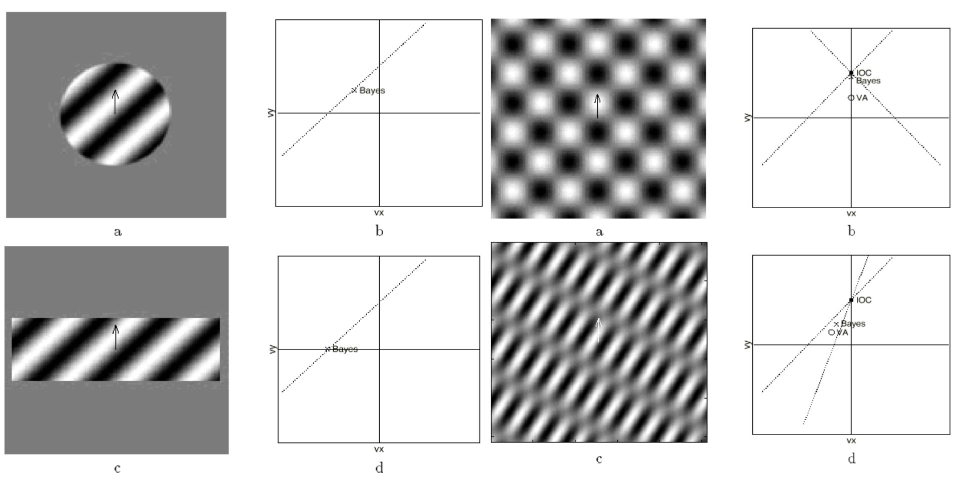


Figure from: Weiss & Adelson, 1998

Efficient coding

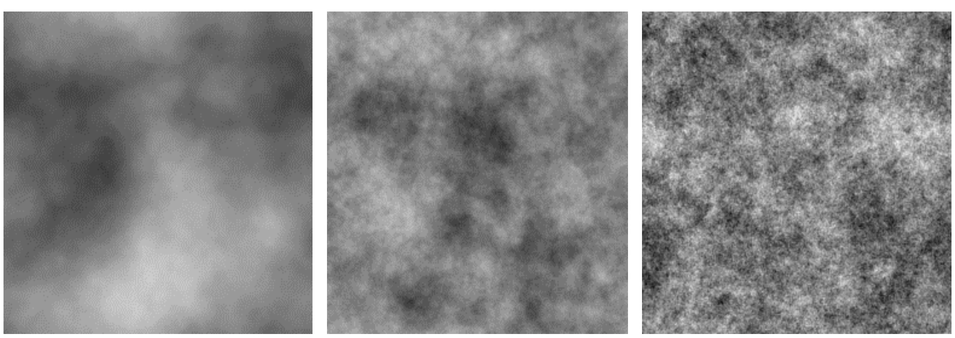
Natural image statistics

- Shannon’s guessing game (Kersten, 1986)
- Tuning of the human visual system, 2nd order statistics (Knill, Field & Kersten, 1990)

Redundancy reduction - Barlow (1959)

- Decorrelation, PCA
  - Olshausen & Field (1996)
  - Simoncelli, Heeger
- Minimum entropy, factorial codes, ICA
  - Bell & Sejnowski, 1995

Tuning of human vision to the statistics of images: Fractal image discrimination



Knill, Field & Kersten, 1990

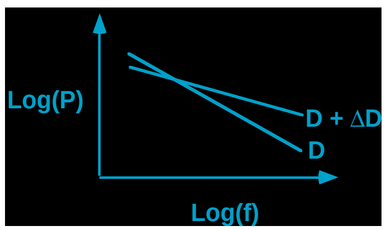
Fractal image discrimination

How well is the human visual system tuned to the correlational structure of images?

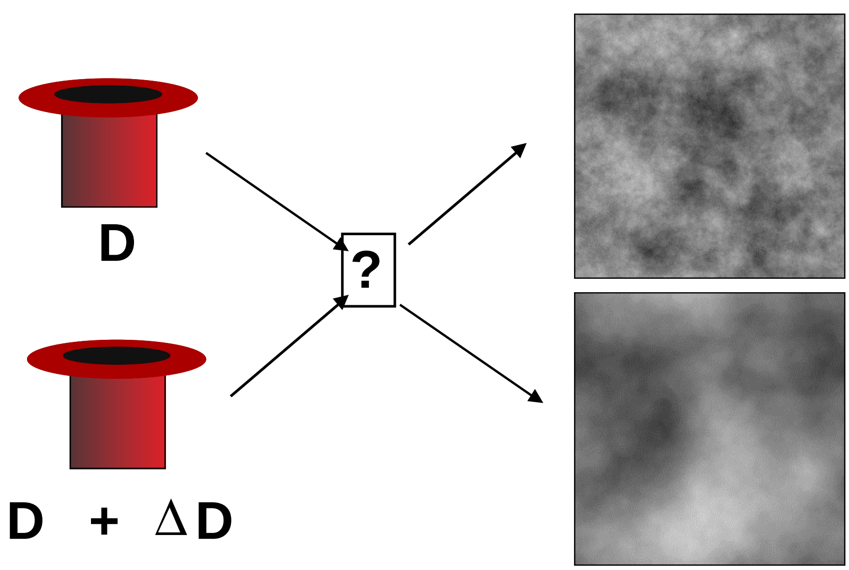
Scale invariant subset of class of images defined by their correlation function:

Random fractals:

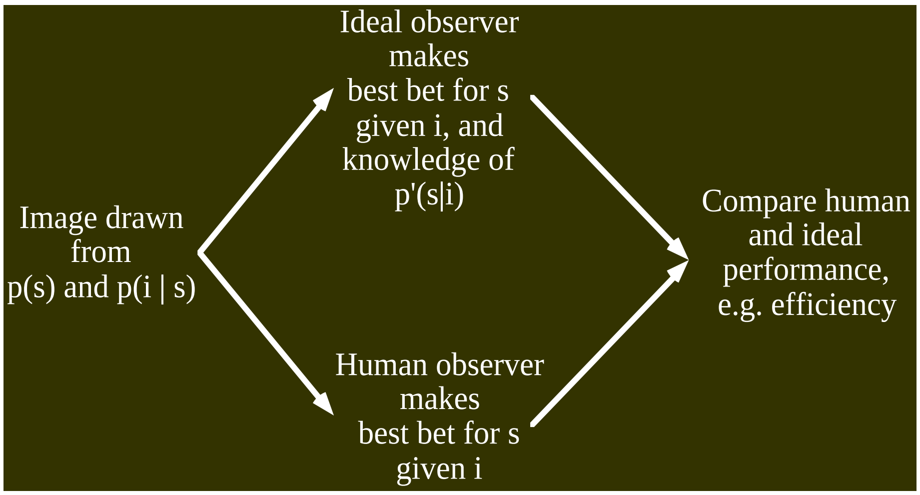
Log(power spectrum) = (2D - 8) Log (spatial frequency)



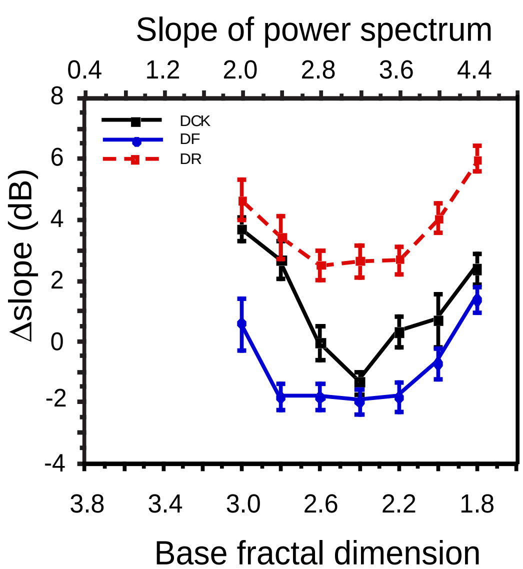
Fractal image discrimination - the task



Ideal observer analysis



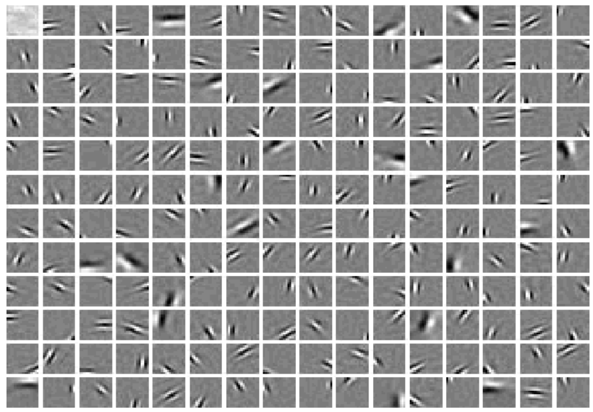
Human fractal image discrimination



Statistical efficiency

$$E = \frac{\Delta D_I^2}{\Delta D_H^2} \approx 10\%$$

Sparse coding



$$\sum_{x,y}\left[I(x,y)-\sum_i a_i\,\phi_i(x,y)\right]^2 + \sum_i S\left(\frac{a_i}{\sigma}\right)$$

Figure from: Olshausen & Field, 1996

Efficient coding: Image density estimation

Learning & density estimation
- PCA, ICA

➡ **Minimax entropy learning**
- Zhu, Wu, Mumford (1997)

Minimax entropy learning

Maximum entropy to determine $p_M(I)$ which matches the measured statistics, but is "least committal"

$$\{\phi_i(\mathbf{I}) : i = 1, \ldots, N\}$$

$$\sum_{\mathbf{I}} p_M(\mathbf{I})\phi_i(\mathbf{I}) = \psi_i, \quad \text{for } i = 1, \ldots, N$$

$$p_M(\mathbf{I}) = \frac{1}{Z[\lambda]} \exp\{-\sum_{i=1}^{N} \lambda_i \phi_i(\mathbf{I})\},$$

Minimum entropy to determine statistics/features

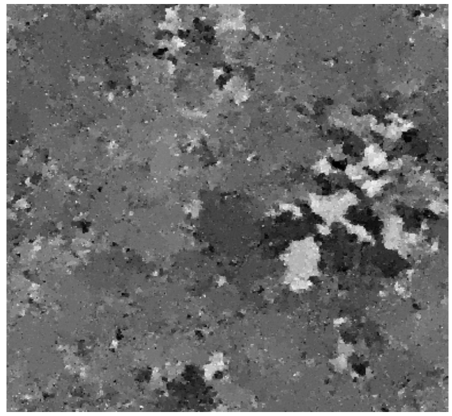
$$\sum_{\mathbf{I}} p(I) \log p_M(\mathbf{I}) = \sum_{\mathbf{I}} p_M(\mathbf{I}) \log p_M(\mathbf{I})$$

=> $D(p(\mathbf{I}) \, | p_M(\mathbf{I})) = \text{entropy}(p_M(\mathbf{I})) - \text{entropy}(p(\mathbf{I}))$

Minimax entropy learning

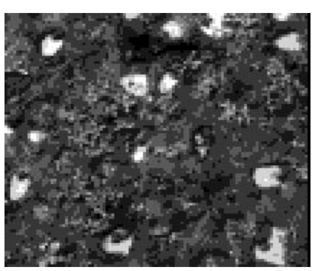
- Feature pursuit
- Examples
  - – Generic prior
  - – Class-specific priors

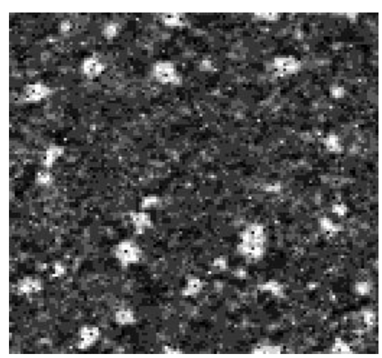
Generic natural image prior



Courtesy: Song Chun Zhu   Zhu & Mumford, IEEE PAMI

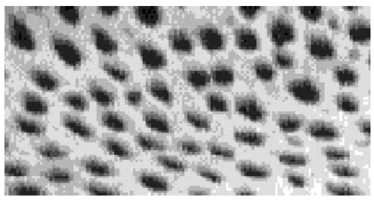
Class-specific prior - “Mud”

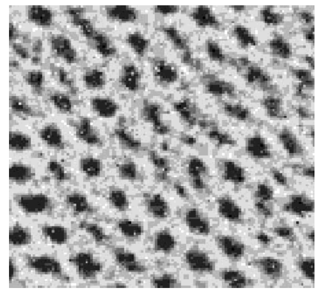




Courtesy: Song Chun Zhu   Zhu, Wu & Mumford, 1997

Class-specific prior: Cheetah





Zhu, Wu, Mumford, 1997

Relation to the brain?

New density estimation tools to test hypotheses of human image coding

- Efficiency of human processing of generic & class-specific textures

See Eero Simoncelli’s talk tomorrow 8:30 am

Break

Introduction: Computational Vision
- Context
- Working definition of Computational Vision
- History: Perception as inference

Theoretical framework
- Pattern theory
- Bayesian decision theory

Vision overview & examples
- Early: local measurements, local integration, efficient coding
- Intemediate-level: global organizational processes
- High-level: functional tasks

Announcements

NIPS*98 Workshop on Statistical Theories of Cortical Function (Friday, December 4, 1998 : 7:30 am , Breckenridge)

IEEE Workshop on Statistical and Computational Theories of Vision: Modeling, Learning, Computing, and Sampling

June 22, 1999, Fort Collins, CO. (Yellow handout)

Yuille, A.L., Coughlan, J. M., and Kersten, D. Computational Vision: Principles of Perceptual Inference.

http://vision.psych.umn.edu/www/kersten-lab/papers/yuicouker98.pdf

Intermediate-level vision

Generic, global organizational processes

- Domain overlap, occlusion
- Surface grouping, selection
- Gestalt principles

Cue integration

Cooperative computation

Attention

References: (Adelson, 1993; Brainard, & Freeman, 1994; Bülthoff, & Mallot, 1988; Bülthoff, et al., 1988; Clark & Yuille, 1990; Darrell, Sclaroff, & Pentland, 1990; Darrell, & Pentland, 1991; Darrell, & Simoncelli, 1994; Jacobs, Jordan, Nowlan, & Hinton, 1991; Jepson, & Black, 1993; Kersten, & Madarasmi, 1995; Jordan, & Jacobs, 1994; Knill, & Kersten, 1991a; Knill, 1998a; Landy, Maloney, Johnston, & Young, 1995; Maloney, & Landy, 1989; Mamassian, & Landy, 1998; Bülthoff, & Yuille, 1991; Weiss, & Adelson, 1998; Sinha, & Adelson, 1993; Nakayama, & Shimojo, 1992; Wang, & Adelson, 1994; Young, Landy, & Maloney, 1993; Weiss, 1997; Yuille, et al., 1996; Yuille, Stolorz, & Ultans, 1994; Zucker, & David, 1988)

Cue /information integration

Weak, strong coupling (Bülthoff & Yuille, 1996)

Robust statistics (Maloney & Landy, 1989)

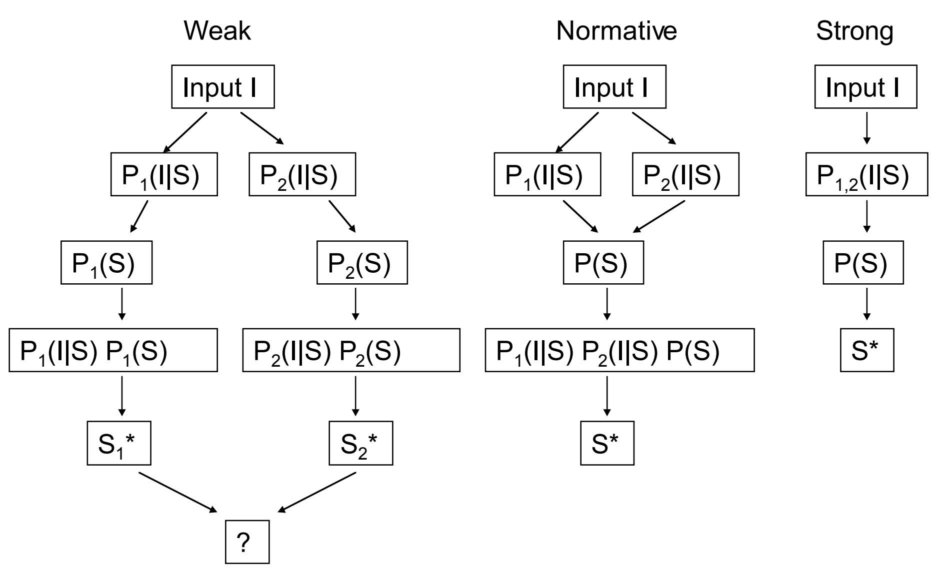
Depth, orientation, shape

- Orientation from texture (Knill, 1998)
  - Confidence-driven cue utilization
- Shape from texture, (Blake, Bülthoff, & Sheinberg 1992a)
  - Cramer-Rao

Cooperative computation

Density Mixtures (e.g. Jacobs, Jordan, Nowlan, Hinton, 1991; Jordan & Jacobs, 1994)

"Strong coupling", Competitive priors (Yuille & Bülthoff, 1996)



Yuille & Bülthoff, 1996

Cooperative computation

Color & illumination (e.g. Brainard & Freeman)

Occlusion, surfaces and segmentation

- Nakayama, Shimojo,1992
- Layers (Darrell & Pentland, 1992; Kersten & Madarasmi, 1995)

Motion segmentation, layers, mixtures

– Selection for smoothing (e.g. Jepson & Black, 1993; Weiss, 1997; Motion, aperture problem revisited)

Shape, reflectance, lighting

– Knill & Kersten (1991); Adelson (1993)

Cooperative computation: Shape, reflectance, lighting

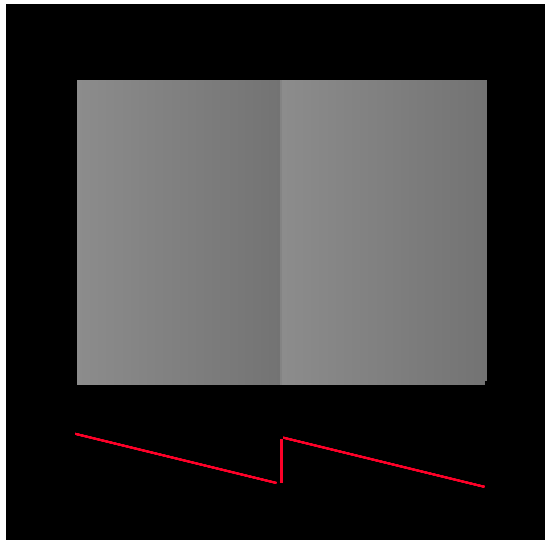
Land & McCann

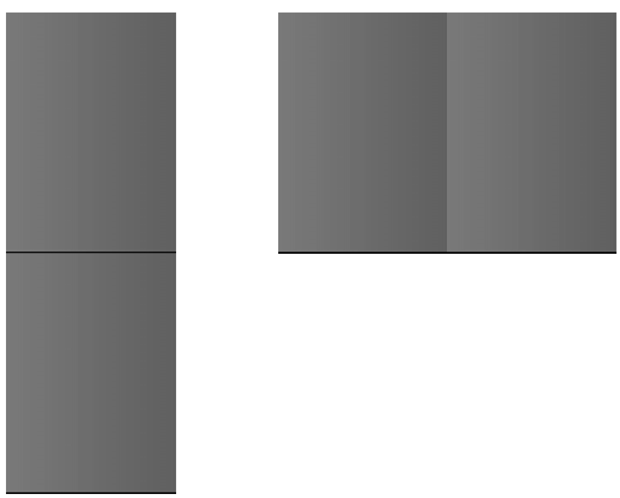
Filter explanation

Shape affects lightness

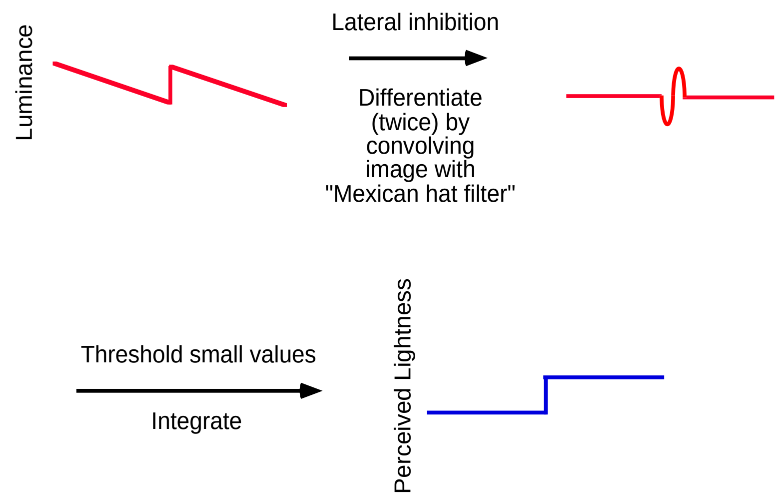
Inverse graphics explanation

Land & McCann's lightness illusion

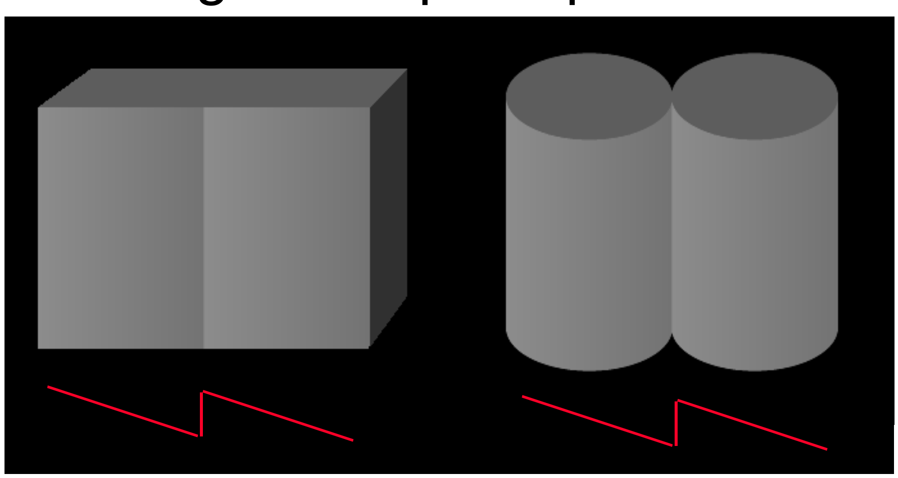




Neural network filter explanation



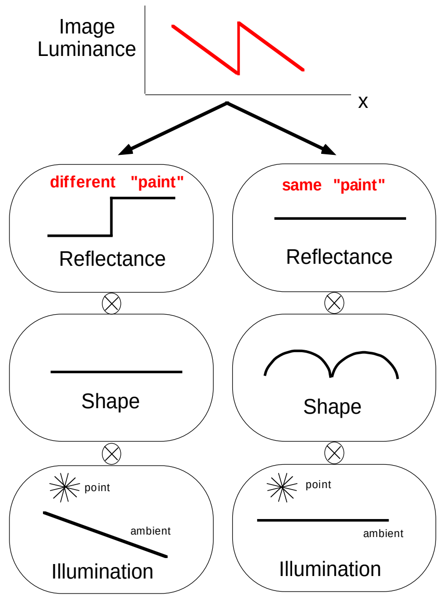
Apparent surface shape affects lightness perception



Knill & Kersten (1991)

Inverse graphics solution

What model of material reflectances, shape, and lighting fit the image data?



Shape and lightness

Functional or "inverse graphics" explanation

- luminance gradients can be caused by smooth changes in shape or smooth changes in illumination

Mechanism

- NOT a simple neural network filter
- Looks more like "inverse 3D graphics"
  - cooperative interaction in the estimation of shape, reflectance, and illumination
  - much of the machinery in the cortex

High-level vision

Functional tasks, viewer-object relations, object-object relations

- Manipulation
- Navigation
- **Spatial layout**
- **Recognition**

References: (Amit, Geman, & Jedynak, 1997b; Amit, & Geman, 1997a; Belhumeur, Hespanha, & Kriegman, 1997; Belhumeur, & Kriegman, 1996; Biederman, 1987; Blake, & Yuille, 1992b; Bobick, 1987; Bülthoff, Edelman, & Tarr, 1995; d'Avossa, & Kersten, 1993; Hallinan, 1994; Geman, & Jedynak, 1993; Heeger, & Jepson, 1990; Kersten, Mamassian & Knill, 1997; Kersten, Knill, Mamassian, Bülthoff, 1996; Langer, & Zucker, 1994; Legge, Klitz, & Tjan, 1997; Liu, Knill, & Kersten, 1995; Liu, & Kersten, 1998; Osuna, Freund, & Girosi, 1997; Tarr, Kersten, & Buelthoff, 1997; Thorpe, Fize, & Marlot, 1996; Tjan, Braje, Legge, & Kersten, 1995; Tjan, & Legge, 1997; Poggio, & Edelman, 1990; Schölkopf, 1997; Ullman, 1996; Ullman, & Basri, 1991; Wolpert, Ghahramani, & Jordan, 1995; Zhu, & Yuille, 1996)

Manipulation

Reach & grasp

- Kalman filtering (e.g. Wolpert, Ghahramani, Jordan, 1995)

Navigation, layout

- Direction of heading
  - – Optic flow: Separating rotational from translational components
    - (e.g. Heeger & Jepson, 1990)
    - Translational component
      - – Cramer-Rao bounds (d'Avossa & Kersten, 1993)
- Orienting/planning

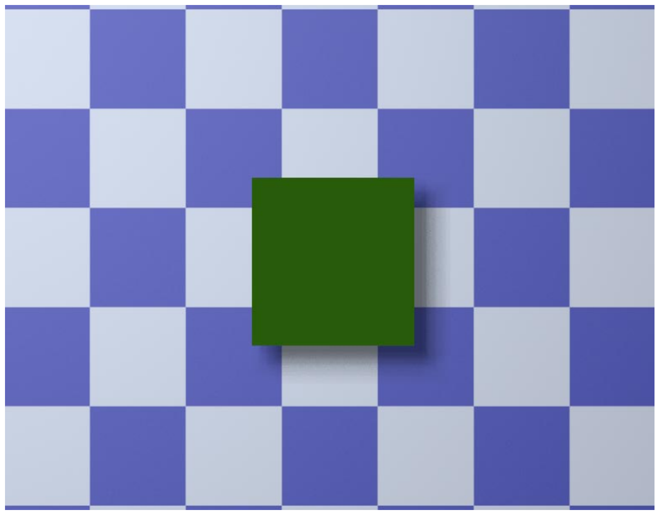
Spatial layout

– Relative object depth/trajectory from shadows

– Qualitative Bayesian analysis

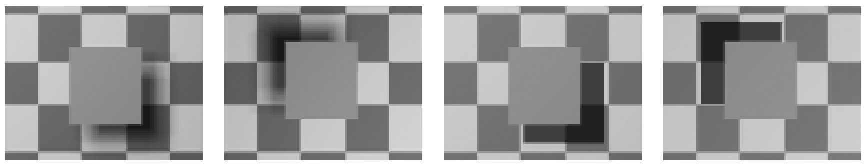
Square-over-checkerboard

Depth change from cast shadows



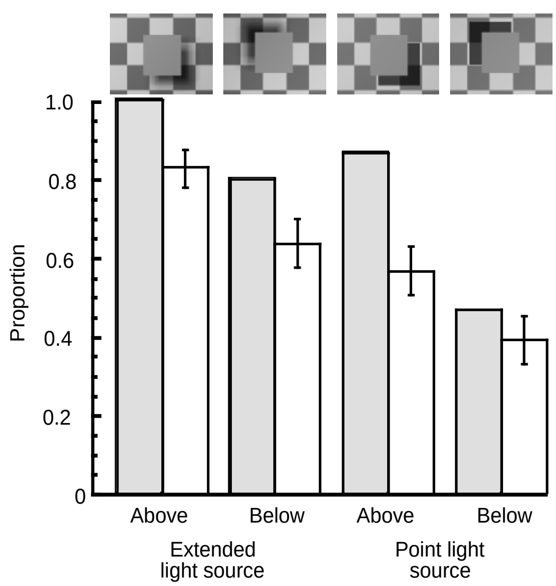
Show shadow video

Shadow motion
vs.
object image motion



http://vision.psych.umn.edu/www/kersten-lab/shadows.html

Kersten, D., Knill, D. C., Mamassian, P. and Bülthoff, I. (1996)



“Square-over-checkerboard” Summary of results

Light from above is better than from below

Dark shadows are better than light

Extended light sources lead to stronger depth illusion

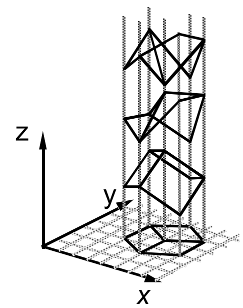
Knowledge required to resolve ambiguity

Piece together a scene model of explicit variables subject to:

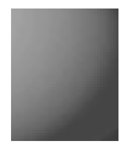
- Consistency with image data
- Prior probabilities
- Robustness over generic variables

Problems of ambiguity

Many 3D shapes can map to the same 2D image

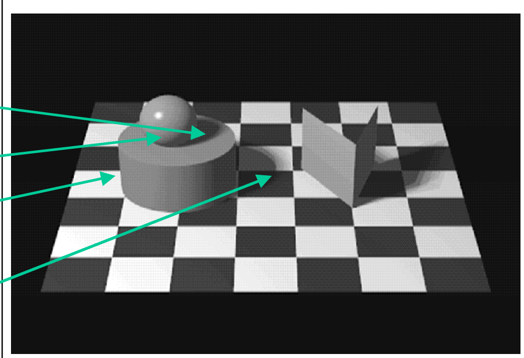


The scene causes of local image intensity change are confounded in the image data

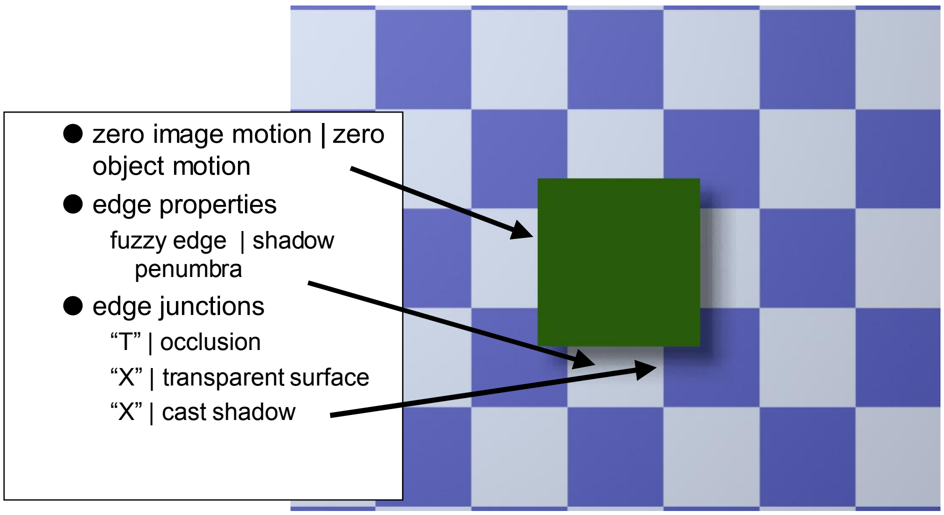


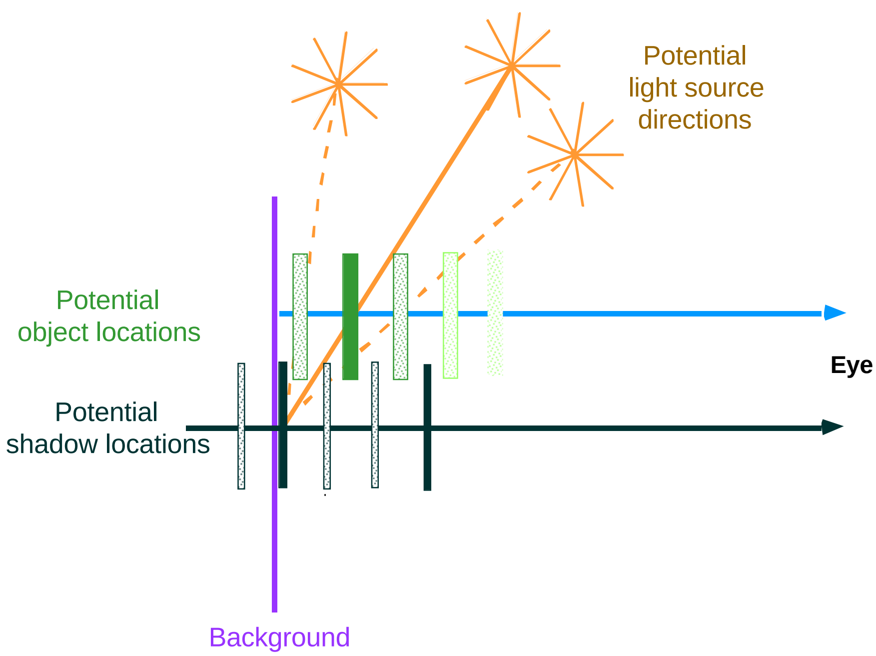
Examples of local image formation constraints

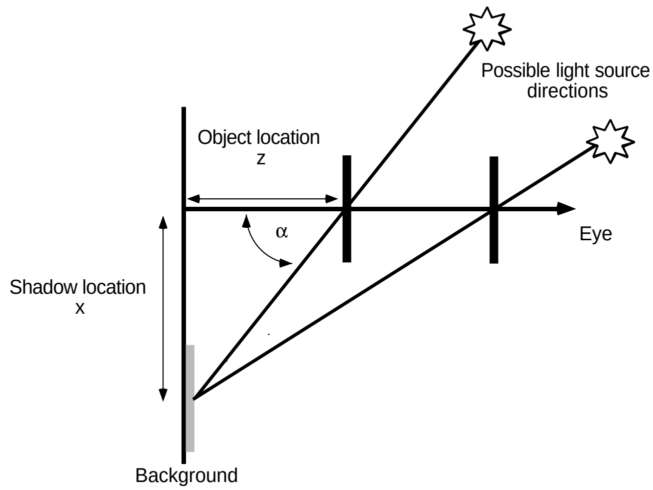
- zero image motion | zero object motion
- edge properties
  - fuzzy edge | shadow penumbra
  - fuzzy edge | surface crease
  - fuzzy edge | attached shadow
- edge junctions
  - "T" | occlusion
  - "T" | accidental alignment
  - "X" | transparent surface
  - "X" | cast shadow



Depth from cast shadows







Possible light source directions
Object location
z
α
Eye
Shadow location
x
Background

Genericity

Perception's model of the image should be robust over variations in generic variables

$$\Delta x = \Delta\alpha \frac{x^2 + z^2}{z}$$

$$z = x$$

See too: Shape from shading, Freeman, 1994;
Viewpoint as a generic variable: Lowe, 1986; 1987; Nakayama & Shimojo, 1992

Object recognition

Variations

- Viewpoint
  - Poggio & Edelman, 1990; Ullman,1996; Buelthoff, Edelman & Tarr, 1995; Liu, Knill & Kersten, 1995)
- Illumination
  - (cf. Belhumeur & Kriegman, 1996)
- Articulation
  - (Zhu & Yuille, 1996)
- Within class variations: categories
  - Bobick, 1987; Belhumeur, Hespanha, Kriegman, 1997)

Viewpoint

How do we recognize familiar objects from unfamiliar views?

3D transformation matching
(really smart)
View-combination
(clever)
View-approximation
(dumb?)

Liu, Knill & Kersten, 1995; Liu & Kersten, 1998ç

3D transformation matching
(really smart)

Explicit 3D knowledge
- Model of 3D object in memory
- Verify match by:
  - 3D rotations, translations of 3D model
  - Project to 2D
  - Check for match with 2D input

Problems
- Requires top-down processing
  i.e. transformations on memory representation, rather than image
- Predicts no preferred views

View-combination
(clever)

Implicit 3D knowledge
- Verify match by:
  - Constructing possible views by interpolating between stored 2D views
  - Check for match with 2D input
    - Basri & Ullman

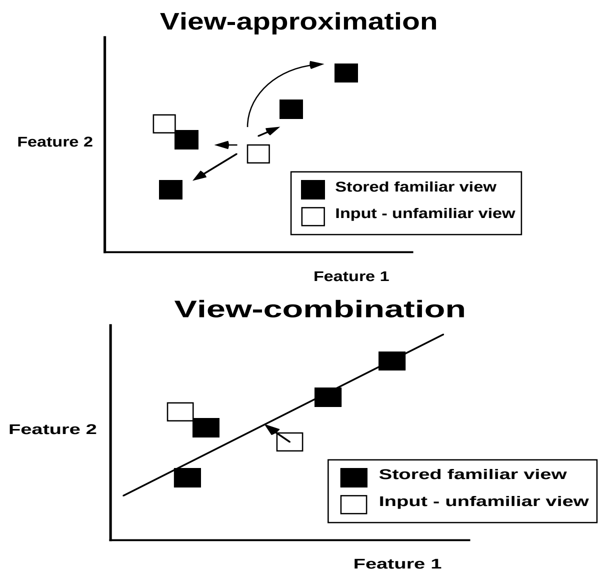
Problems
  - Hard to falsify psychophysically-- view-dependence depends on interpolation scheme

Advantages
  - Power of “really smart” 3D transformations but with simple transformations

View-approximation
(dumb?)

Little or no 3D knowledge
  - Familiar 2D views treated independently
- Verify match by:
  - Comparing incoming novel 2D view with multiple familiar 2D views stored in memory

Advantages
- Simple computation
- Psychophysics with novel objects
  - Rock & DiVita, Bülthoff & Edelman, Tarr et al.
- View-dependence in IT cells
  - Logothetis et al.



View-approximation

Range of possible models

- 2D template nearest neighbor match
- 2D transformations + nearest neighbor match
- 2D template + optimal match
- 2D transformations + optimal match

2D transformations + optimal matching

2D rigid ideal observer

allows for:

- translation
- rigid rotation
- correspondence ambiguity

2D affine ideal observer

allows for:

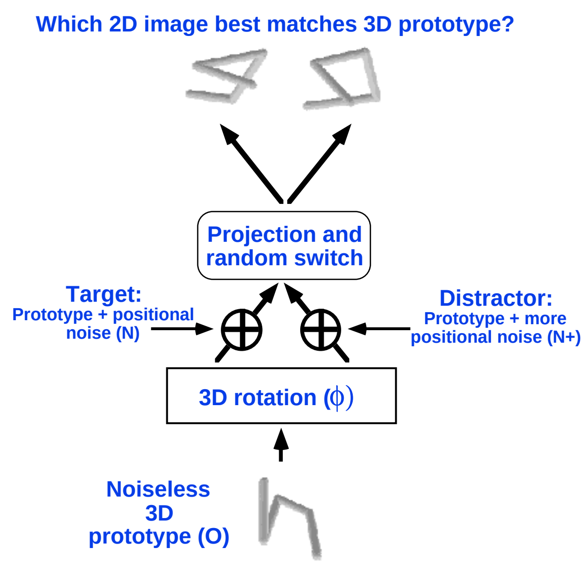
- translation
- scale
- rotation
- stretch
- correspondence ambiguity

Ideal observer analysis

Statistical model of information available in a well-defined psychophysical task

Specifies inherent limit on task performance

Liu, Knill & Kersten, 1995



Optimal Matching

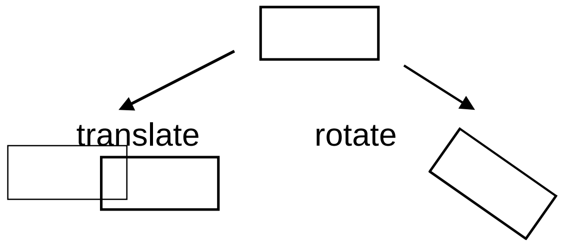
2D/2D sub-ideal -- 2D rigid transformations to match stored templates $\mathbf{T}_i$

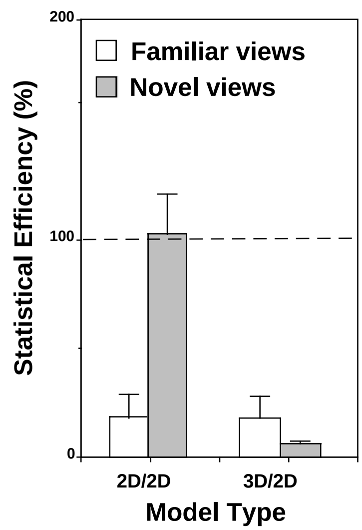
$$p_t(\mathbf{I}) = \sum_{i=1}^{11} \int_0^{2\pi} [p(\mathbf{I} - R_\phi(\mathbf{T}_i)p(R_\phi(\mathbf{T}_i))d\phi$$

3D/2D ideal -- 3D rigid transformations of object **O**

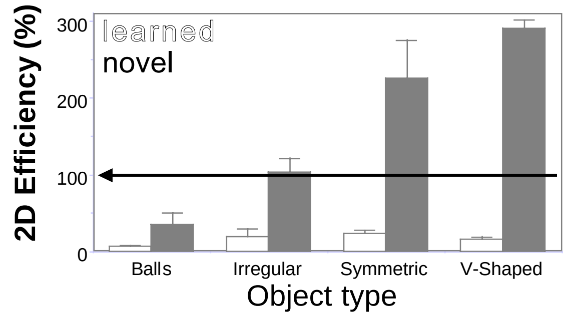
$$p_t(\mathbf{I}_k) = \int p(\mathbf{N}_p = \mathbf{I} - F_\Phi(\mathbf{O}))p(\Phi)d\Phi$$

2D rigid ideal

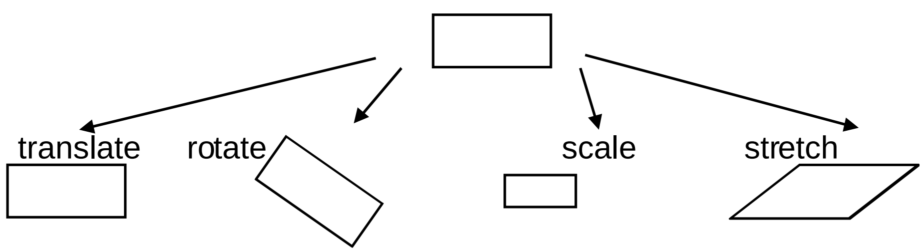




Humans vs. 2D rigid ideal: Effect of object regularities



2D affine ideal

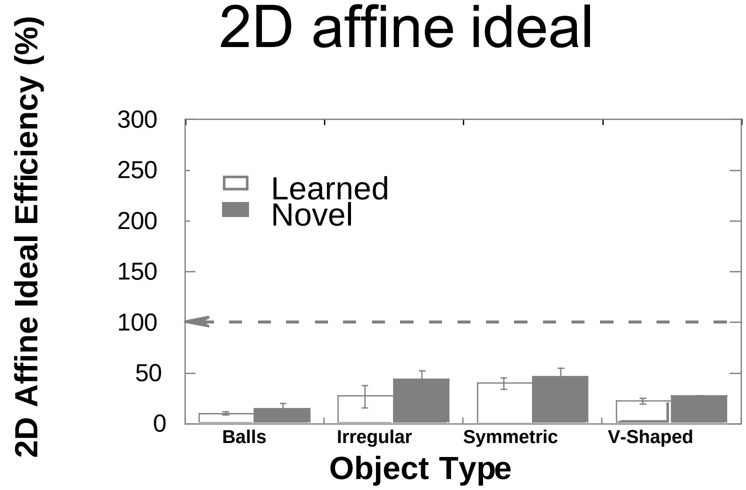


$$\begin{pmatrix} x_s^1 & x_s^2 \cdots & x_s^n \\ y_s^1 & y_s^2 \cdots & y_s^n \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_t^1 & x_t^2 \cdots & x_t^n \\ y_t^1 & y_t^2 \cdots & y_t^n \end{pmatrix} + \begin{pmatrix} t_x & t_x \cdots & t_x \\ t_y & t_y \cdots & t_y \end{pmatrix}$$

$$.\ p(\mathbf{S} \mid \mathbf{T}) = \int da\ db\ db\ dc\ dd\ dt_x\ dt_y\ t_x\ t_y)\ p(\mathbf{S} \mid a\ b\ c\ d\ t_x\ t_y, \mathbf{T})\ p(a\ b\ c\ d\ t_x\ t_y)$$

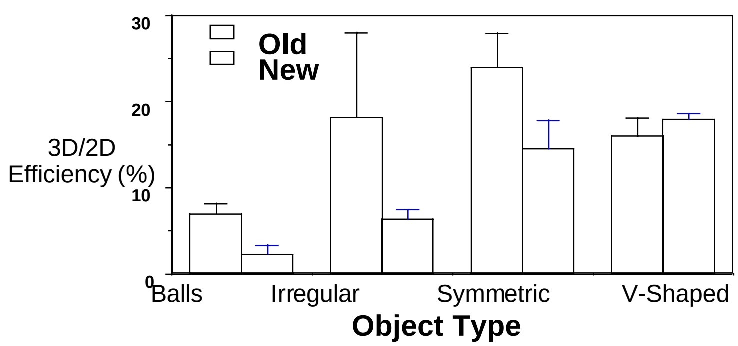
Liu & Kersten, 1998

Humans vs. 2D affine ideal



Liu & Kersten, 1998

Humans vs. "smart" ideal: Effect of object regularity



Peak efficiency relative to "really smart" ideal is 20% for familiar views, but less for new ones.

Results

Relative to 2D ideal with rigid rotations

Human efficiency > 100%

Relative to 2D affine

Efficiency for novel views is bigger than for familiar views

Efficiency for novel views increases with object class regularity

Conclusions

3D transformation ideal

– View-dependency for subordinate-level type task

2D rigid & affine ideals

– view-approximation models unlikely to account for human performance

More 3D knowledge either in the memory representation or matching process is required to account for human performance

Cutting the Gordian Knot: Initial fast access given natural images

Attention allocation

- 20 questions, minimum entropy selection
  - Geman & Jednyak (1993)
  - Mr. Chips ideal observer model for reading (Legge,, Klitz, & Tjan, 1997)

Support vector machines

- Face recognition/detection (Osuna, Freund & Girosi, 1997)
- Object recognition (Schölkopf, B., 1997)

Principles of Perceptual Inference: Key points I (Yuille, Coughlan & Kersten)

- Vision is decoding input image signals in order to extract information and determine appropriate actions
- Natural images consist of complex patterns; but there are regularities and, in particular, a limited number of transformations which constantly appear
- In Bayesian models the objects of interest, both in the image and in the scene, are represented by random variables. These probability distributions should represent the important properties of the domain and should be learnt or estimated if possible. Stochastic sampling can be used to judge the realism of the distributions

Key points II

- Visual inference about the world would be impossible if it were not for regularities occurring in scenes and images. The Bayesian approach gives a way of encoding these assumptions probabilistically. This can be interpreted in terms of obtaining the simplest description of the input signal and relates to the idea of vision as information processing
- The Bayesian approach separates the probability models from the algorithms required to make inferences from these models. This makes it possible to define ideal observers and put fundamental bounds on the ability to perform visual tasks *independently* of the specific algorithms used.
- Various forms of inference can be performed on these probability distributions. The basic elements of inference are marginalization and conditioning.

Key points III

Probability distributions on many random variables can be represented by graph structures with direct influences between variables represented by links. The more complex the vision problem, in the sense of the greater direct influence between random variables, the more complicated the graph structure

The purpose of vision is to enable an agent to interact with the world. The decisions and actions taken by the agent, such as detecting the presence of certain objects or moving to take a closer look, must depend on the importance of these objects to the agent. This can be formalized using concepts from decision theory and control theory.

Computer vision modelers assume that the uncertainty lies in the scene and pay less attention to the image capturing process. By contrast, biological vision modelers have paid a lot of attention to modeling the uncertainty in the image measurements -- and less on the scene.

Yuille, A.L., Coughlan, J. M., and Kersten, D. Computational Vision: Principles of Perceptual Inference.
http://vision.psych.umn.edu/www/kersten-lab/papers/yuicouker98.pdf

Limited number of copies available here