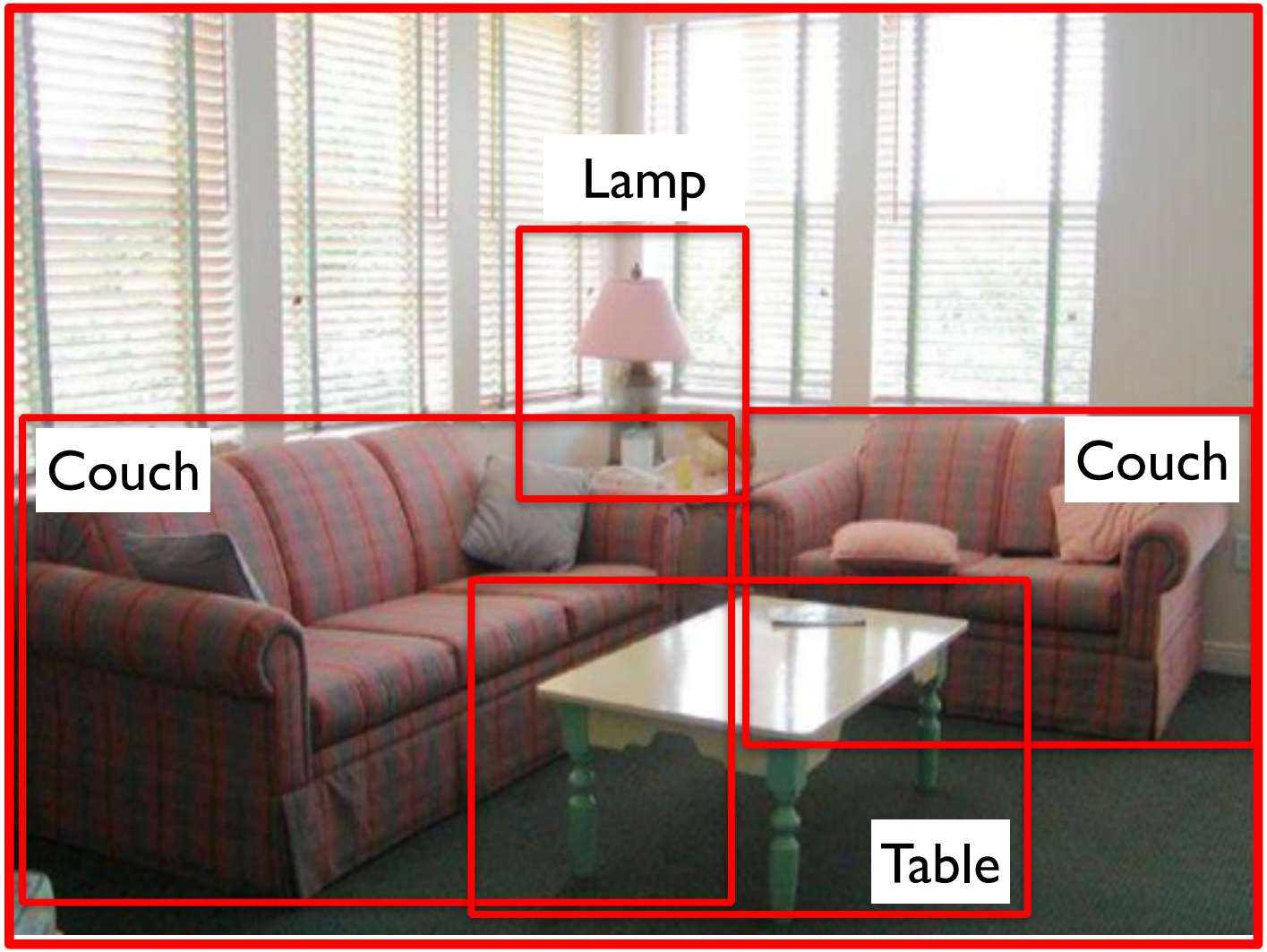
Lecture 2: Theories and History

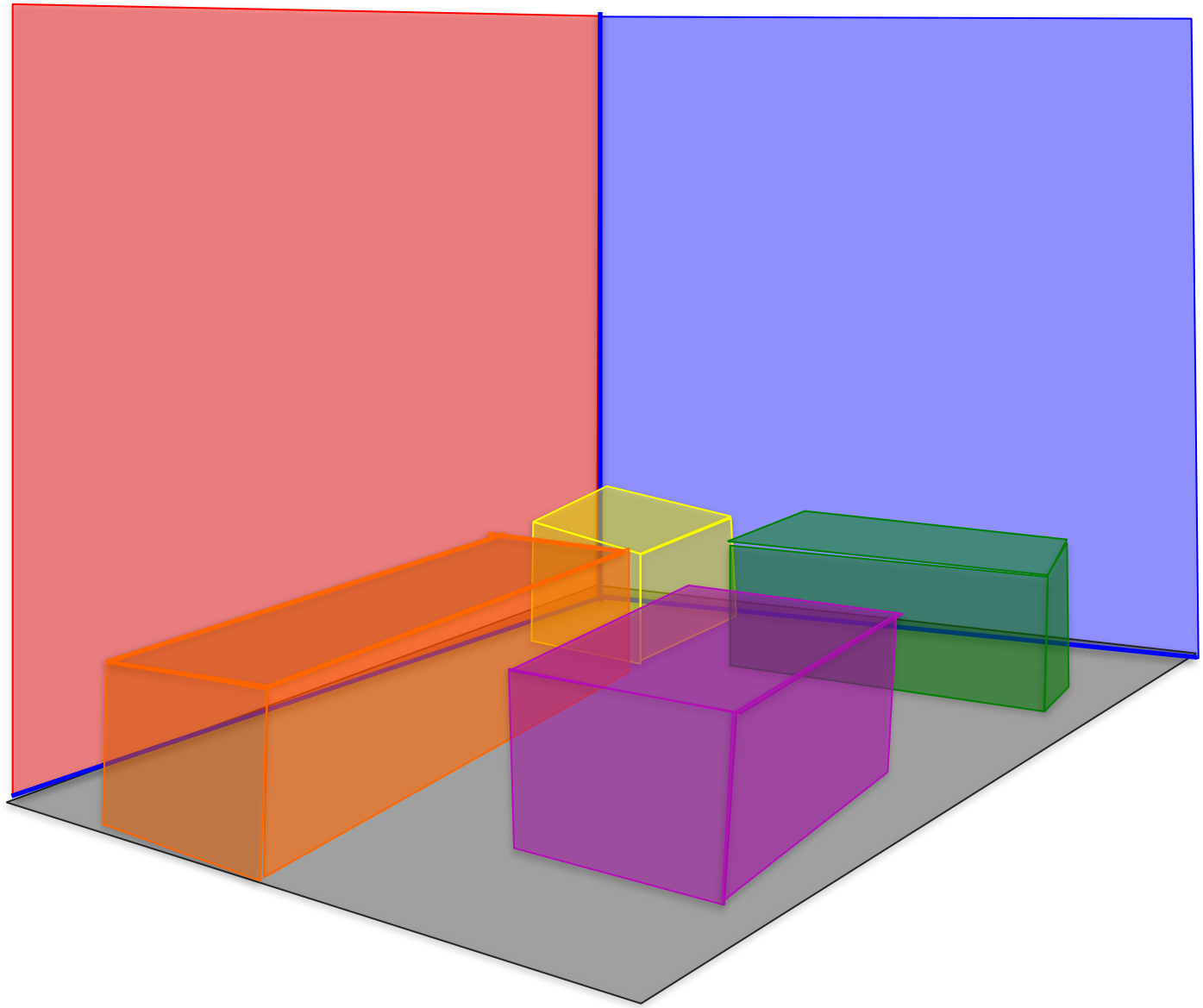
fundamental questions

What does it mean to understand?

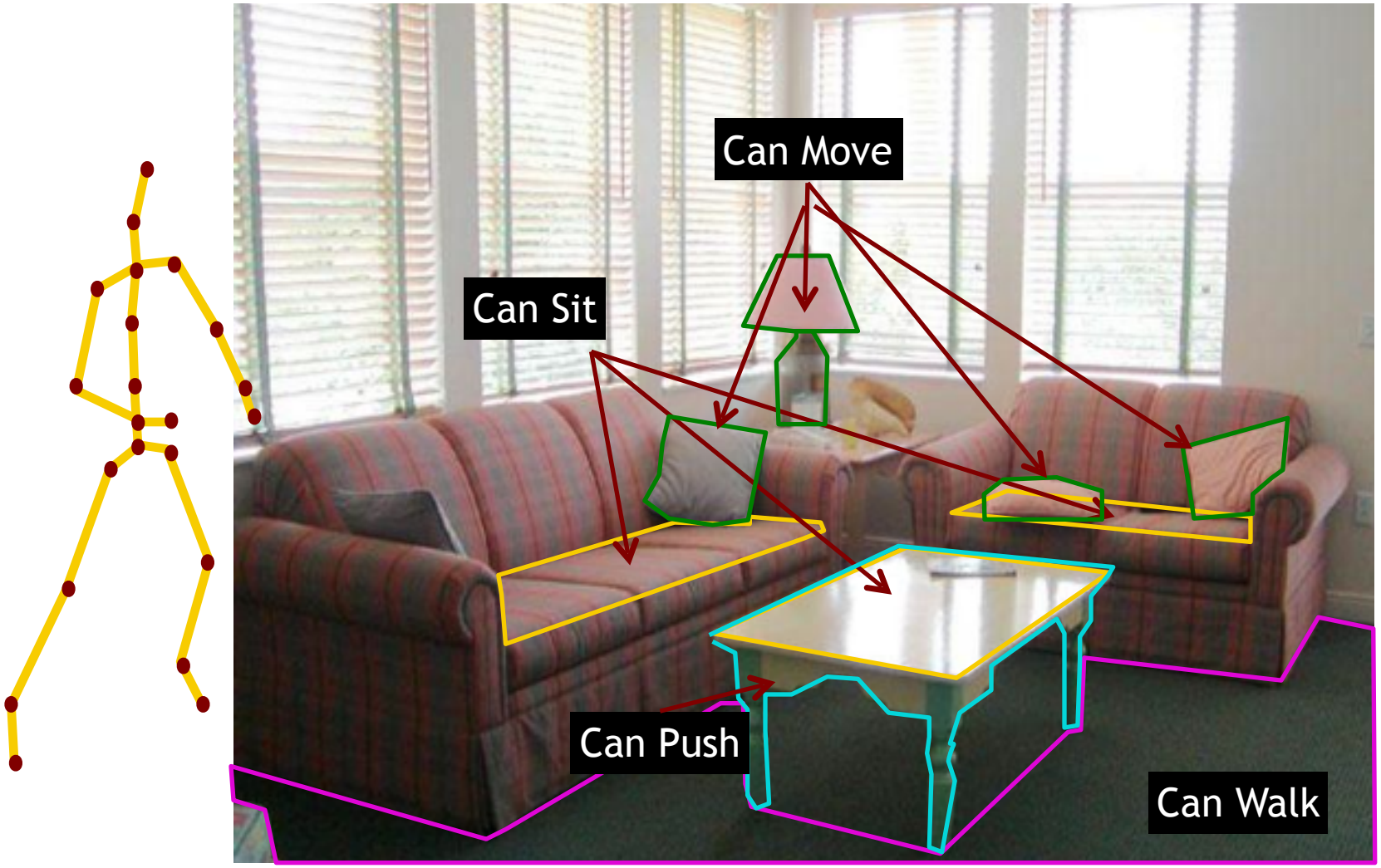
semantics



3D understanding



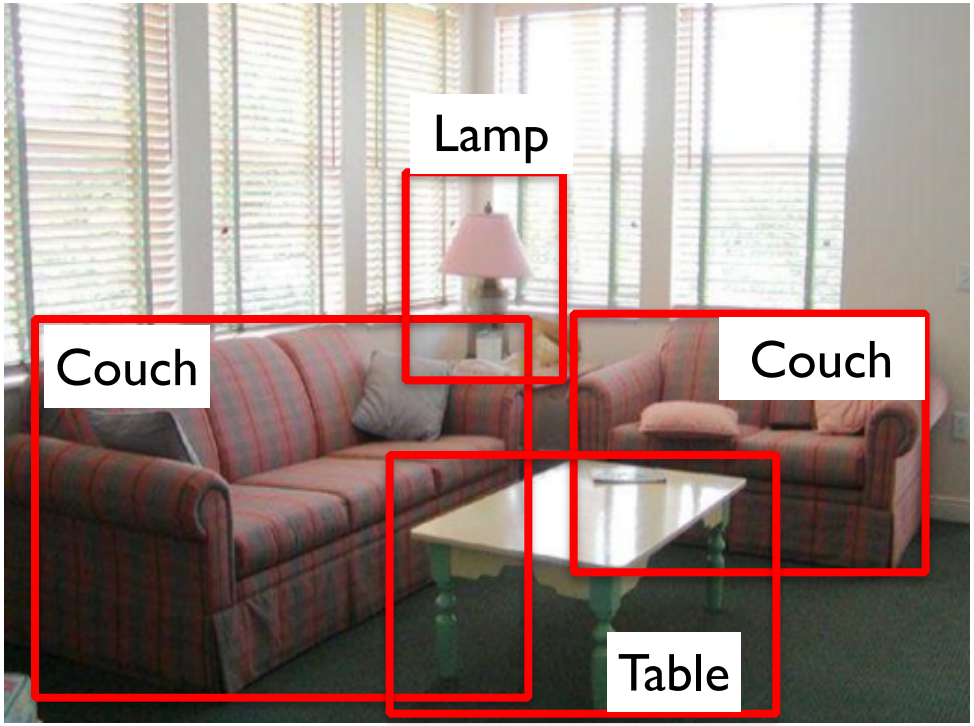
functions – what can we do?

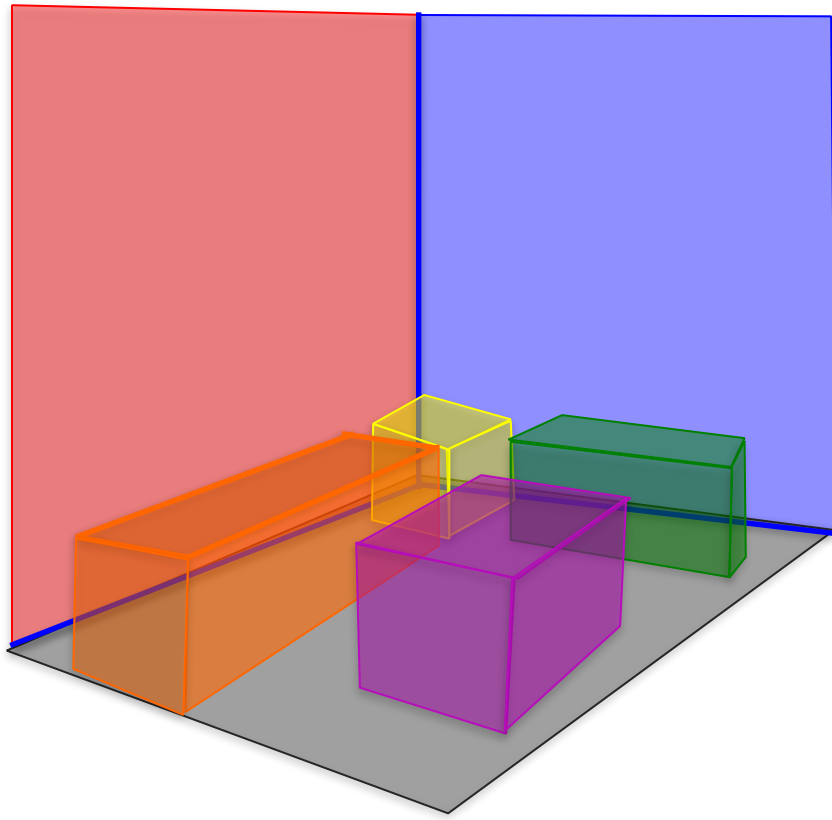


fundamental questions

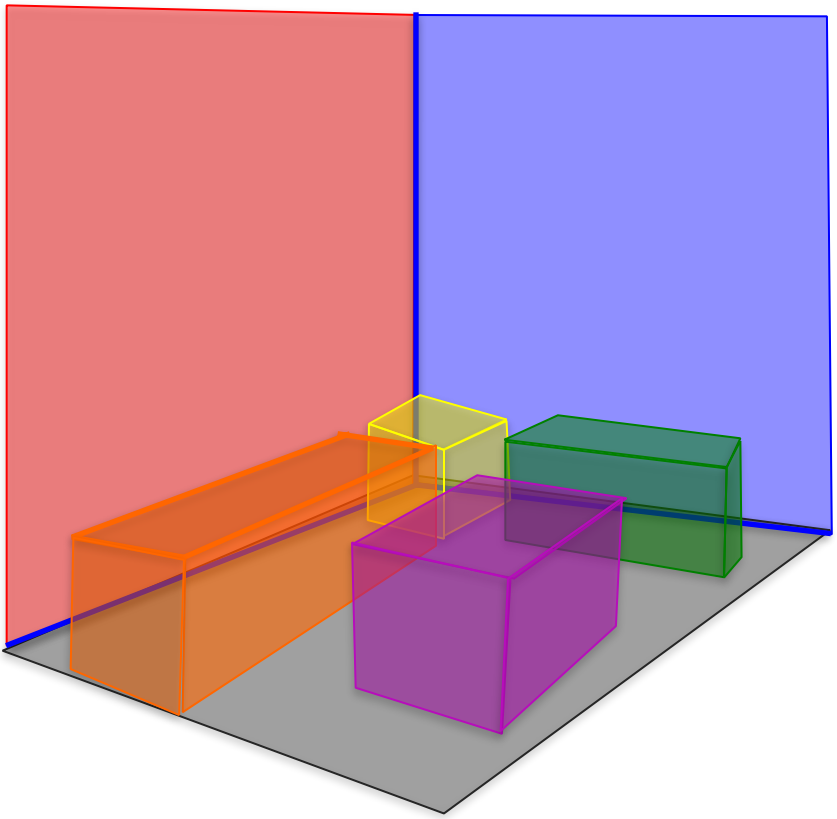
how are these tasks related?

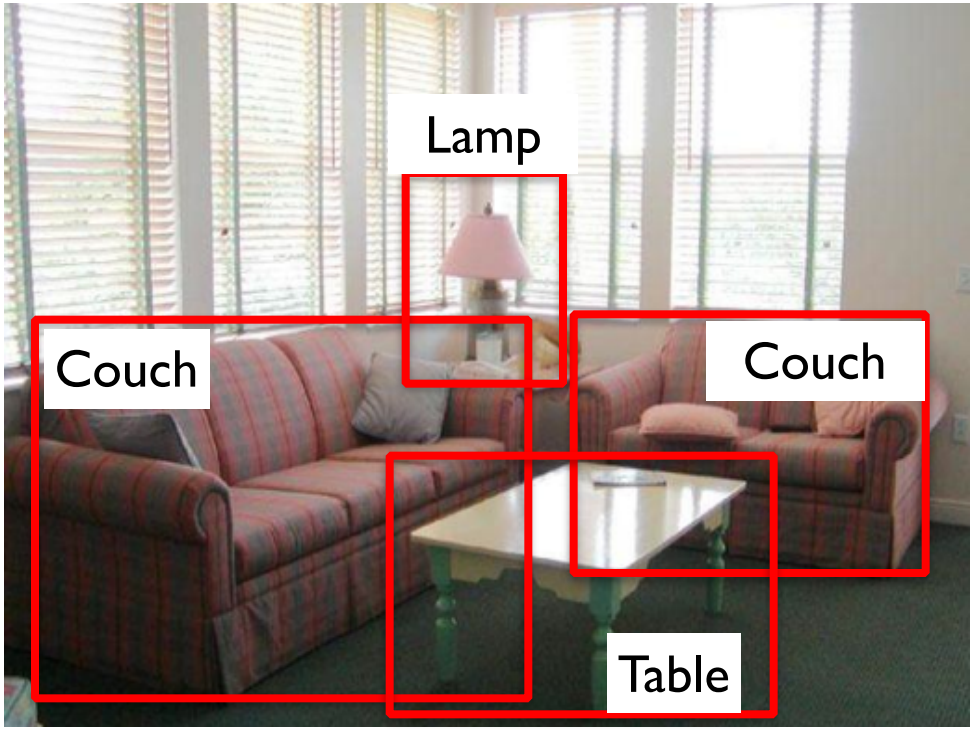
semantics -> 3D





3D -> semantics

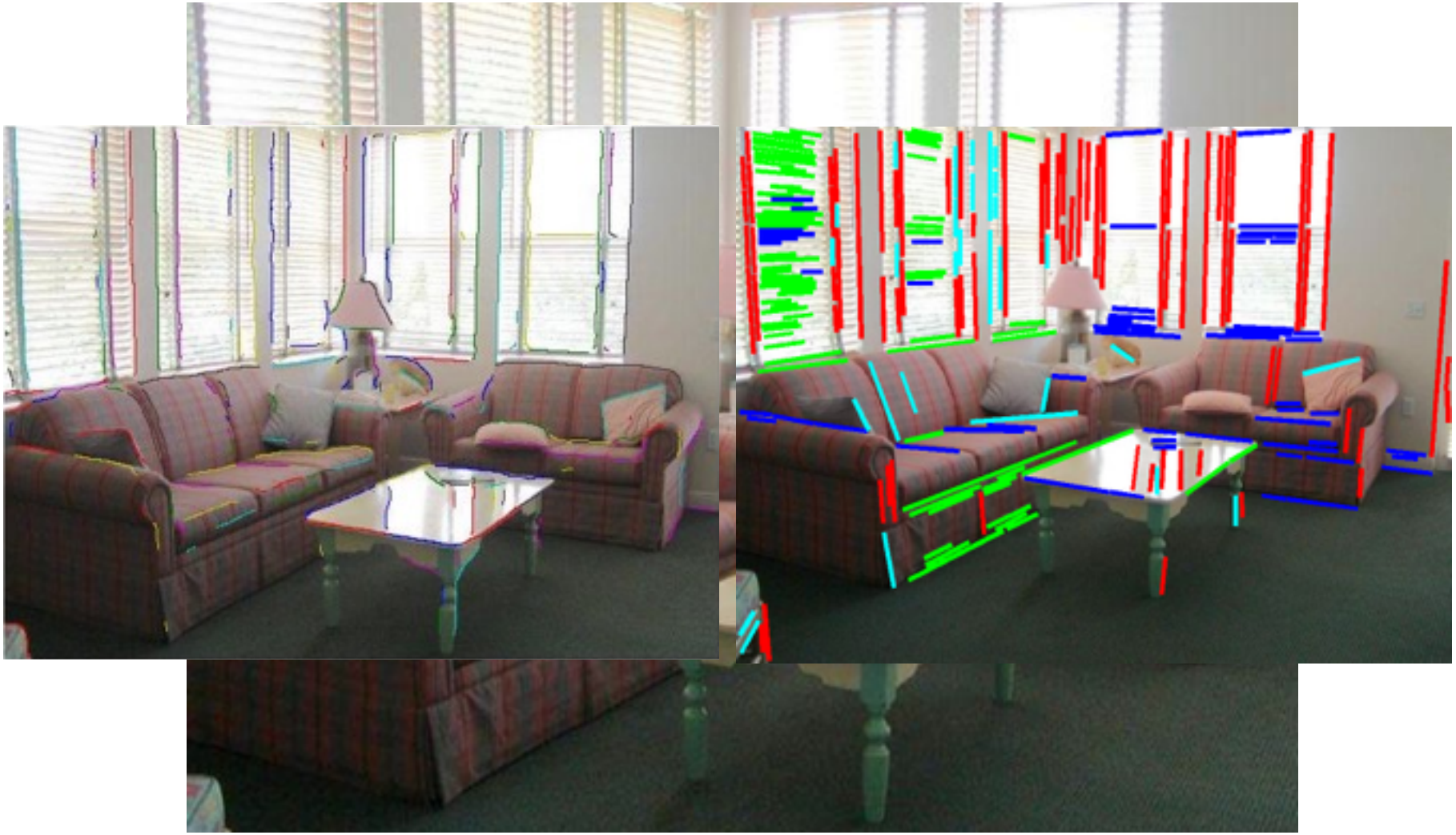




fundamental questions

What are the processes involved?

.ines, contours, segments



object detection

matching scenes

how do we answer all these questions?

how do we answer all these questions?

Understanding Human Vision

Most Slides from Palmer

Goal: Understand Perception

Approach 1: Discover all the facts related to perception !

Is that even possible?

Understanding a process as complex as perception requires more than listing facts

Theories

What is a Theory?

Theory

An integrated set of related hypotheses about underlying mechanisms or principles that organize, explain, and predict facts about the phenomena of a given domain.

What is a Theory?

Three important functions of a theory:

- Integrate “old” facts
- Predict “new” facts
- Lead to “understanding”

Three criteria for evaluating a theory:

- Logical consistency
- Empirical adequacy
- Parsimony (“Occam’s razor”)

Theories of Vision

Categorization of Theories of Vision

- "*Why do things looks as they do?"*
  -- Kurt Koffka (1935)

- Nativism vs. Empiricism
  - "Because we were born *(evolved)* to see them that way" vs.
  - "Because we have learned to see them that way"
- Atomism vs. Holism
  - "because of the way each pixel appears" vs.
  - "because of the way the entire scene appears"
- Organism vs. Environment
  - "Because we are the way we are" vs.
  - "Because the world is the way it is"

4th Axis

- How do you derive the theory?
  - Introspection (Conscious Inference)
  - Behavior (Measurement of Human Performance)

  - -Koffka: why do the things look the way they do?
  - -What does vision enable us to do?

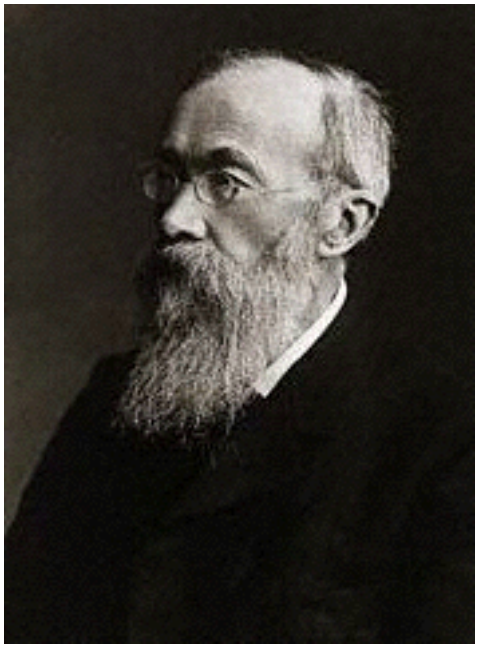
Structuralism

THEORY	NATIVISM vs. EMPIRICISM	ATOMISM vs. HOLISM	ORGANISM vs. ENVIRONMENT	PRINCIPAL ANALOGY	METHOD
Structuralism					
Gestaltism					
Ecological Optics					
Constructivism					

Structuralism

Derived from philosophy of British Empiricists (e.g., Locke, Berkeley, Hume, and Mills).

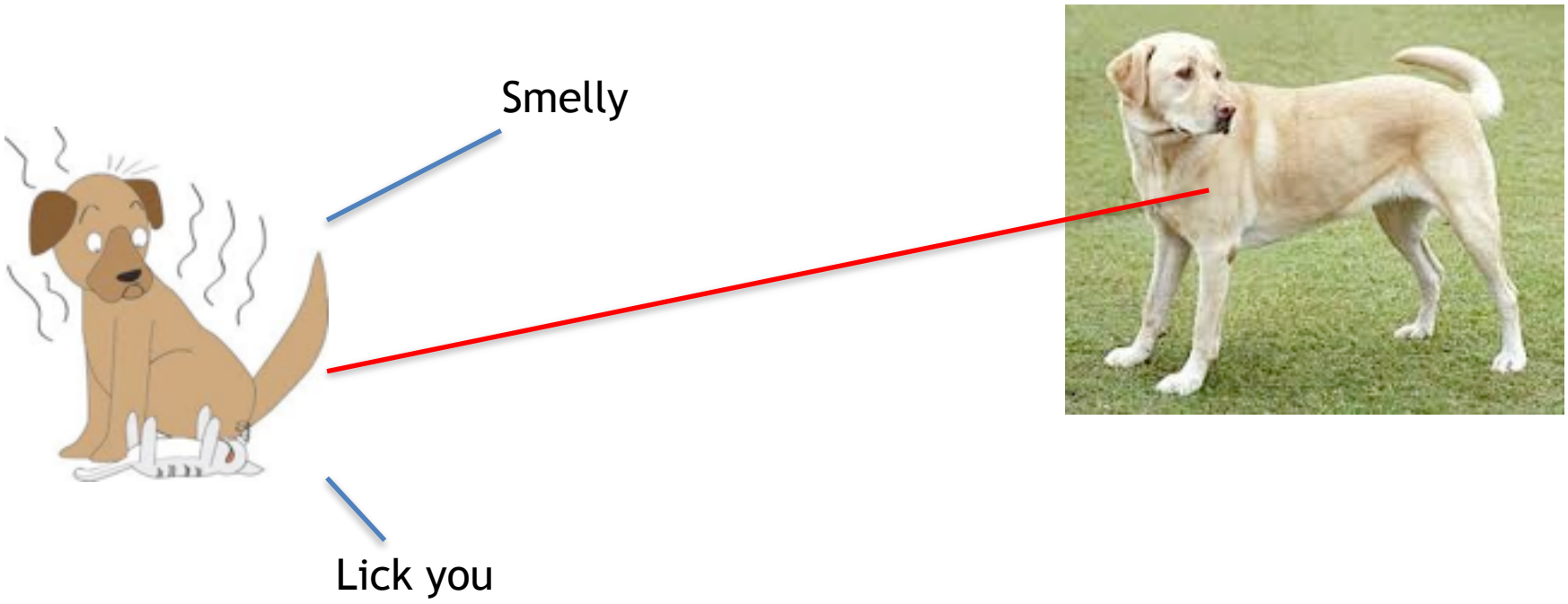
Proposed by Wilhelm Wundt, the father of modern Psychology.



Structuralism

Structuralism:

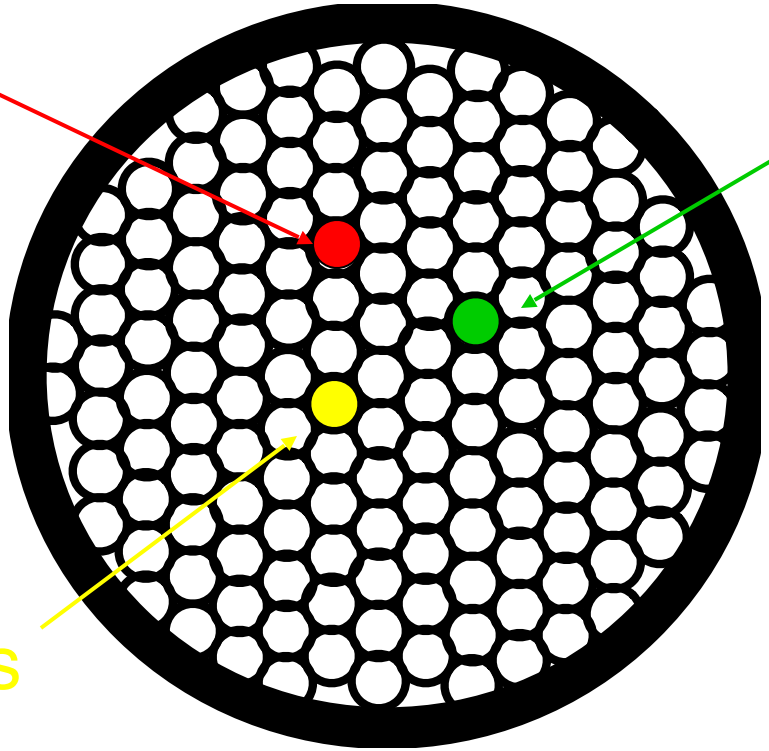
Perception results from the association of basic sensory atoms in memory via repeated, prior joint occurrences.



Structuralism

Sensory Atoms

Redness
at (x1,y1)



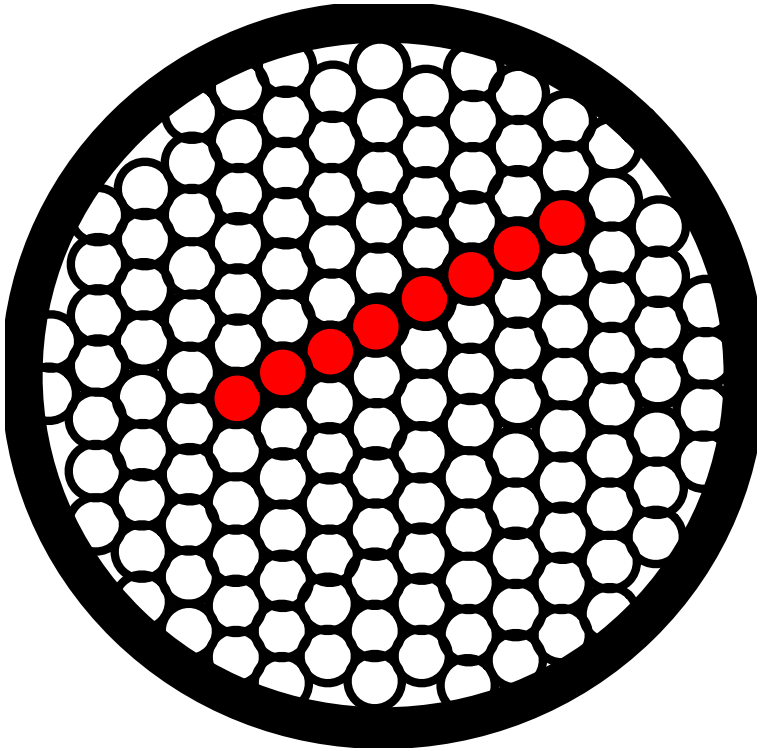
Greenness
at (x3,y3)

Yellowness
at (x2,y2)

Retinal mosaic

Structuralism

Perceptual Complexes

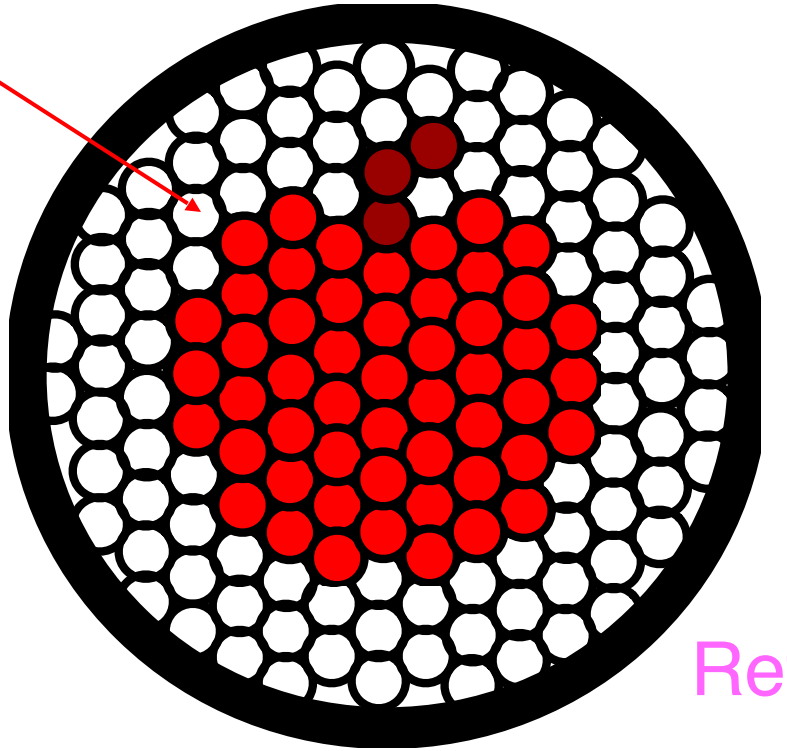


Retinal mosaic

Structuralism

Perceptual Complexes

Red apple
at (x0,y0)

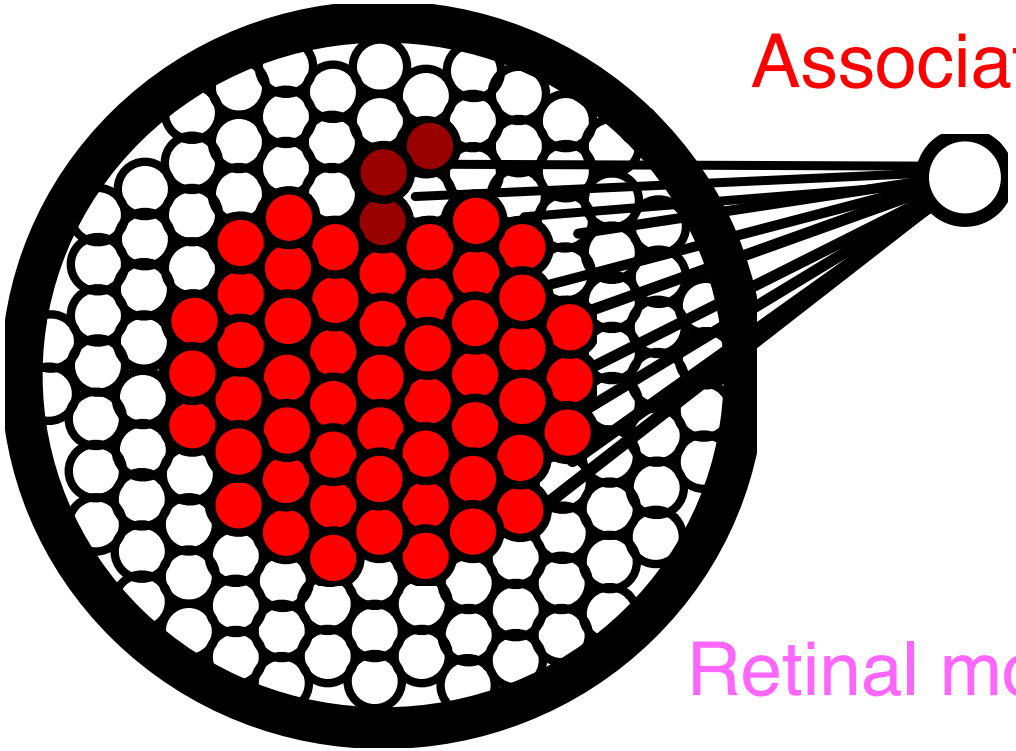


Retinal mosaic

Structuralism

Perceptual Complexes

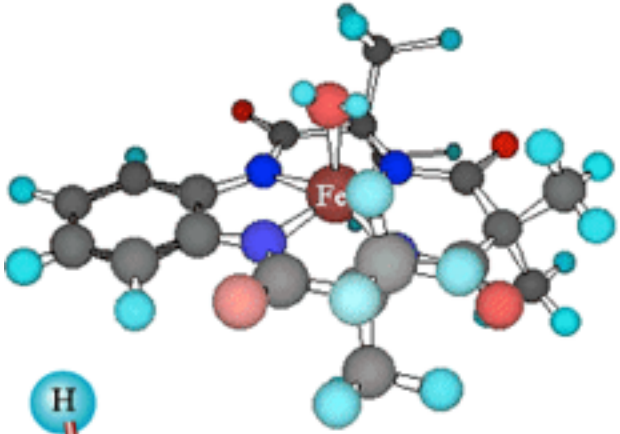
Red apple
at (x0,y0)

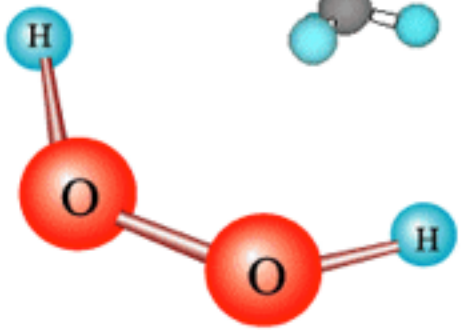


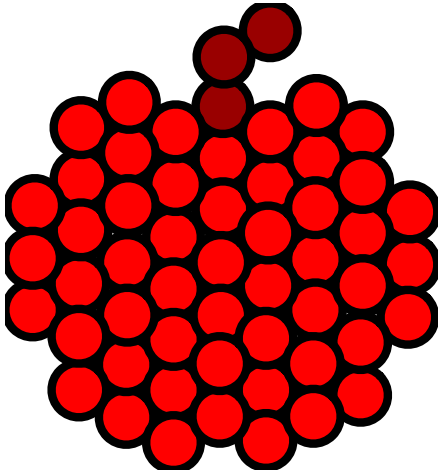
Structuralism

Chemical Analogy

Perceptions are made of basic sensory experiences just as molecules are made of basic atoms.







Structuralism

Training Experts to discover the elementary units of perception

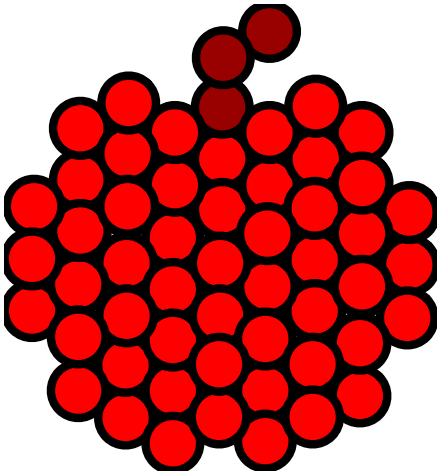
Nature of training led to different conclusions – One of the reasons for the downfall

Structuralism

THEORY	NATIVISM vs. EMPIRICISM	ATOMISM vs. HOLISM	ORGANISM vs. ENVIRONMENT	PRINCIPAL ANALOGY	METHOD
Structuralism	Empiricism	Atomism	Organism	Chemistry	Trained Introspection
Gestaltism					
Ecological Optics					
Constructivism					

Gestaltism

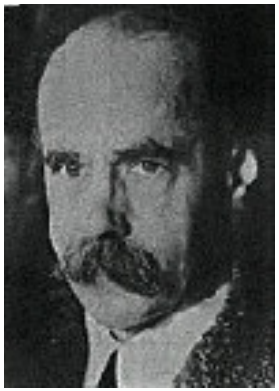
Whole is different from sum of its parts



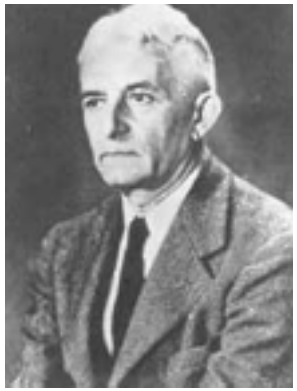
Gestaltism

Gestaltism:

Perception results from the interaction between the intrinsic structure of the stimulus and the intrinsic structure of the brain.



Max Wertheimer



Wolfgang Köhler

Kurt Koffka

Gestaltism

The Gestalt movement in perceptual theory was primarily a reaction against Structuralism:

Rejected atomism
Rejected empiricism
Rejected associationism

Successful in arguing against Structuralism, but less successful in promoting its own theoretical agenda.

© Stephen E. Palmer, 2002

Gestaltism

Principles of Gestalt Theory

Holism:

The whole is different from the sum of its parts.

Prägnanz:

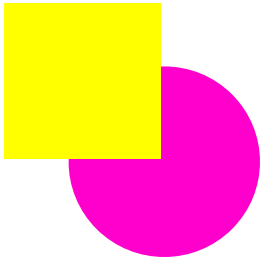
The percept will be as "good" as the prevailing conditions allow, I.e. simplest explanation

Nativism:

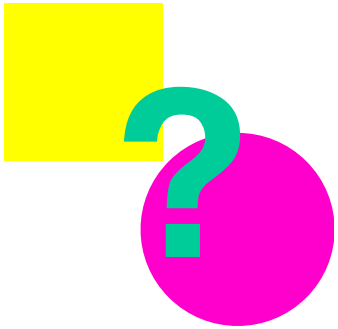
Not a total rejection of learning, but rejection of its primacy.

Gestaltism

Prägnanz: the percept will be as "good" as the prevailing conditions allow



What is this?



square & circle?

square & pacman?

squigit & pacman?

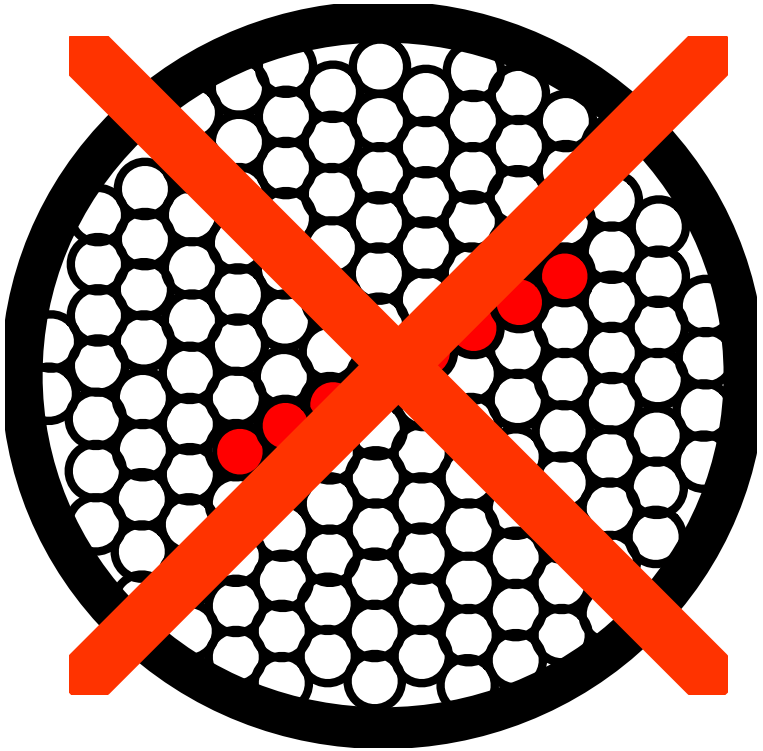
etc….

Gestaltism

Holism: The whole is <u>different</u> from the sum of its parts.

<u>Emergent properties:</u>
Features of a configuration that are not features of its components, e.g.:

- length
- orientation
- curvature
- closure
- connectedness



Emergent Properties

A

B

C

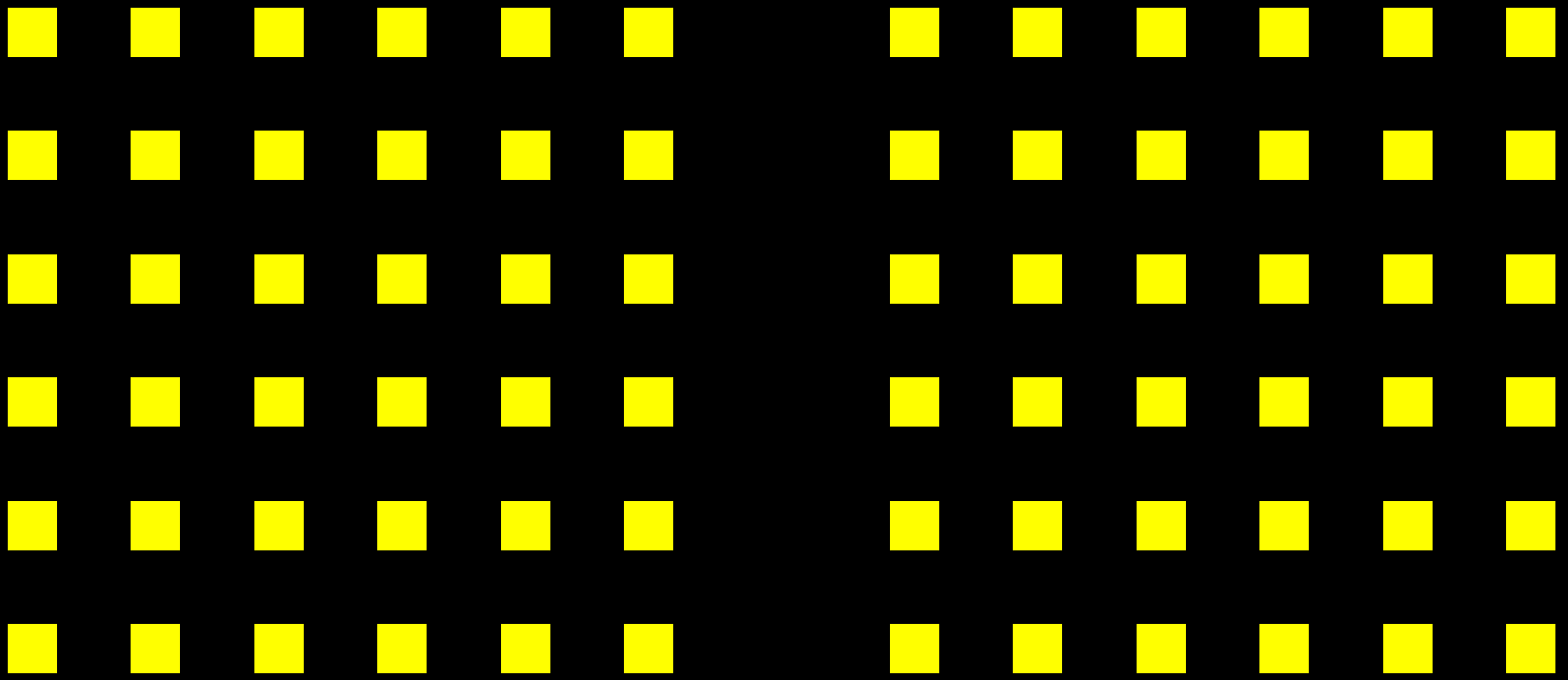
The Problem of Organization

The problem of grouping:

"I stand at the window and see a house, trees, sky." Theoretically, I might say that there were 327 brightnesses and nuances of color. Do I have "327"? No. I have sky, house, and trees. (Wertheimer, 1923)

Perceptual Grouping

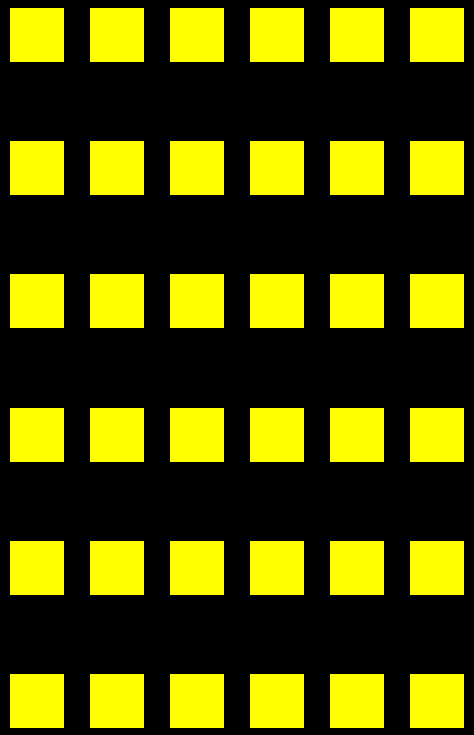
Wertheimer's "laws" of grouping

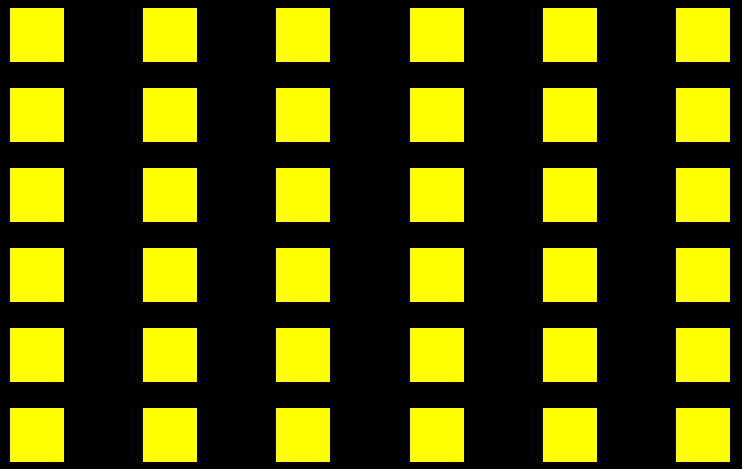


© Stephen E. Palmer, 2002

Perceptual Grouping

Proximity

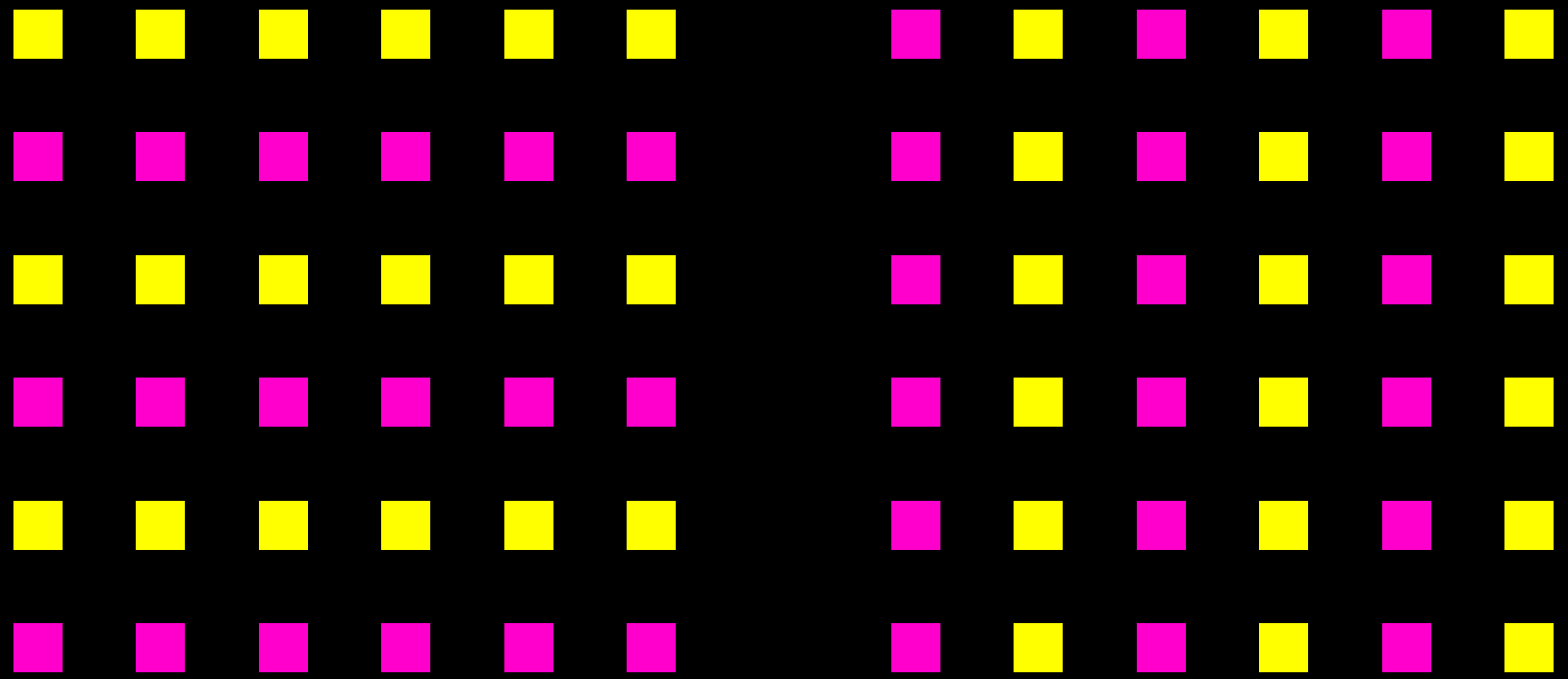




© Stephen E. Palmer, 2002

Perceptual Grouping

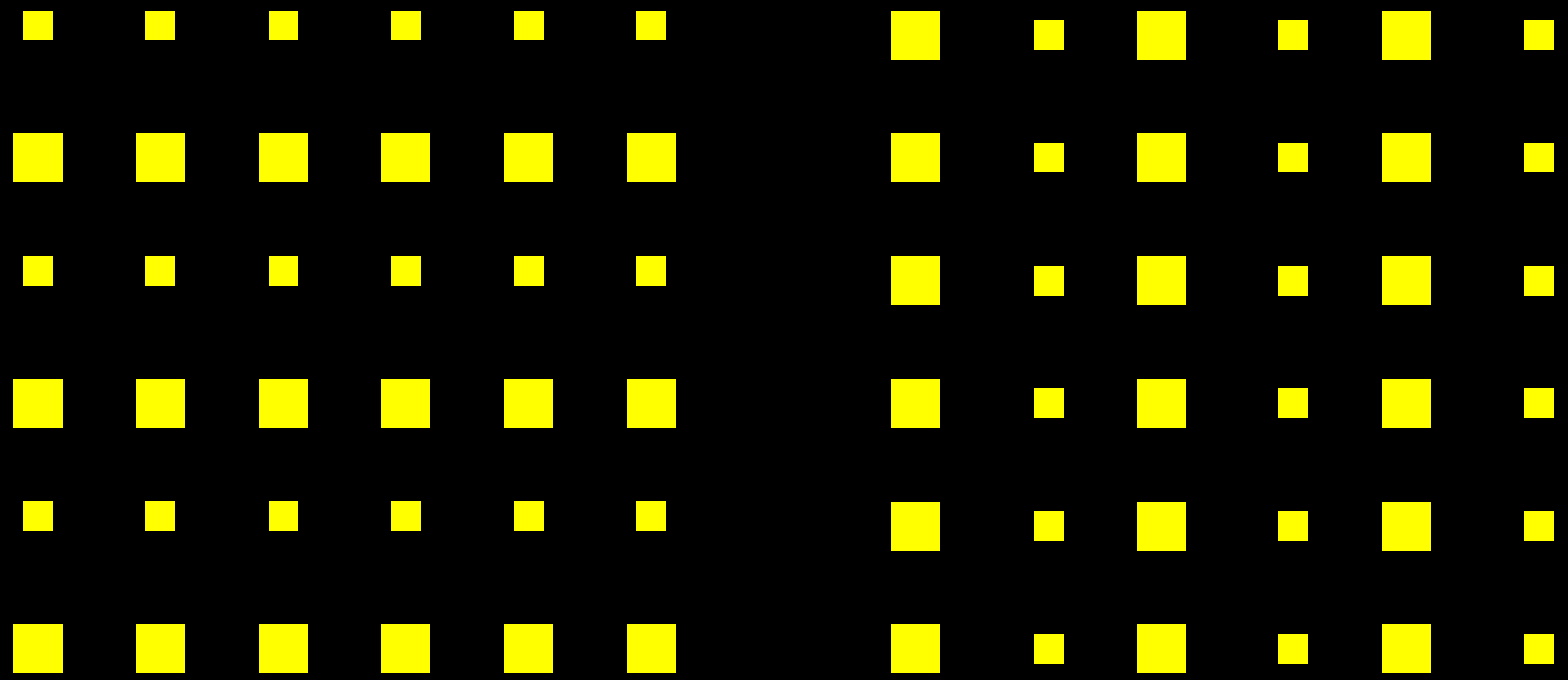
Color Similarity



© Stephen E. Palmer, 2002

Perceptual Grouping

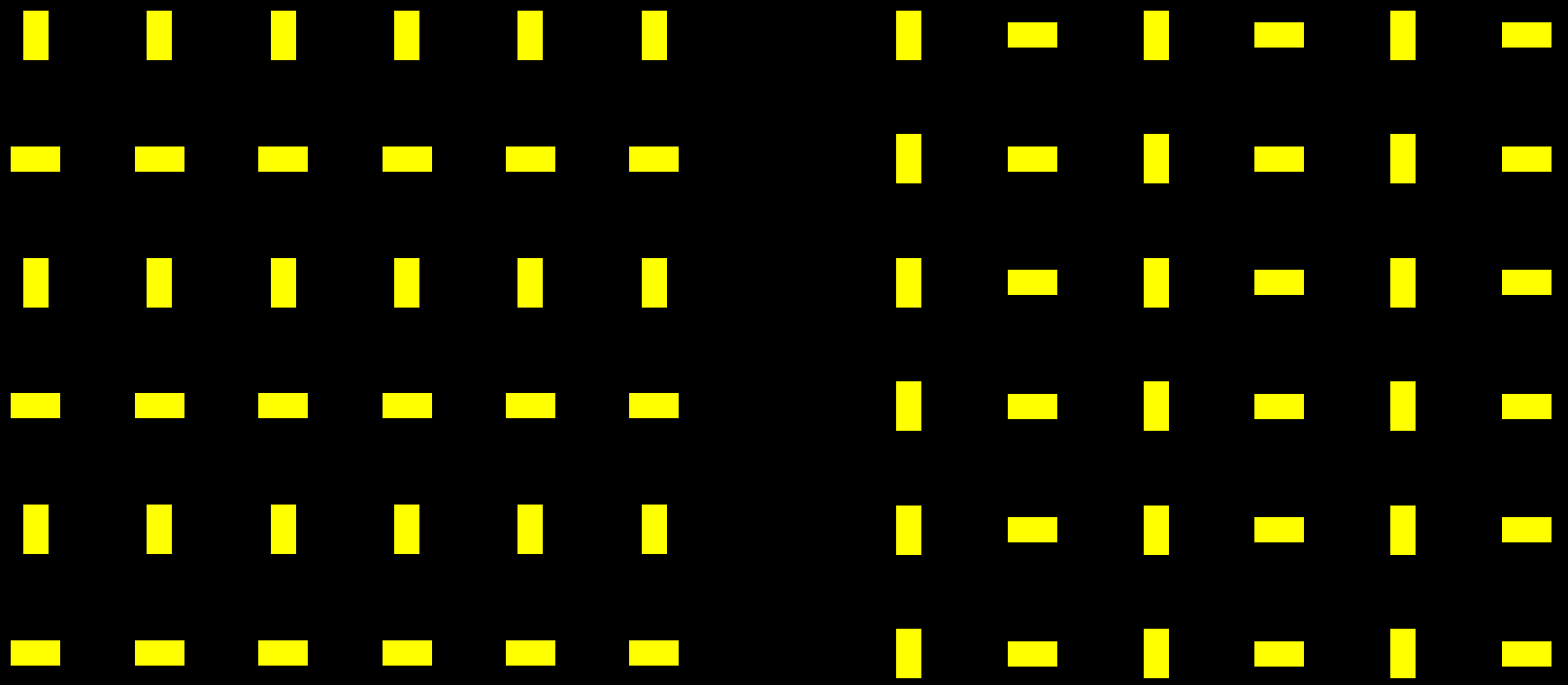
Size Similarity



© Stephen E. Palmer, 2002

Perceptual Grouping

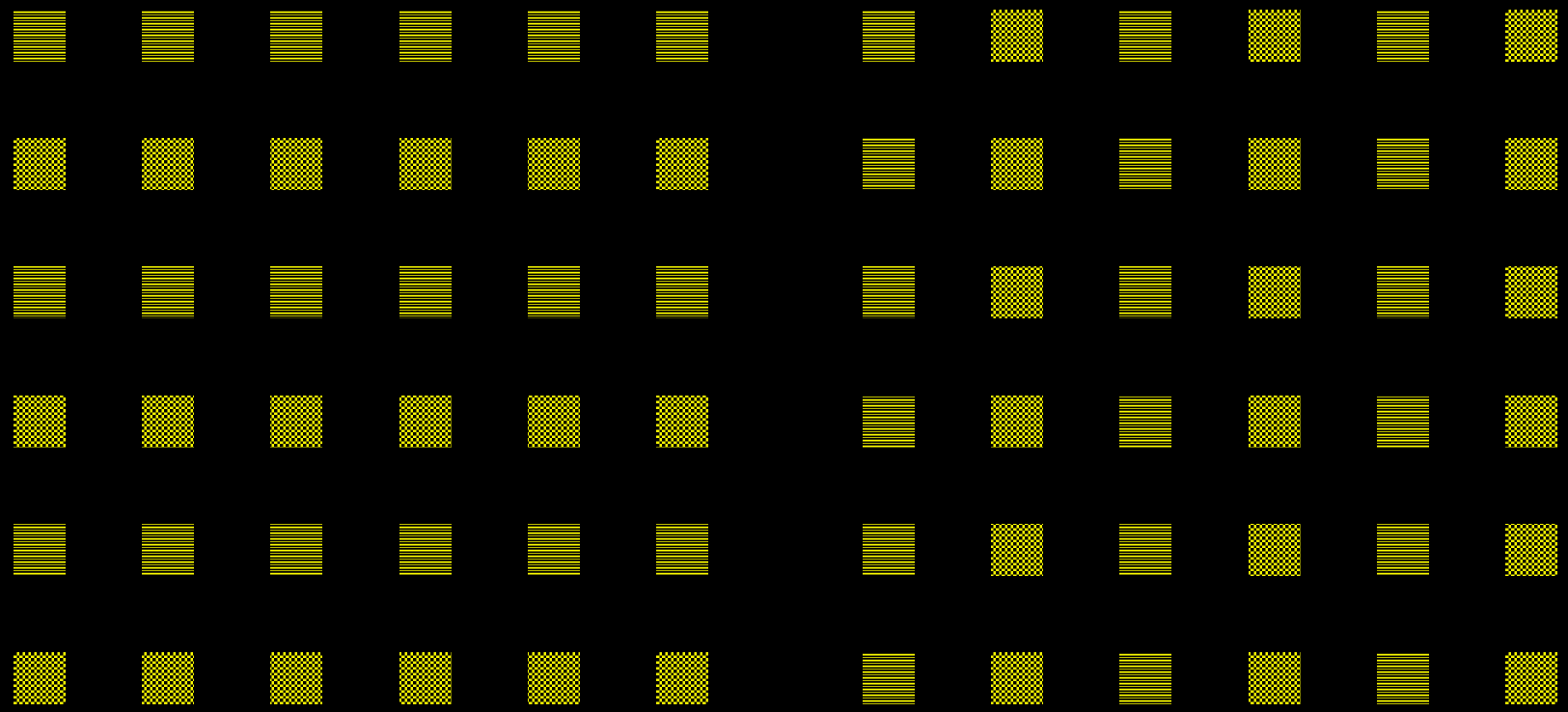
Orientation Similarity



© Stephen E. Palmer, 2002

Perceptual Grouping

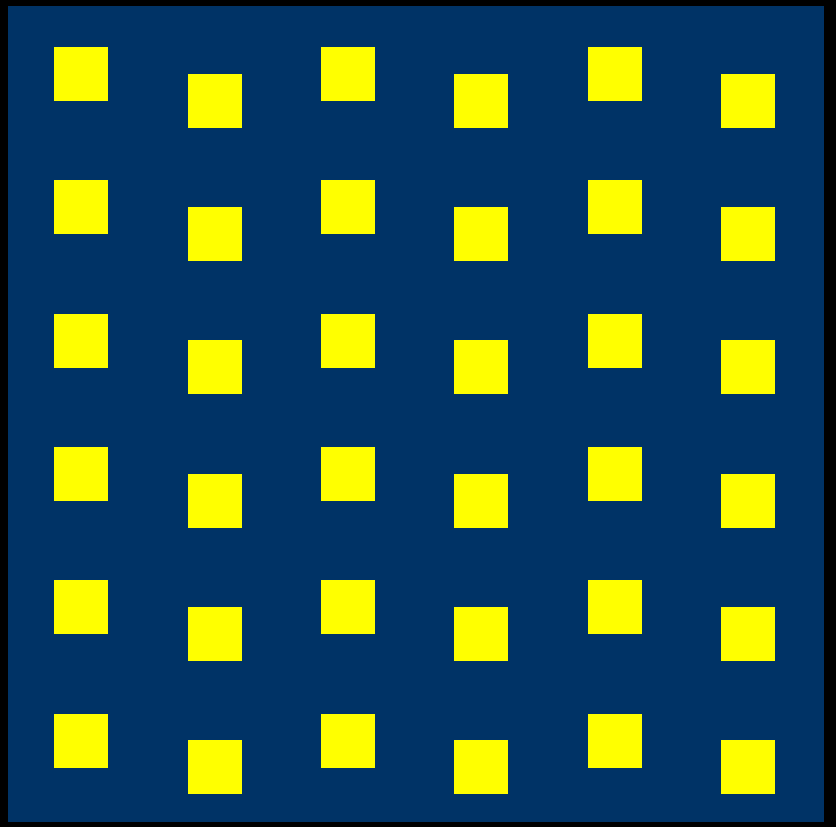
Similarity of texture



© Stephen E. Palmer, 2002

Perceptual Grouping

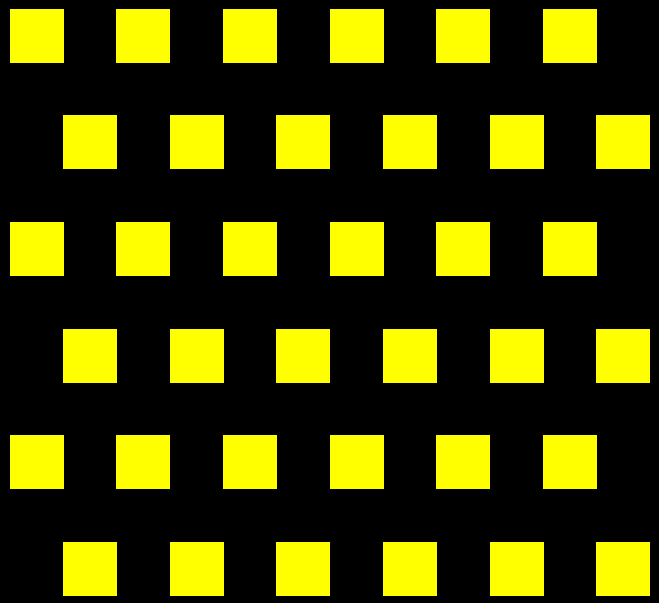
Common Fate

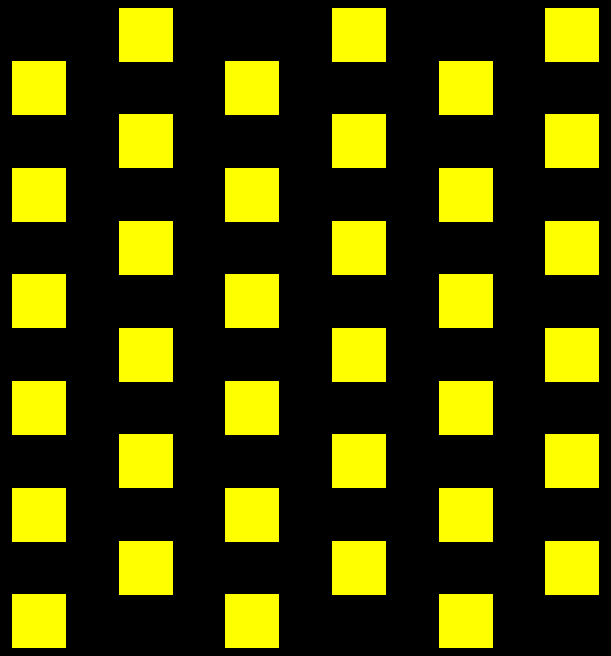


© Stephen E. Palmer, 2002

Perceptual Grouping

Good Continuation

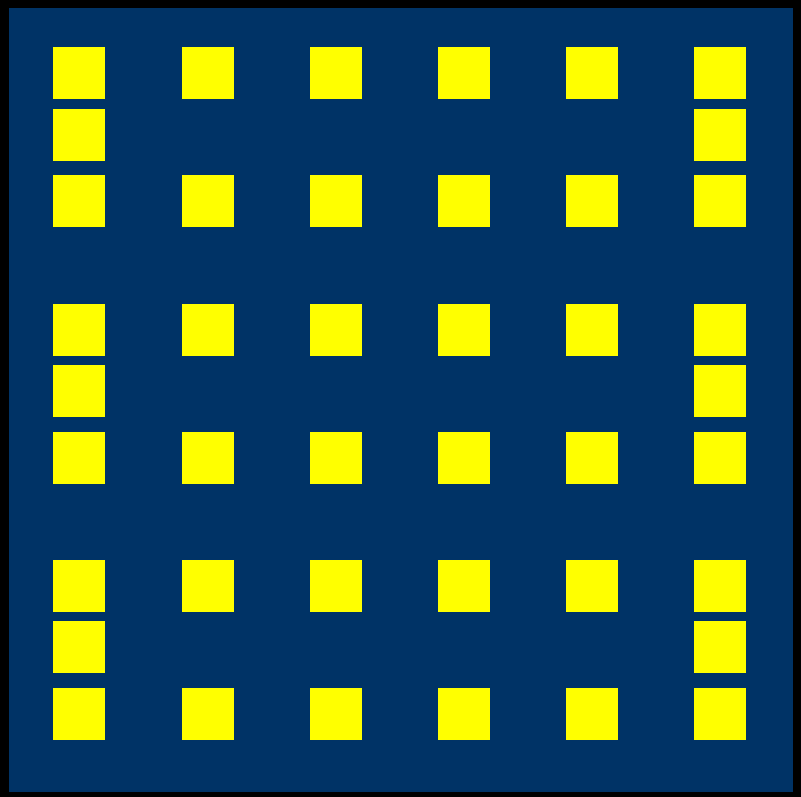


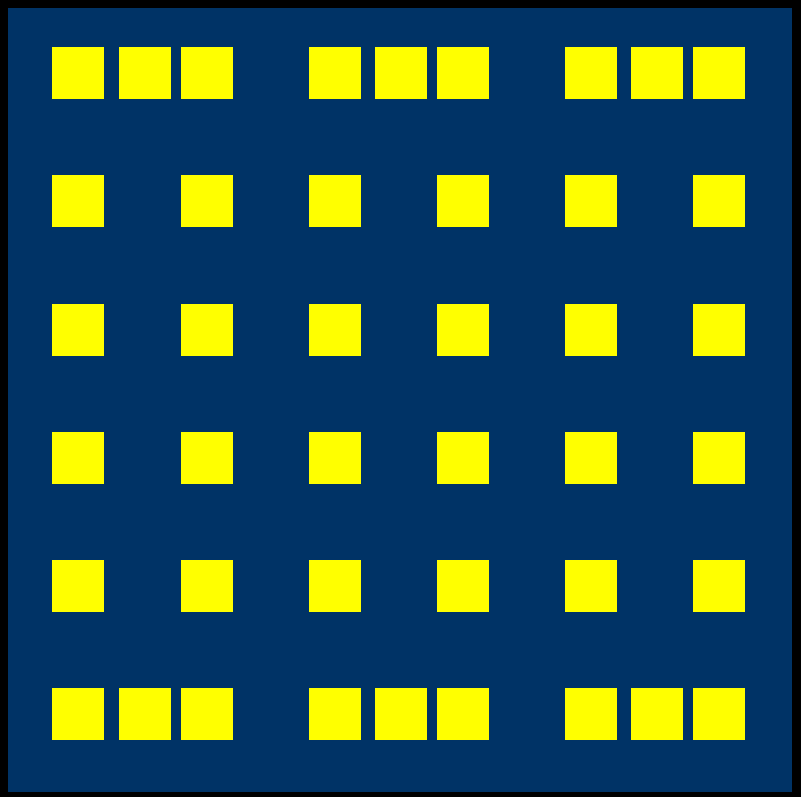


© Stephen E. Palmer, 2002

Perceptual Grouping

Closure

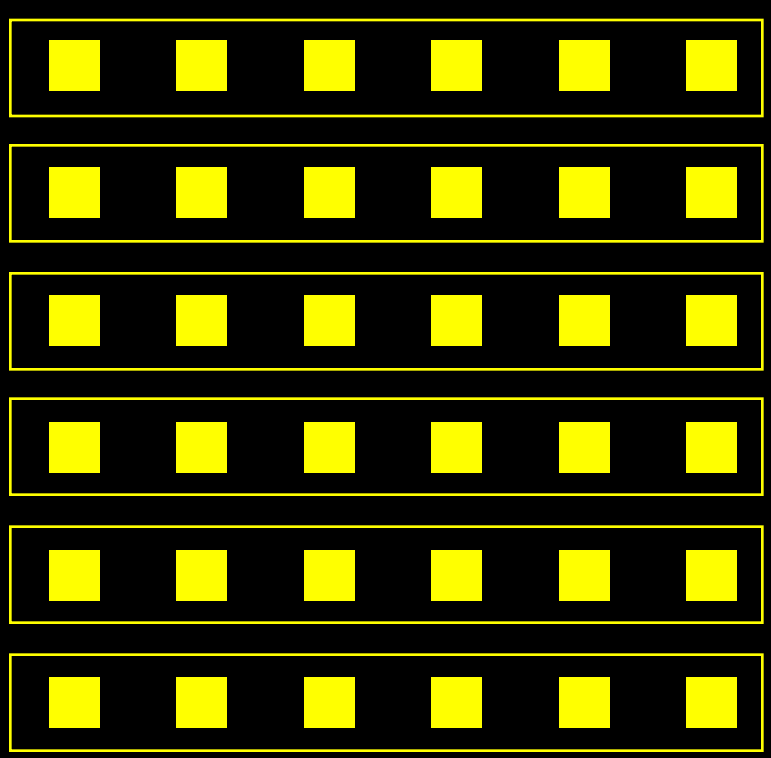


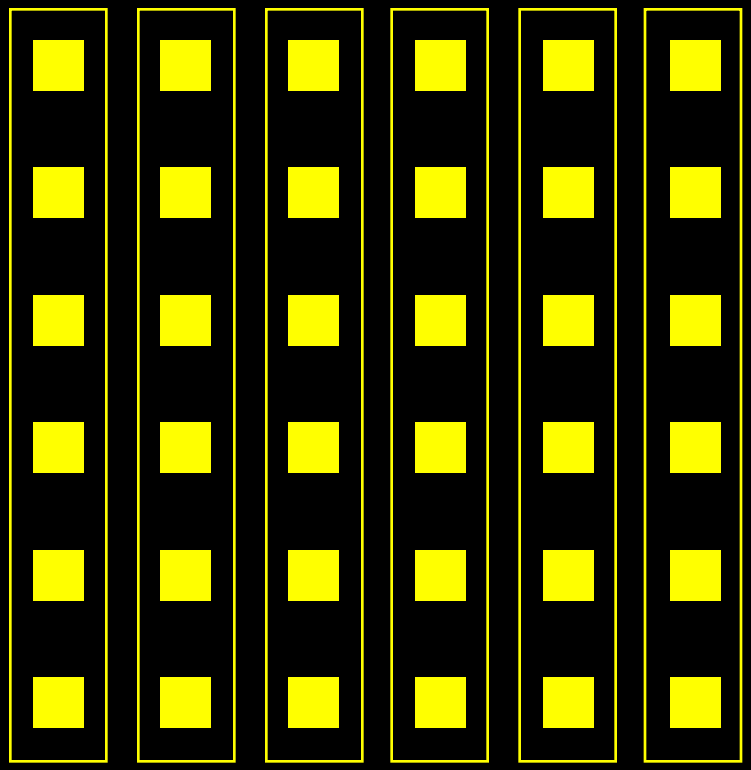


© Stephen E. Palmer, 2002

Perceptual Grouping

Common Region

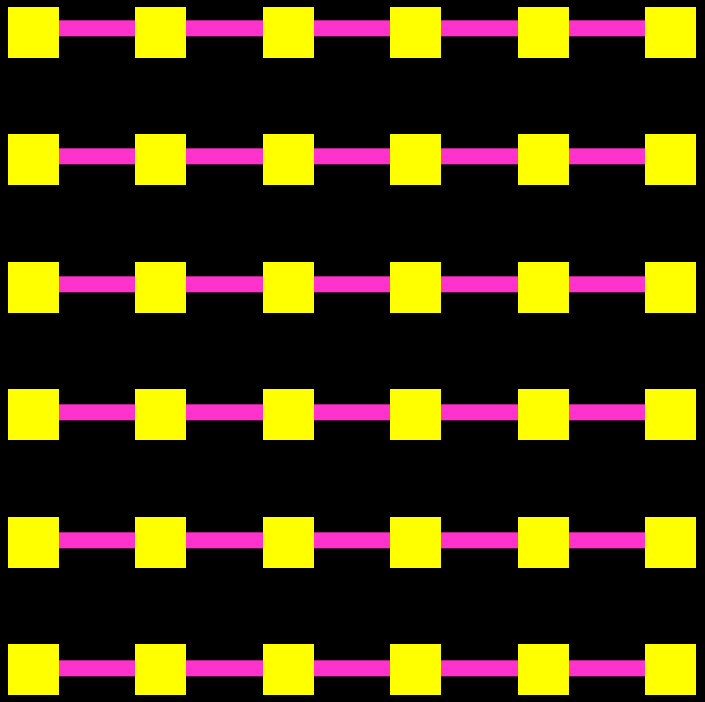


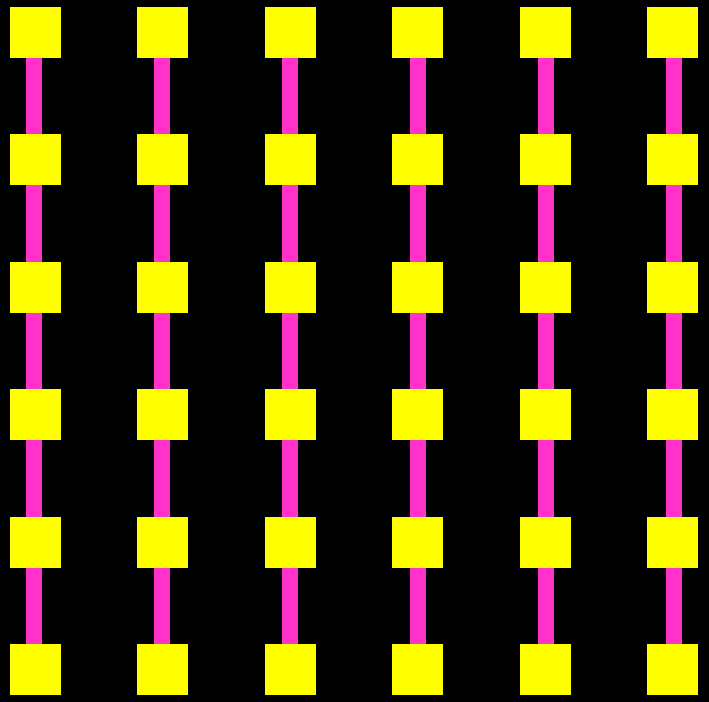


© Stephen E. Palmer, 2002

Perceptual Grouping

Element Connectedness





© Stephen E. Palmer, 2002

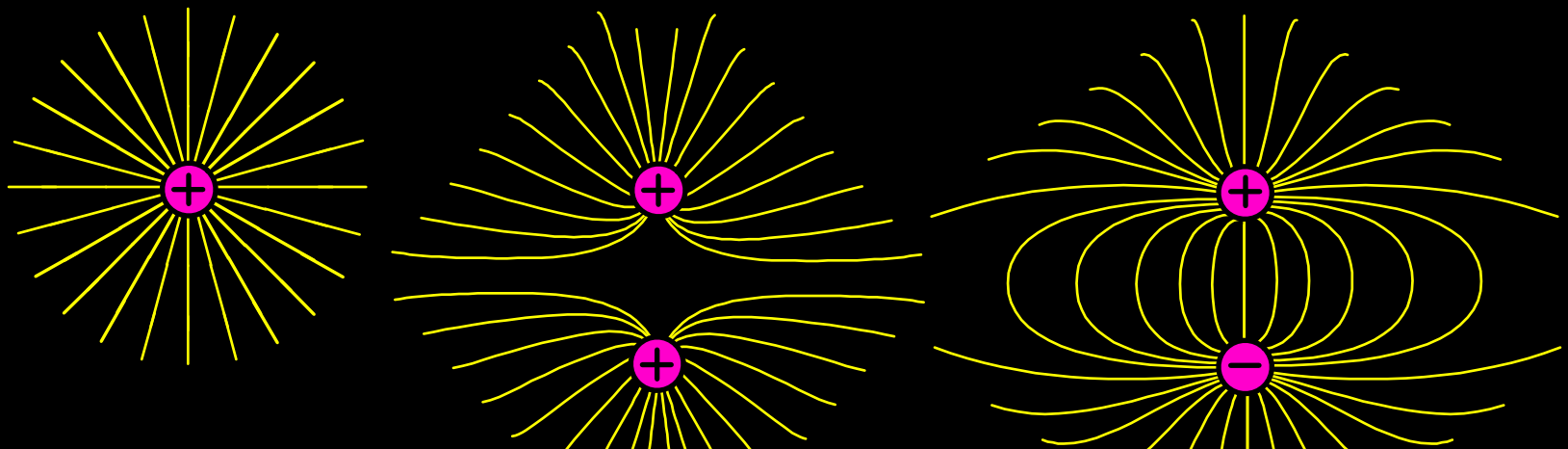
Perceptual Grouping

Past Experience

© Stephen E. Palmer, 2002

Gestaltism

Field Theoretic Analogy



Field of a single positive charge

Field of a pair of positive charges

Repulsion

Field of a positive & a negative charge

Attraction

© Stephen E. Palmer, 2002

Gestaltism

THEORY	NATIVISM vs. EMPIRICISM	ATOMISM vs. HOLISM	ORGANISM vs. ENVIRONMENT	PRINCIPAL ANALOGY	METHOD
Structuralism	Empiricism	Atomism	Organism	Chemistry	Trained Introspection
Gestaltism	Nativism	Holism	Organism	EM Fields	Naïve Introspection
Ecological Optics					
Constructivism					

Ecological Optics

Ecological Optics: Perception is the direct apprehension of the visible environment by extracting invariants in the dynamic ambient optic array and the affordances of objects.

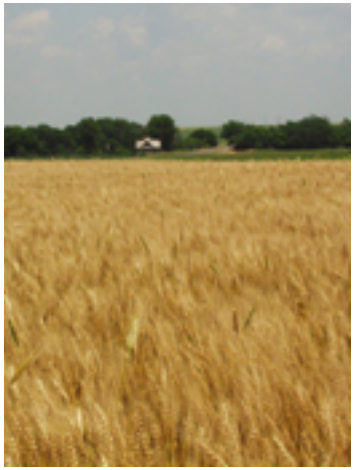
James J. Gibson: Ask not what's inside your head, but what your head's inside of! (Ecological optics is a theory of stimulus structure.)

How does the world structure the light stimulus received by the brain?

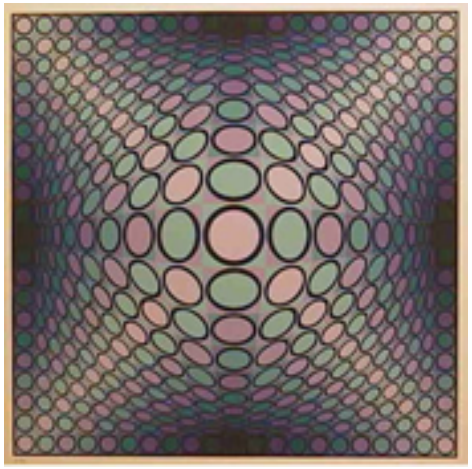
Ecological Optics

Texture Gradients are gradual changes in the size & shape of texture elements in an image when a plane recedes in depth.

Tile floor



Wheat field

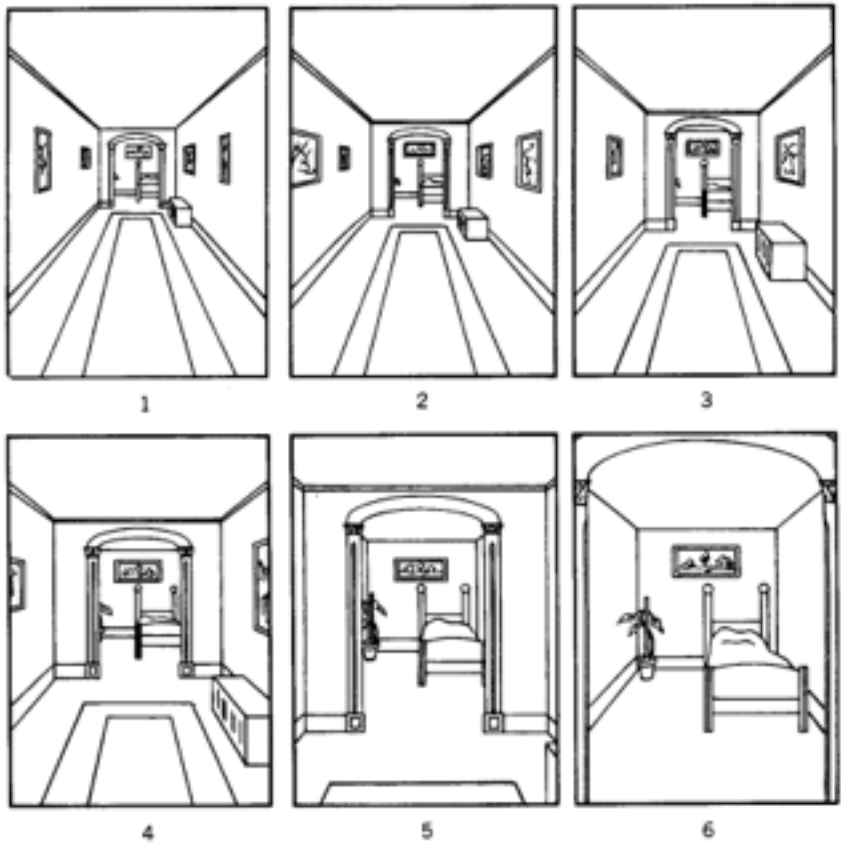


Geometric surface

© Stephen E. Palmer, 2002

Ecological Optics

Perception of the 3-D world is NOT ambiguous for an actively exploring organism.

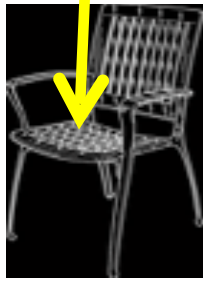


Direct Perception: Perception is not mediated by processes like unconscious thinking

affordances

- **Affordances:** Opportunities of interaction in the scene.
- Infer functions rather than identities.

Sittable



Ecological

THEORY	NATIVISM vs. EMPIRICISM	ATOMISM vs. HOLISM	ORGANISM vs. ENVIRONMENT	PRINCIPAL ANALOGY	METHOD
Structuralism	Empiricism	Atomism	Organism	Chemistry	Trained Introspection
Gestaltism	Nativism	Holism	Organism	EM Fields	Naïve Introspection
Ecological Optics	Nativism	Holism	Environment	Resonance	Ecological Analysis
Constructivism					

Constructivism

Constructivism: Perception is the result of unconscious inferences about the scene most likely to have caused the retinal image or event.

Hermann von Helmholtz originated the idea of unconscious inference and the likelihood principle.

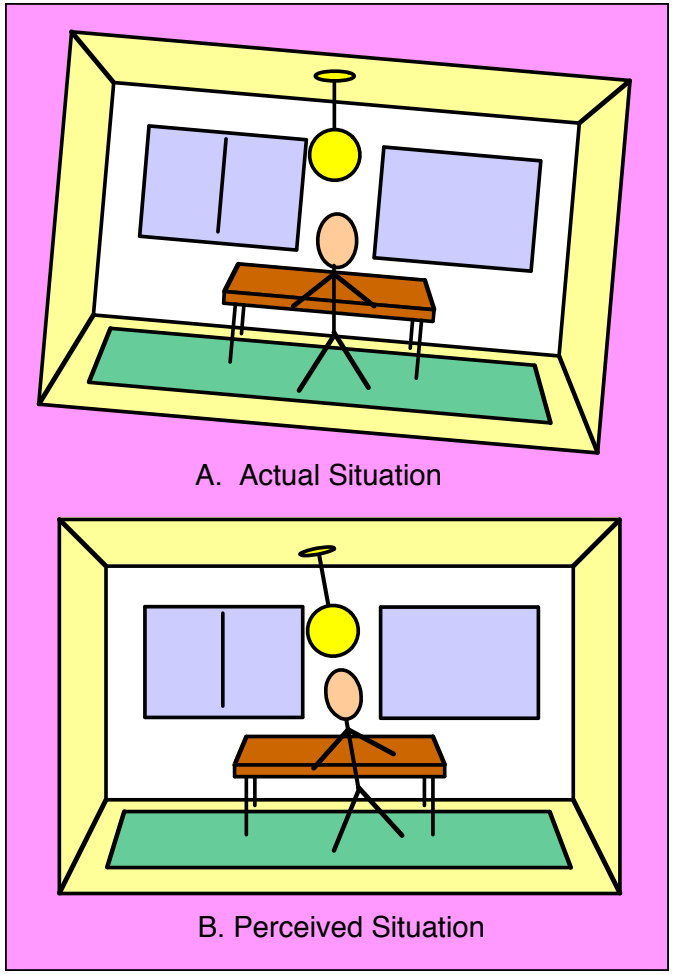
Constructivism

Unconscious Inference: the process of recovering environmental information by logically combining retinal information with heuristic assumptions.

Tilted room illusion:
If you assume that the walls and floor of the room are vertical and horizontal, then you must be tilted —and you feel that way!

Constructivism

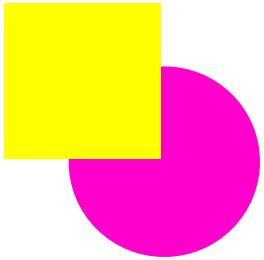
Tilted room illusion



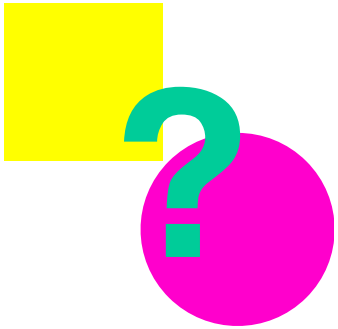
Constructivism

Prägnanz vs. Likelihood

What governs what we see: goodness or probability?



What is this?



square & circle?

square & pacman?

squigit & pacman?

etc….

Constructivsm

THEORY	NATIVISM vs. EMPIRICISM	ATOMISM vs. HOLISM	ORGANISM vs. ENVIRONMENT	PRINCIPAL ANALOGY	METHOD
Structuralism	Empiricism	Atomism	Organism	Chemistry	Trained Introspection
Gestaltism	Nativism	Holism	Organism	EM Fields	Naïve Introspection
Ecological Optics	Nativism	Holism	Environment	Resonance	Ecological Analysis
Constructivism	both	both	both	Logical Inference	Behavior

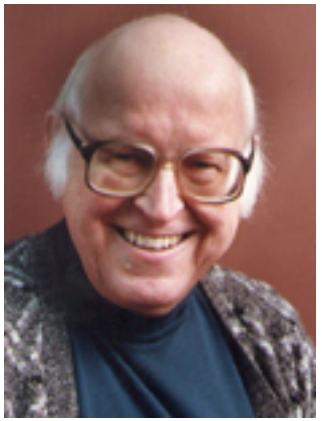
So which theory is correct?

Probably none of them!

Or maybe all of them, to some degree!

It's good to keep them in mind, when designing your algorithms!

"You can't play 20-questions with nature and win!"
(Allen Newell, 1973)



© Stephen E. Palmer, 2002

Recap-Theories of Human Vision

- Importance of theories
- Issues to answer in creating a theory for vision
  - Empiricism vs. Nativism
  - Organism vs. Environment
  - Atomism vs. Holism
- Theories
  - Structuralism
  - Gestaltism
  - Ecology
  - Constructivism

Enter the computers

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Computational Thinking

+ Real Images rather than made up Stimuli

+ No more vague theories. Needed more details.

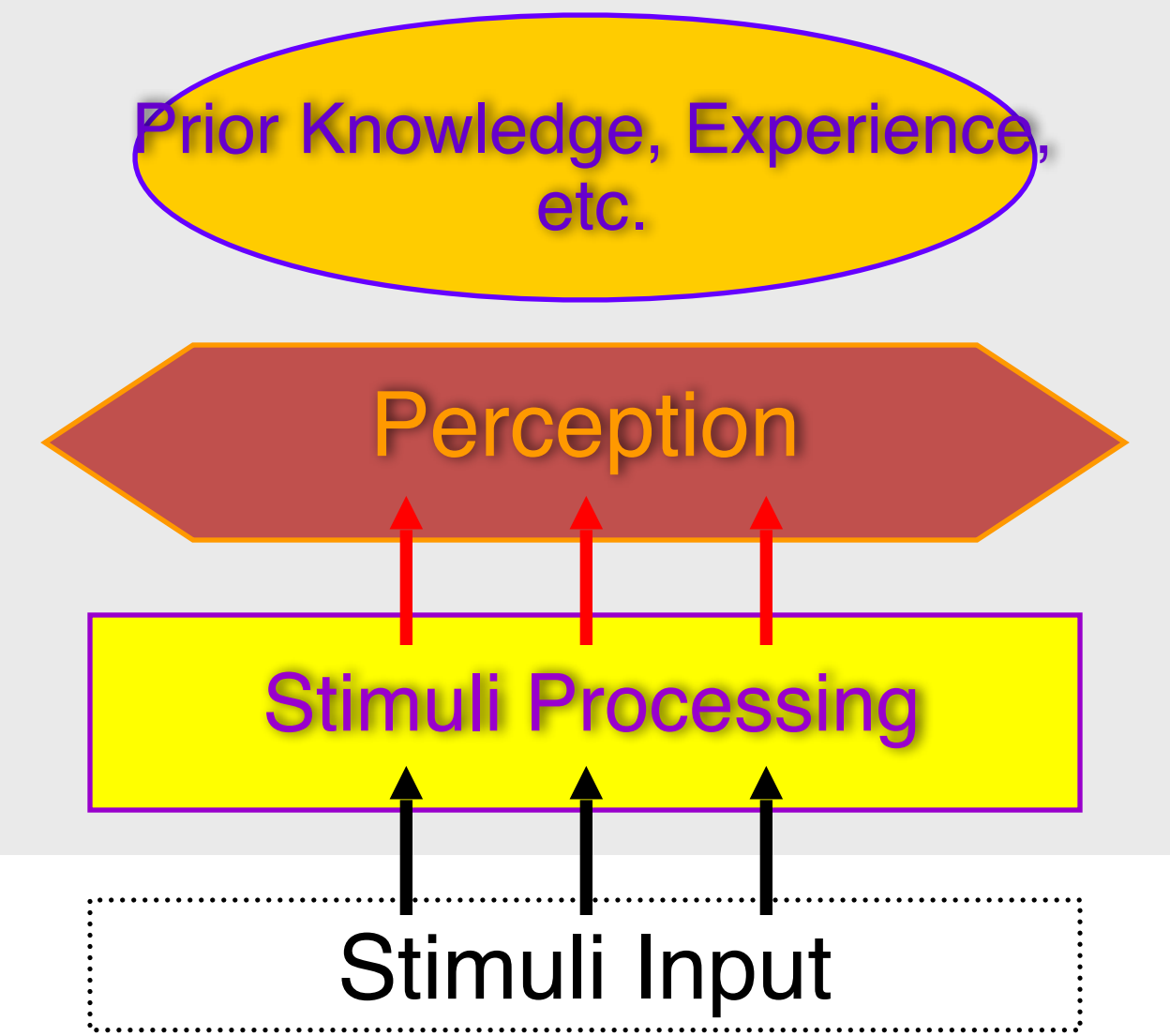
+ Principles of Information Processing

Top-down vs. Bottom-up

Top-Down versus Bottom-Up Perception

- Top-Down
  - Perceive the whole and then individual parts as needed.
  - Experience-driven as opposed to stimulus or input-data driven.
  - Quick and highly inferential but also a source of misperception.
- Bottom-up
  - Perceive the individual parts and organize them into a whole, if possible.
  - Information available in the stimulus itself.

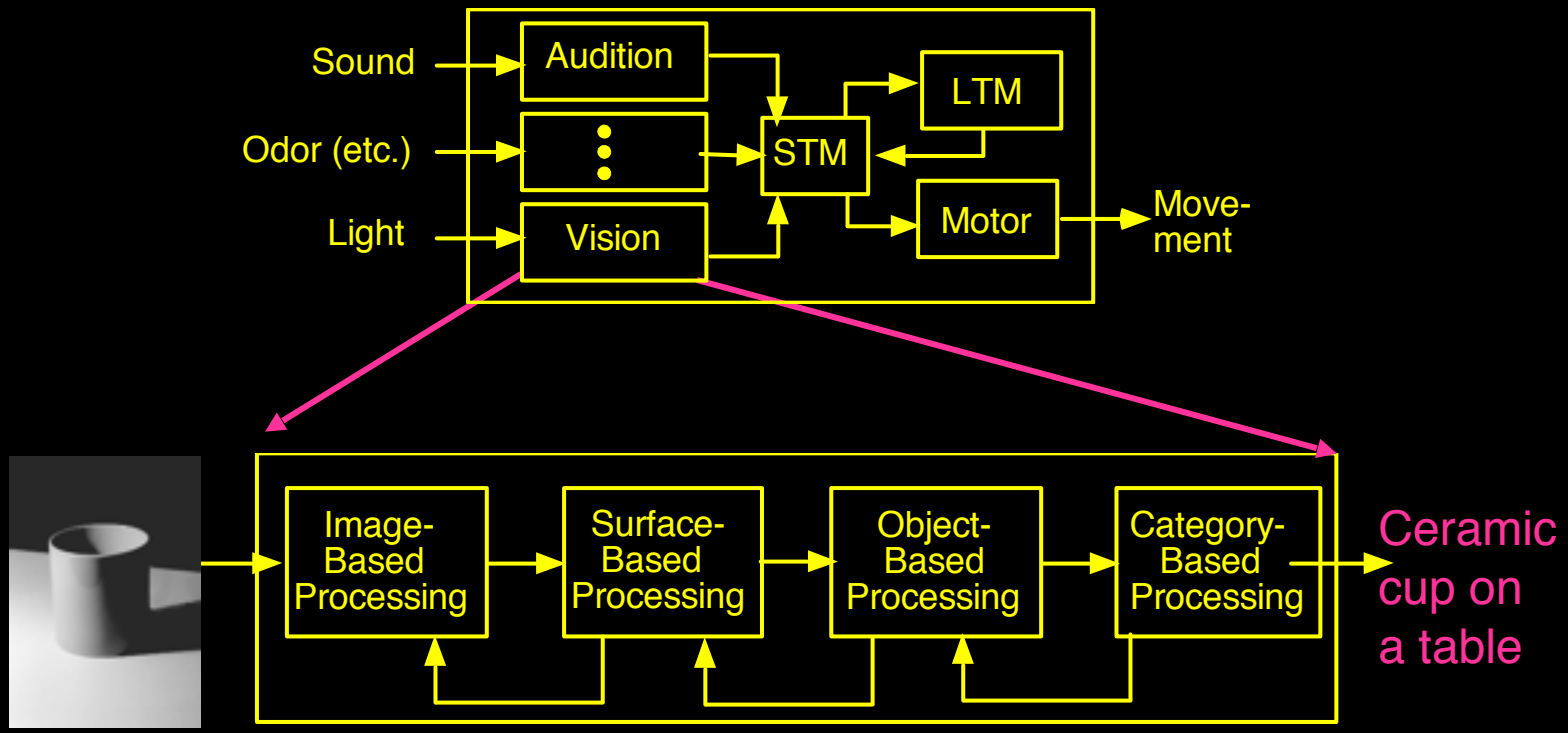
Bottom-Up Processing



Created by Dr. Gordon Vessels 2005

David Marr’s Theory

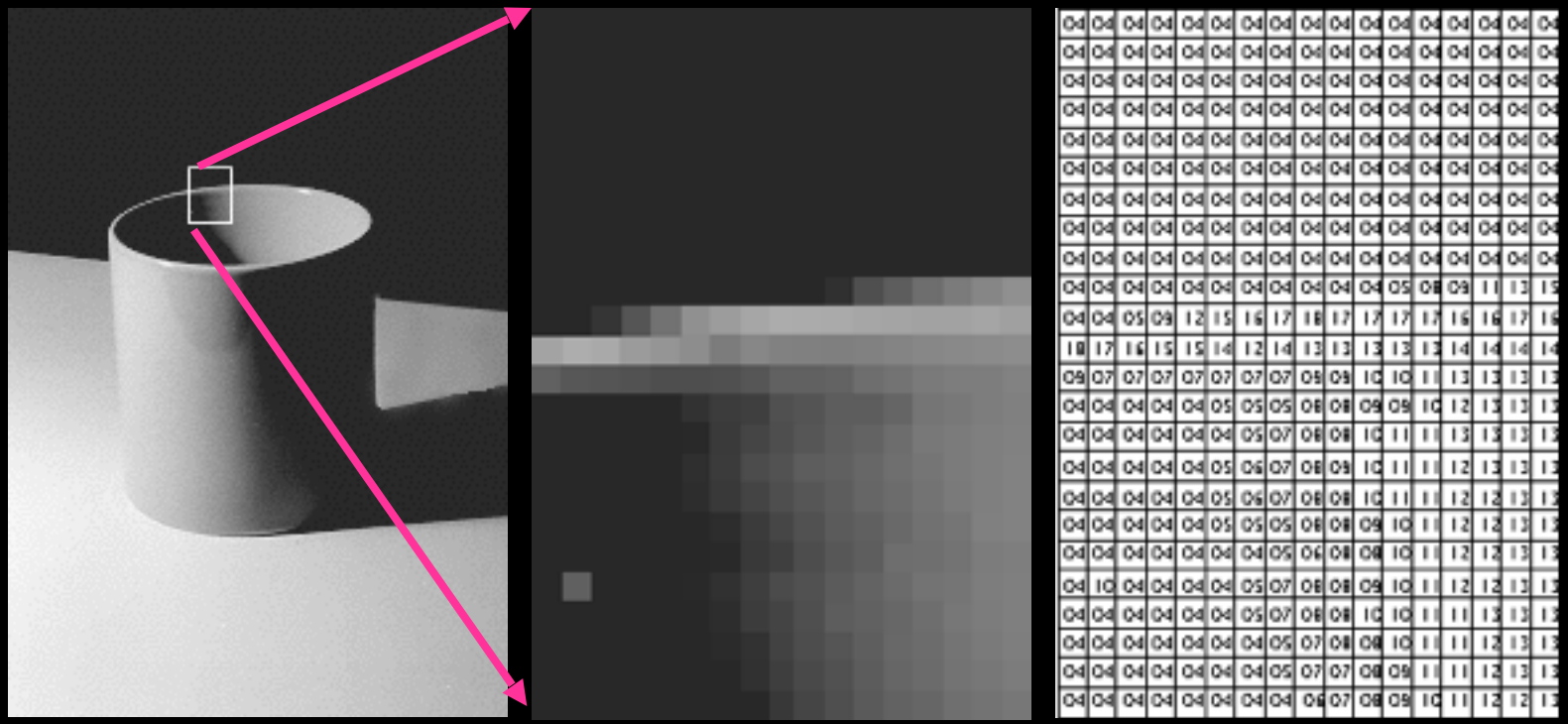
Four Stages of Visual Perception



© Stephen E. Palmer, 2002

Four Stages of Visual Perception

The Retinal Image

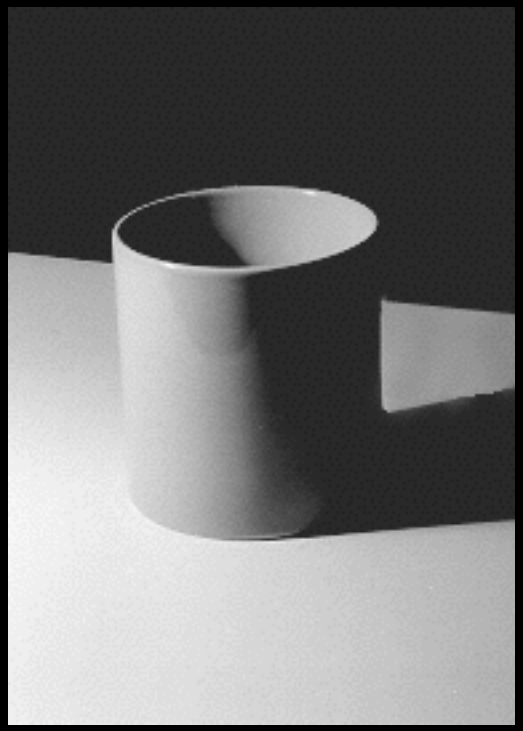


An Image (blowup) Receptor Output

© Stephen E. Palmer, 2002

Four Stages of Visual Perception

Retinal Image

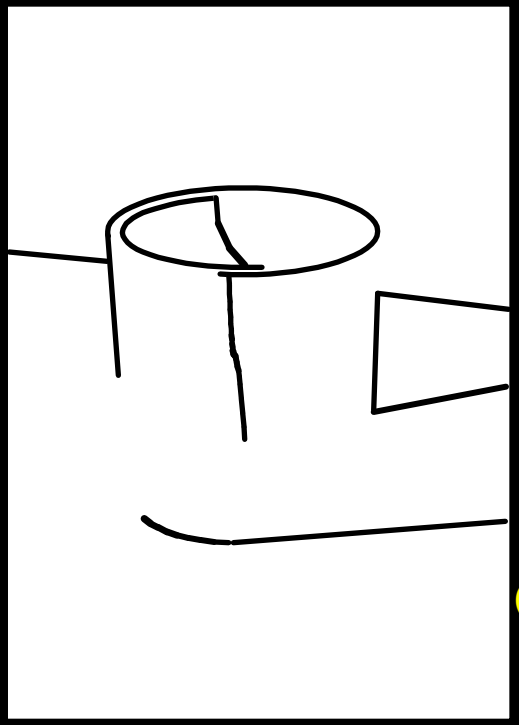


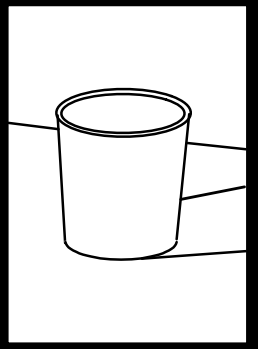
An Image

Image-based processes

Edges
Lines
Blobs
etc.

Image-based Representation





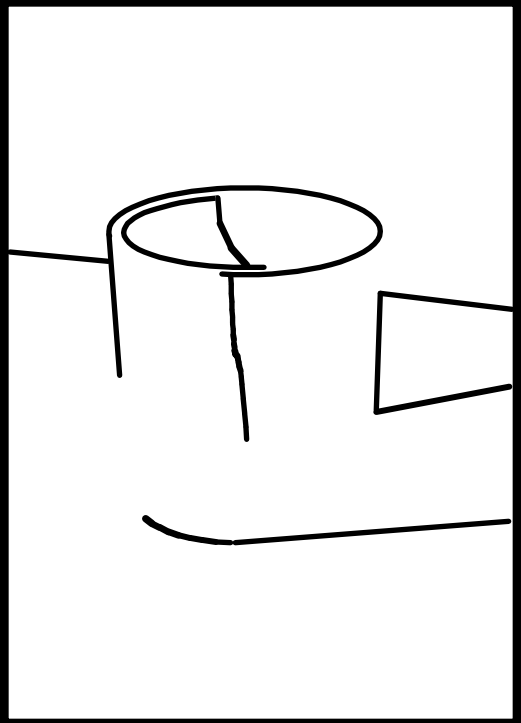
(Line Drawing)

Primal Sketch (Marr)

© Stephen E. Palmer, 2002

Four Stages of Visual Perception

Image-based Representation

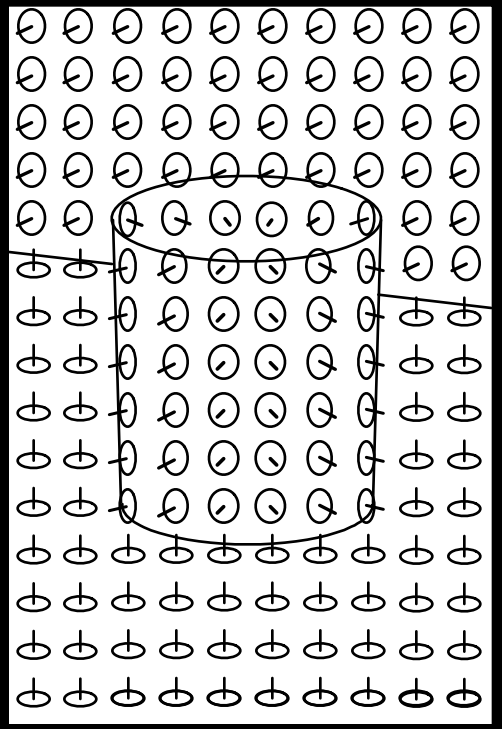


Primal Sketch

Surface-based processes

Stereo
Shading
Motion
etc.

Surface-based Representation



2.5-D Sketch

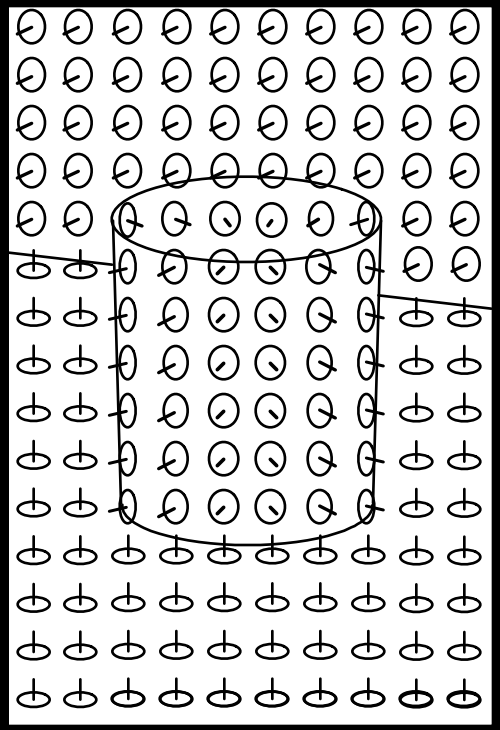
© Stephen E. Palmer, 2002

Surfaces etc.

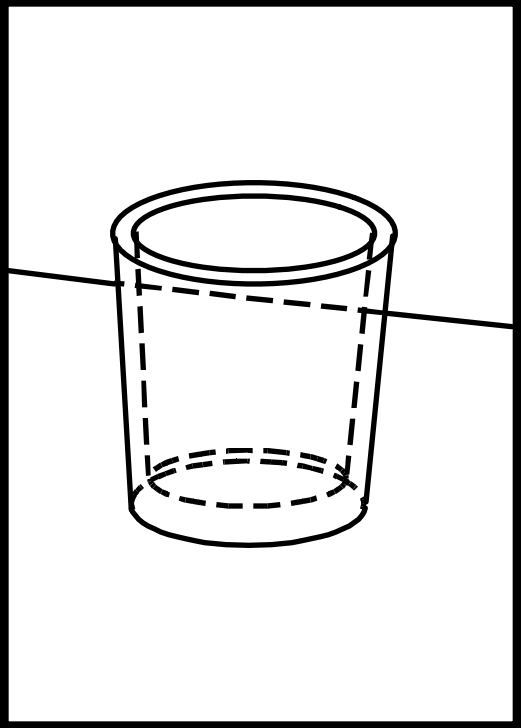
- Initially proposed by Gibson (1950).
- Departure from Object-based 3-D representations.
- But what is this 2.5D and how it is inferred (especially in 2D images).

Four Stages of Visual Perception

Surface-based Representation



Object-based Representation



Object-based processes

Grouping
Parsing
Completion
etc.

2.5-D Sketch

Volumetric Sketch

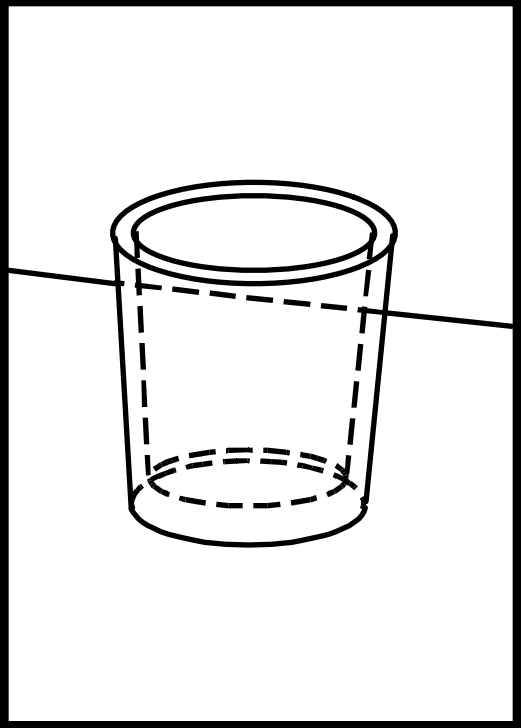
© Stephen E. Palmer, 2002

Object-based

- **Move from surfaces to volumes.**
- **Seeing the hidden layer.**
- **But how do you move from surface to volumes? Bottom-up or Top-down?**

Four Stages of Visual Perception

Object-based Representation



Volumetric Sketch

Category-based processes

Pattern-Recognition

Spatial-description

Category-based Representation

Category: cup

Color: light-gray

Size: 6”

Location: table

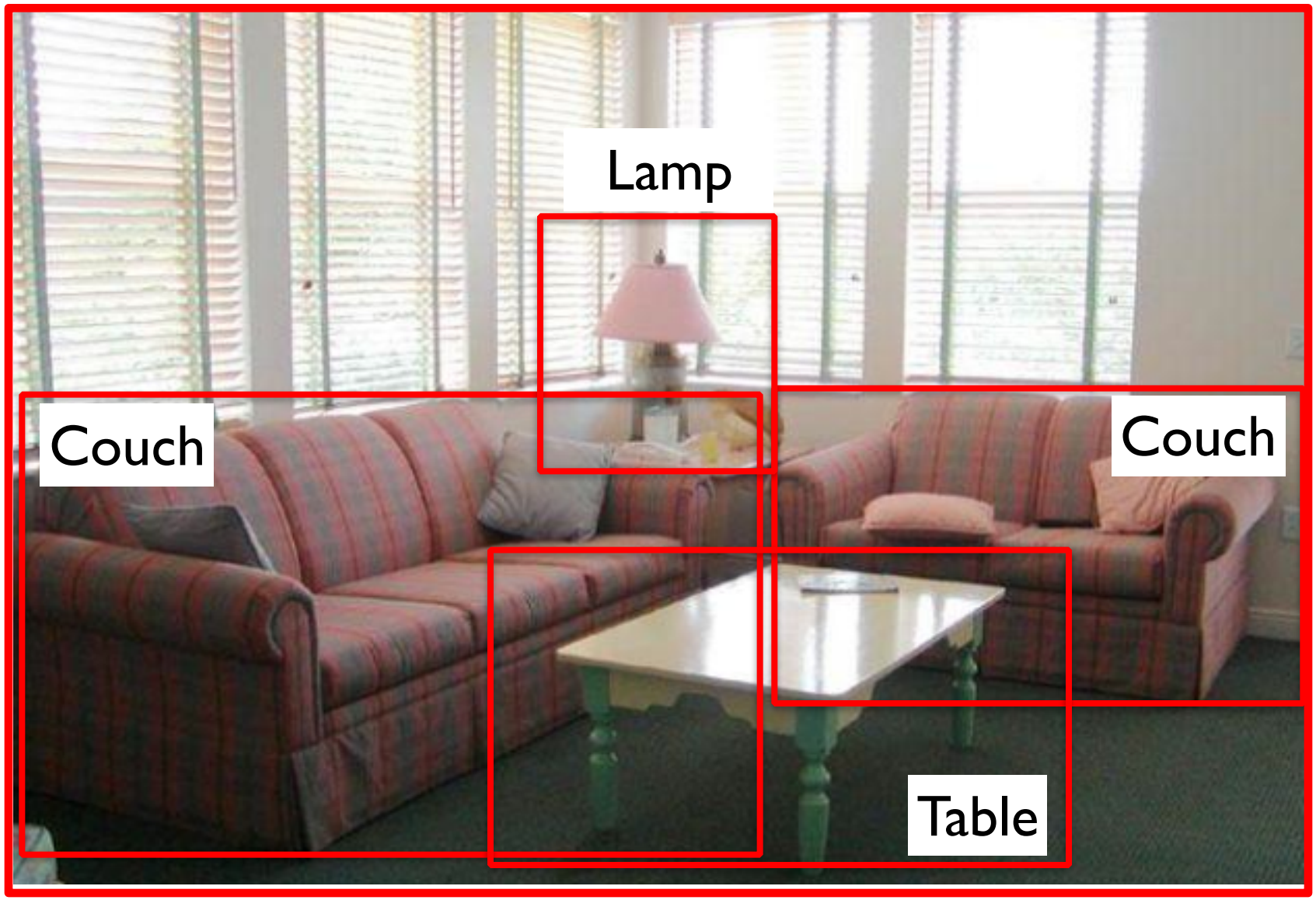
Basic-level Category

© Stephen E. Palmer, 2002

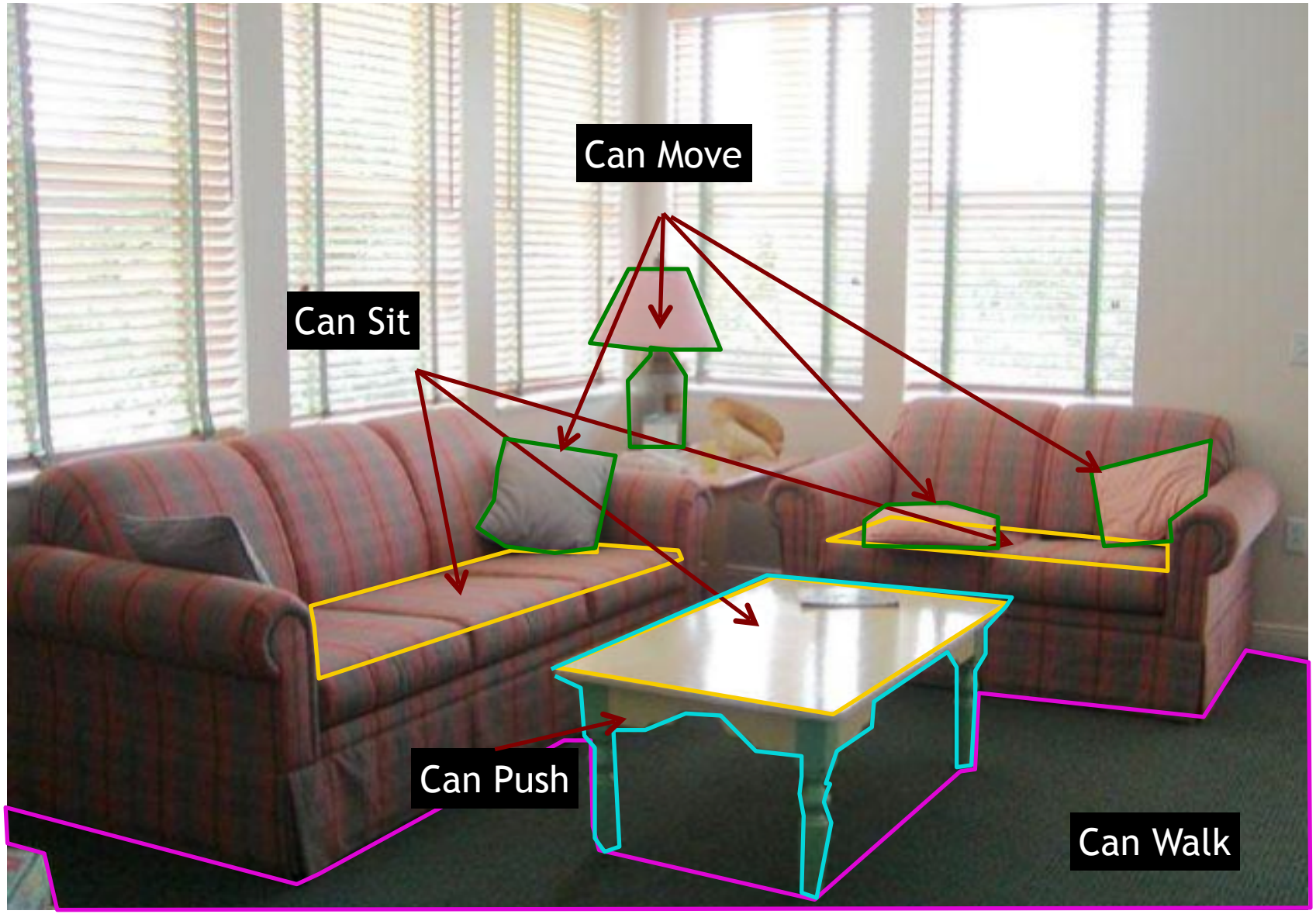
Category-based

- Move from volumes to categories.
- Essentially this step relates to association of object with memory elements.
- But what are the right categories?

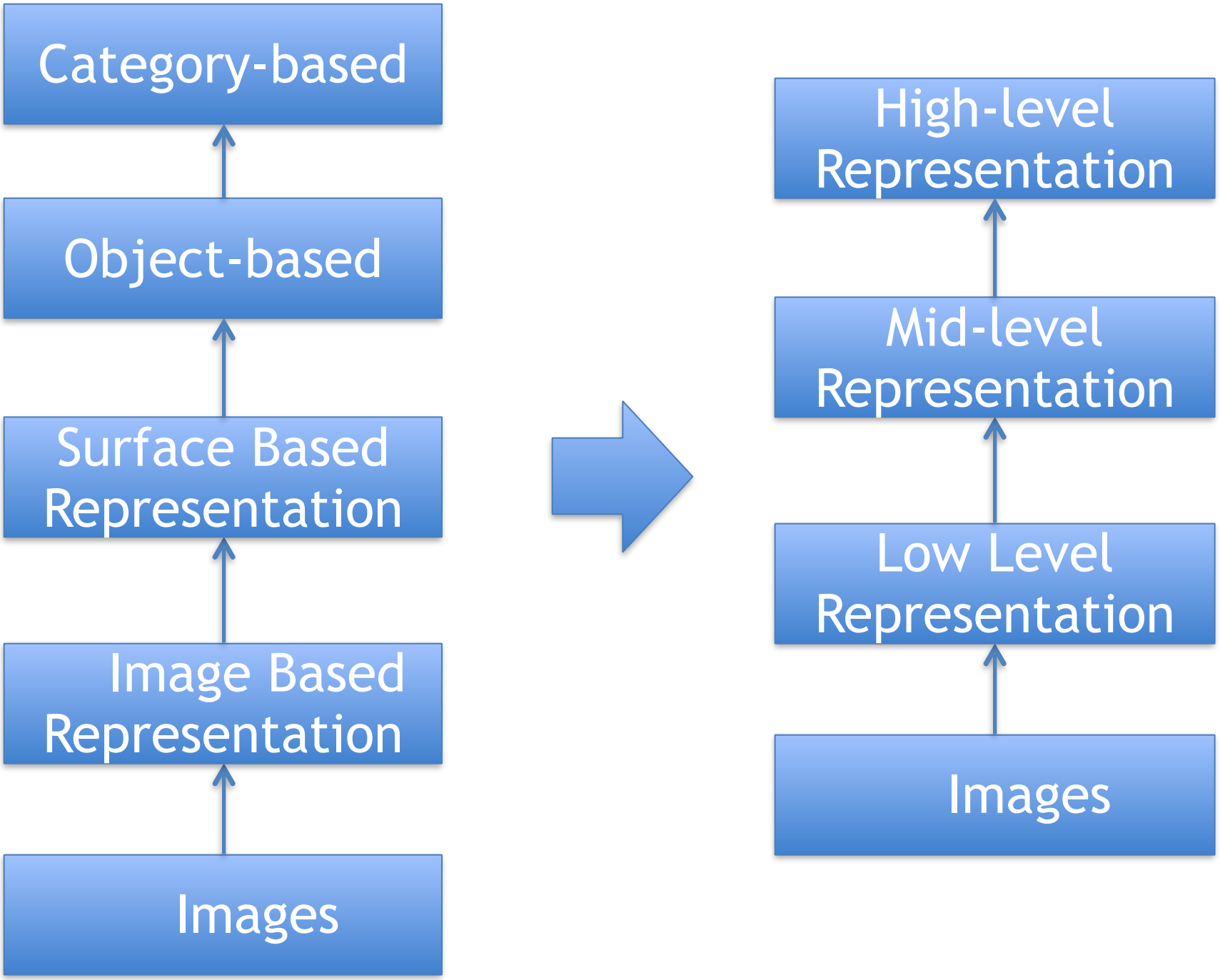
Semantic?



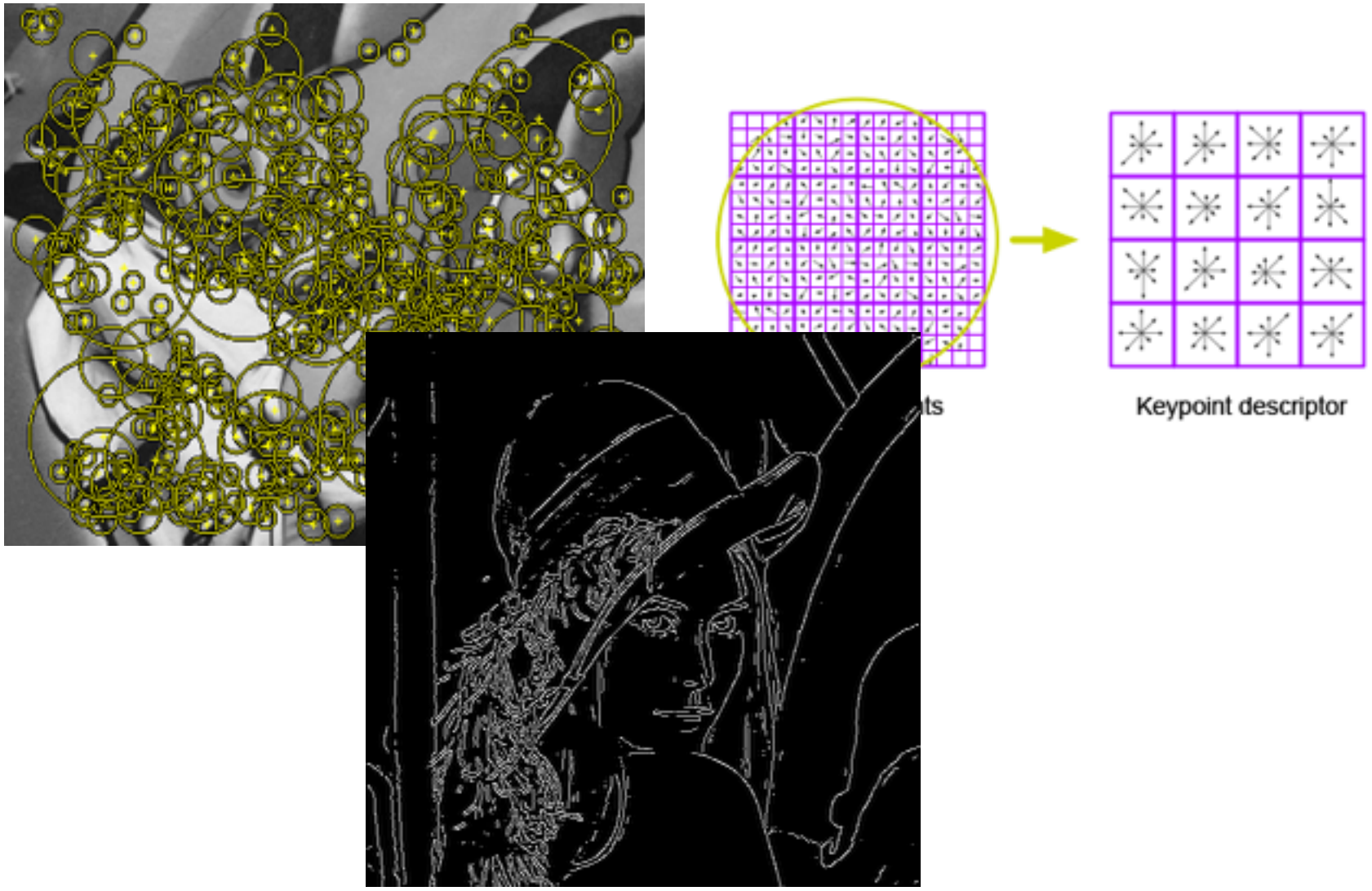
Functional?



Exemplars maybe?

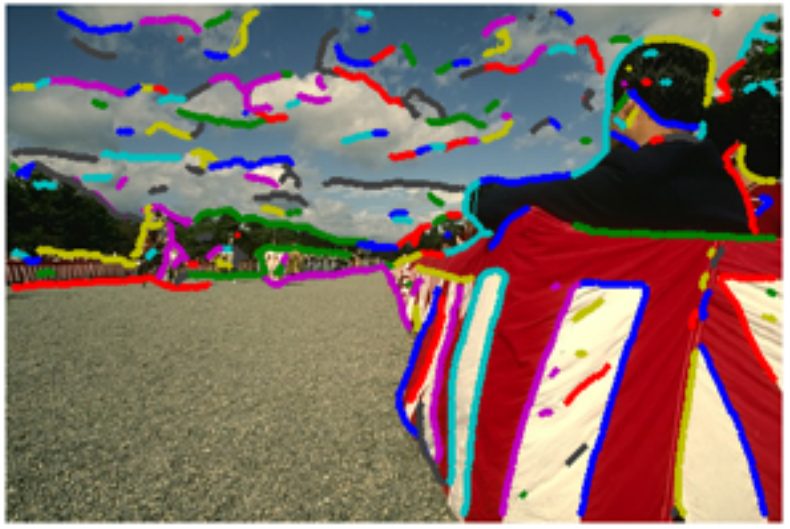

Category-based
Object-based
Surface Based Representation
Image Based Representation
Images
High-level Representation
Mid-level Representation
Low Level Representation
Images

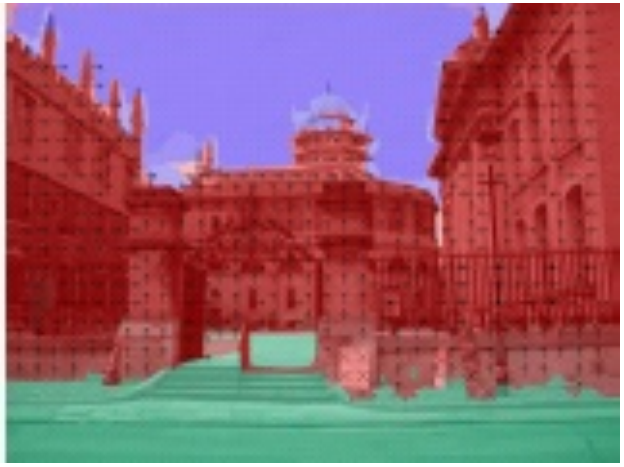
Low-Level Representation



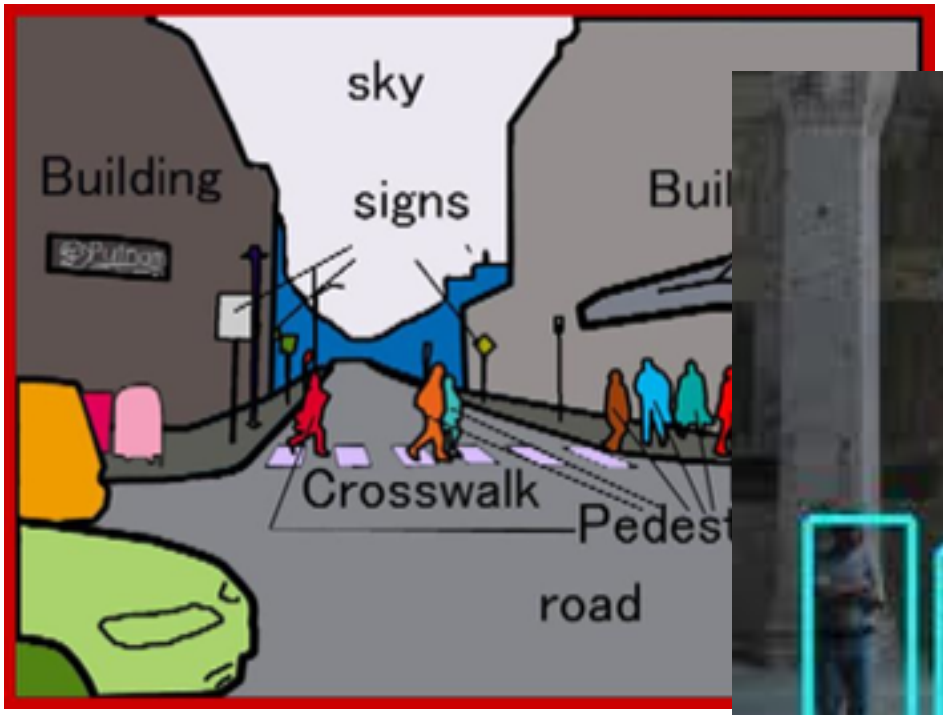
Mid-level Representations

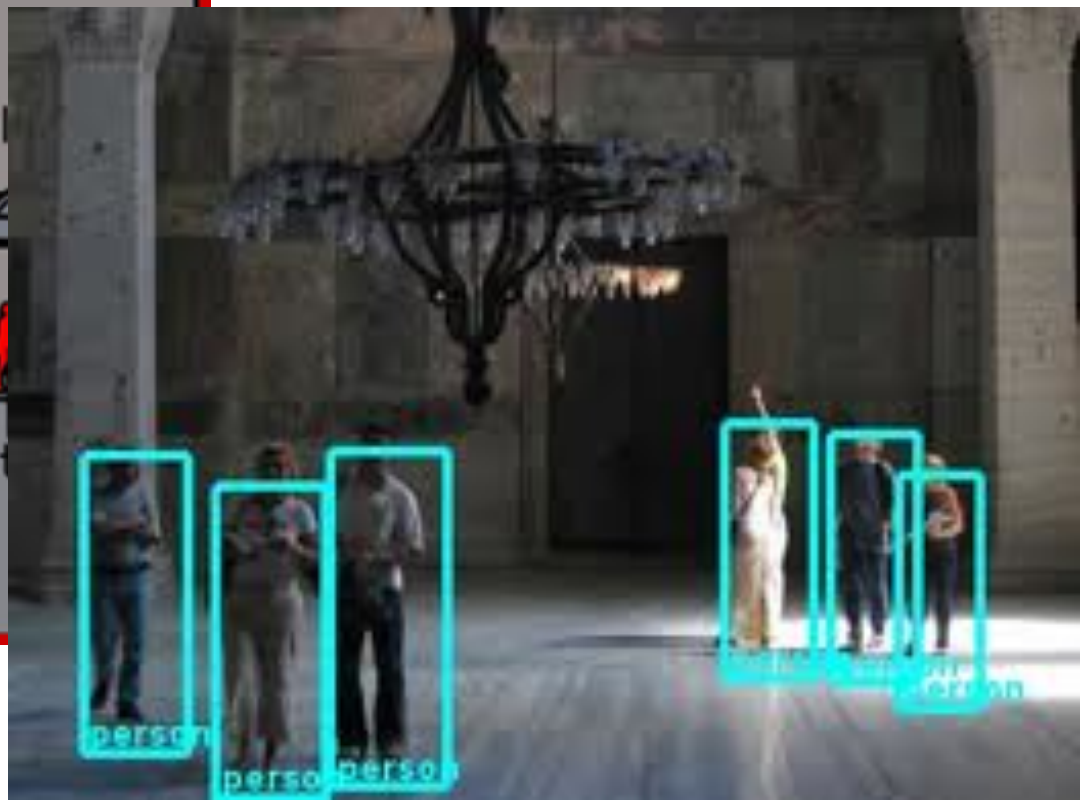
- Grouping, 3D Properties



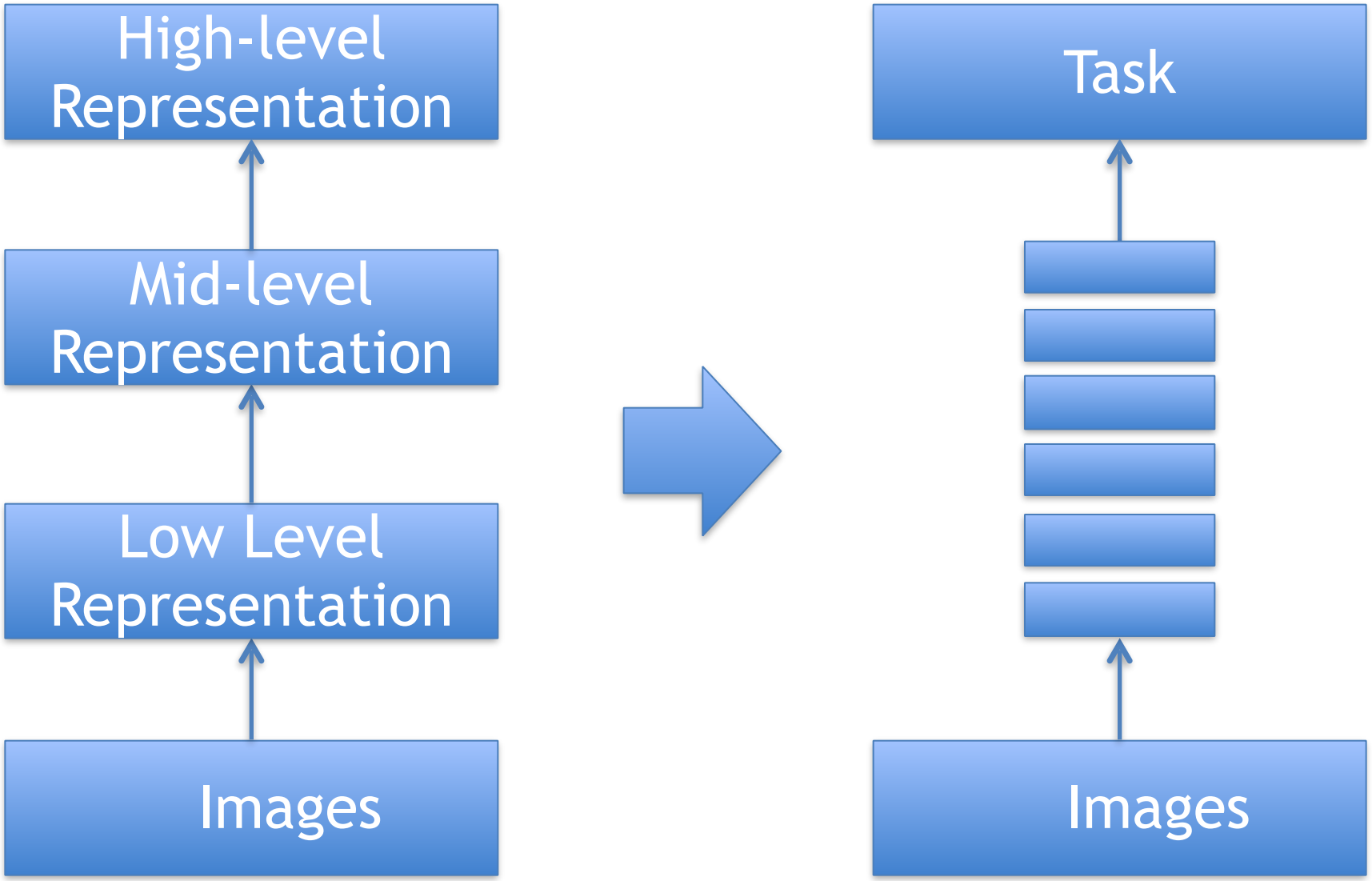


High-level Representation





- But what are the boundaries?
- How does the information flow?


High-level Representation
Mid-level Representation
Low Level Representation
Images
Task
Images

In Search of Objects: 50 years of wondering

Geometric Era

- Influenced by Ecological theory: Computer Vision started by geometric modeling the Visual World (Objects).

Geometric Era

- Influenced by Ecological theory: Computer Vision started by geometric modeling the Visual World (Objects).
    - Invariance to Viewpoint (Geometric Modeling allowed prediction under perspective).

Geometric Era

- Influenced by Ecological theory: Computer Vision started by geometric modeling the Visual World (Objects).

  - Invariance to Viewpoint (Geometric Modeling allowed prediction under perspective).
  - Invariance to Illumination (Matching was based on edges which were robust to illumination variations.

Geometric Era

- Influenced by Ecological theory: Computer Vision started by geometric modeling the Visual World (Objects).

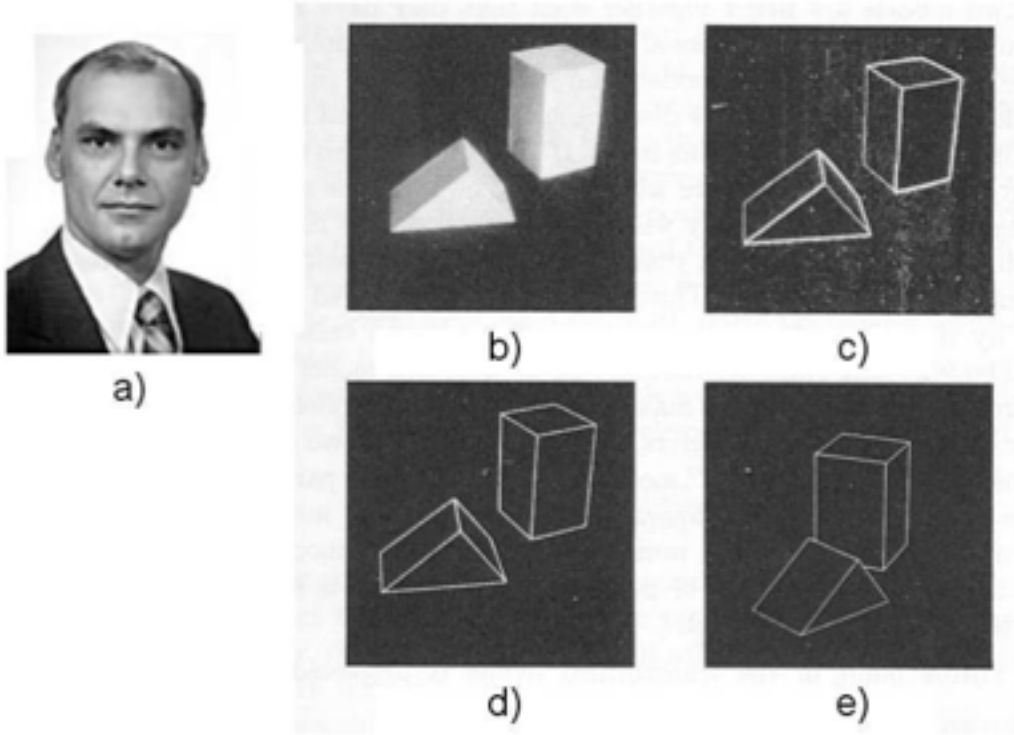
  - Invariance to Viewpoint (Geometric Modeling allowed prediction under perspective).
  - Invariance to Illumination (Matching was based on edges which were robust to illumination variations.
  - Man-made World: Most objects are designed by humans using CAD..therefore rightly modeled using geometric primitives

Geometric Era

- Influenced by Ecological theory: Computer Vision started by geometric modeling the Visual World (Objects).

  - Invariance to Viewpoint (Geometric Modeling allowed prediction under perspective).
  - Invariance to Illumination (Matching was based on edges which were robust to illumination variations.
  - Man-made World: Most objects are designed by humans using CAD..therefore rightly modeled using geometric primitives
  - Well formed theories in geometry.

Blocks World at MIT



L. G. Roberts, *Machine Perception of Three Dimensional Solids*, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

Fig. 1. A system for recognizing 3-d polyhedral scenes. a) L.G. Roberts. b)A blocks world scene. c)Detected edges using a 2x2 gradient operator. d) A 3-d polyhedral description of the scene, formed automatically from the single image. e) The 3-d scene displayed with a viewpoint different from the original image to demonstrate its accuracy and completeness. (b) - e) are taken from [64] with permission MIT Press.)

J. Mundy, Object Recognition in the Geometric Era: a Retrospective, 2006

Why Blocks World?

start with a simplification of the world so that the mathematical models can apply rigorously

Line Labeling and Reasoning

- Essentially there was no learning because there was not enough data.
- Edges were labeled convex and concave and the way these edges combine are using rules.

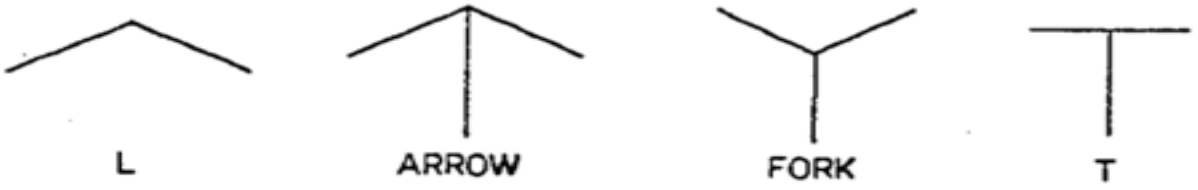
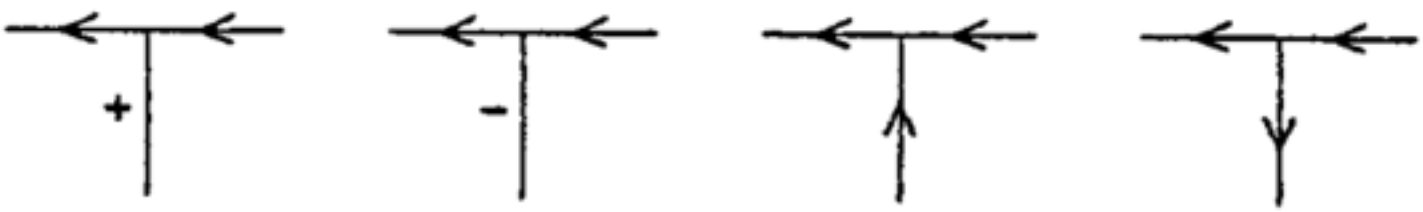
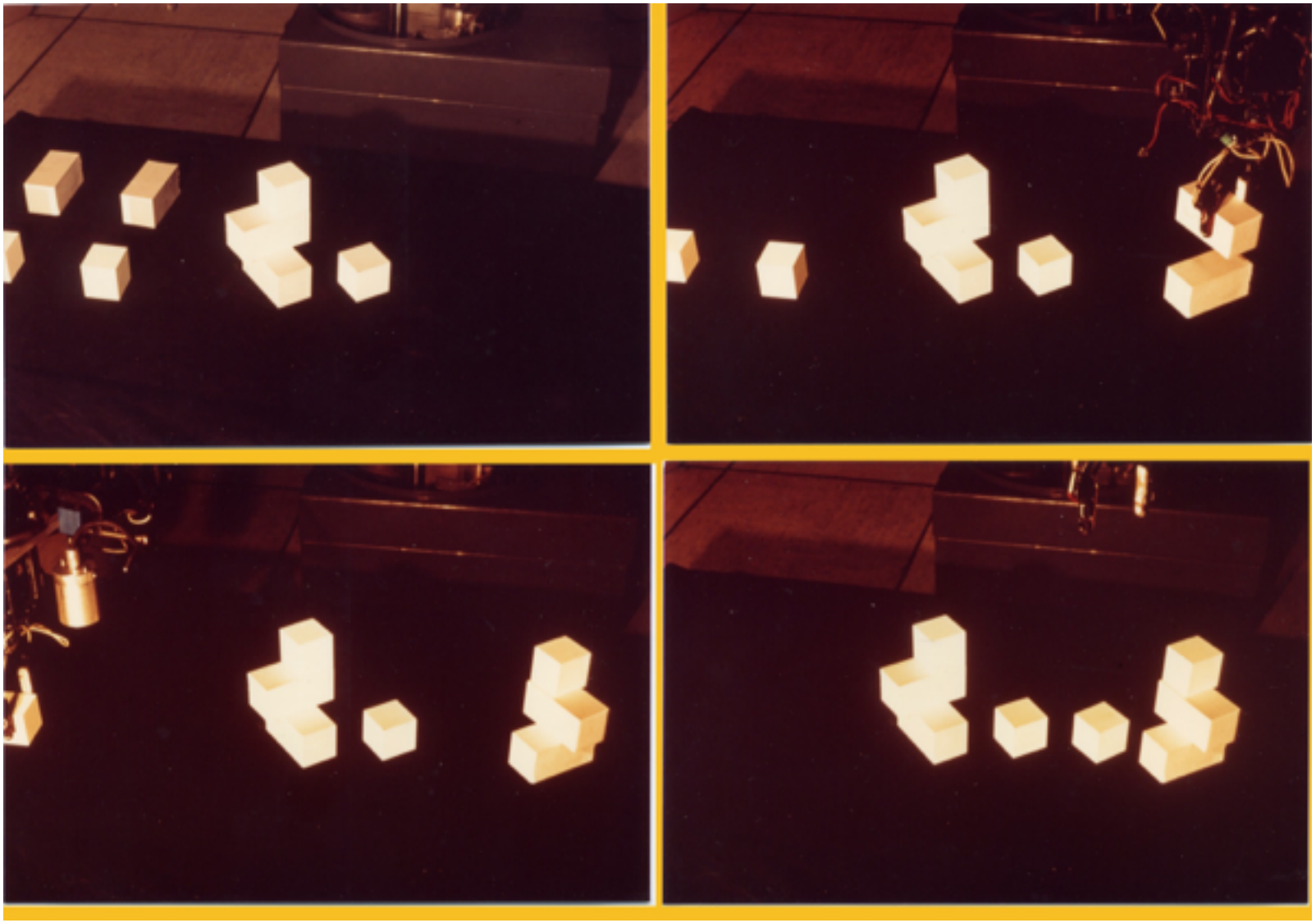


Figure 8 Junction types treated in this paper.



MIT Copy Demo



Why Blocks World?

Decade of Research but culminated soon…assumptions did not translate to real world

- Modeling Assumptions
  - Curved surfaces and boundaries;
  - articulated and moving objects;
- Edge Detection and Segmentation
  - complex background and 3-d texture such as foliage; – specular or mutually illuminating surfaces;
  - multiple light sources and remote shadowing; – transparent or translucent surfaces.
- Occlusion by unknown shapes;

Modeling Generalizations

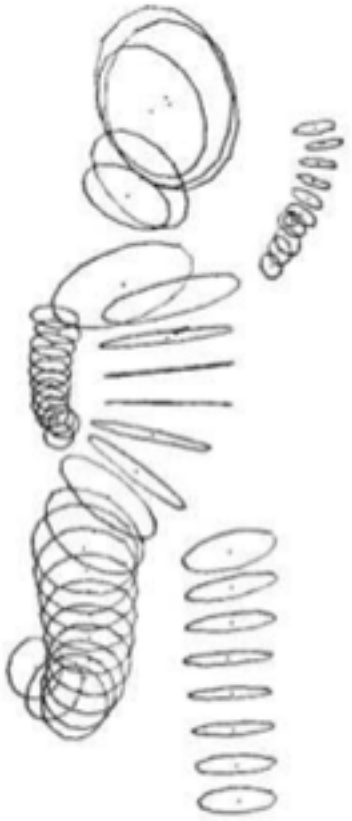
From Blocks World

a)

b)

c)

Generalized Cylinders

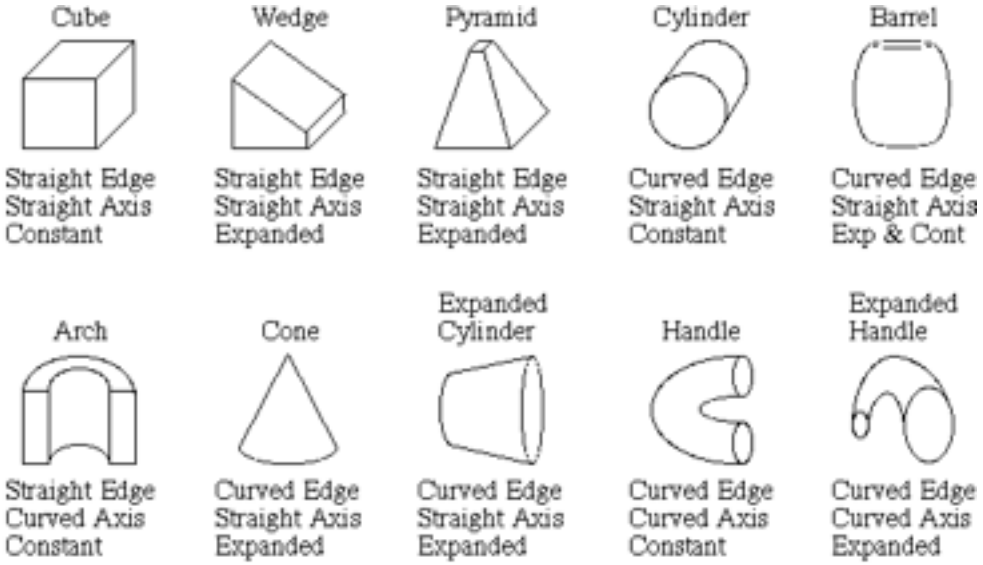


curved shapes can be expressed as a sweep of a variable cross section along a curved axis.

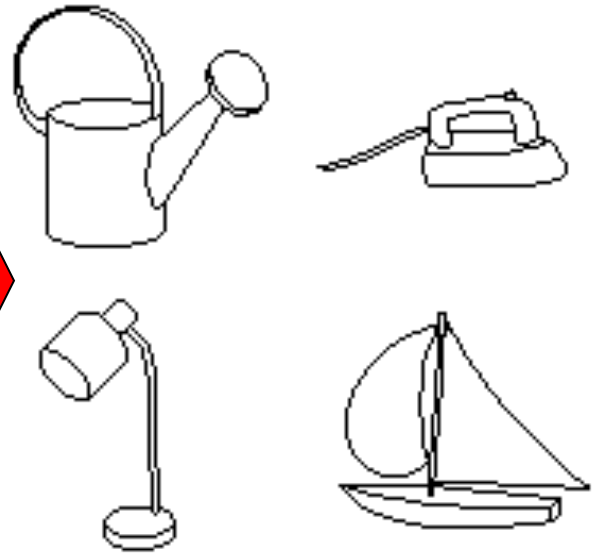
Recognition by components - Geons

Biederman (1987)

Primitives (geons)



Objects



http://en.wikipedia.org/wiki/Recognition_by_Components_Theory

Edge/Grouping Generalization

Aspect Graph

- Until 1970-80's, research focused on object-centered representation.
- No focus on modeling image intensity variations (model every-thing from edges)
- Alternative to 3D representation of objects
- How would appearance change under different viewpoint?

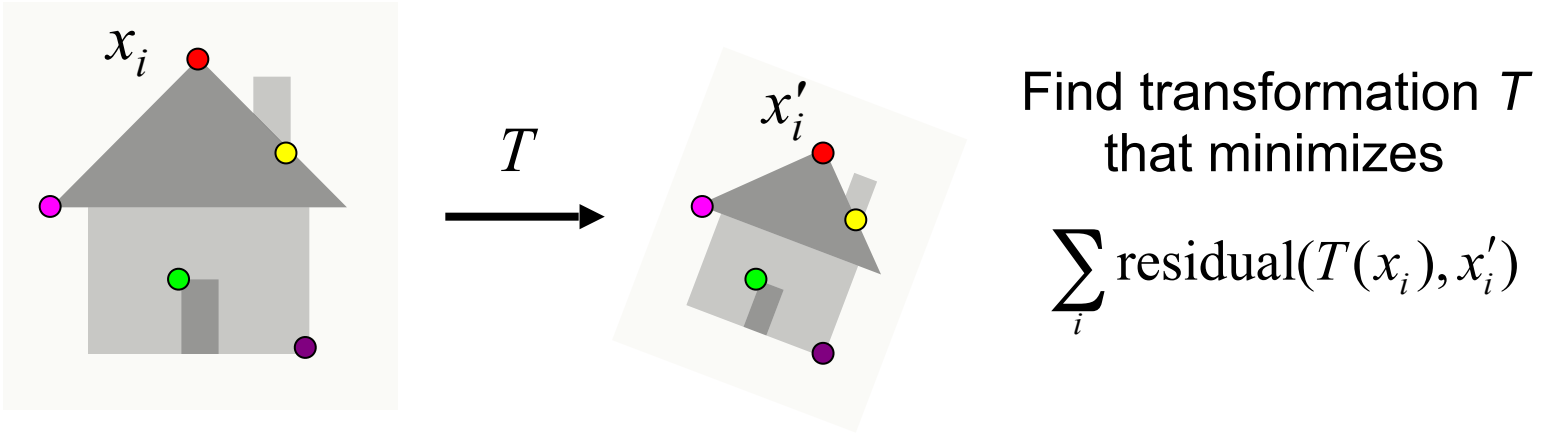
Aspect Graph

- Network of object appearances under different viewpoint.
- The nodes of the graph represent object views that are adjacent to each other on the unit sphere of viewing directions but differ in some significant way.
- the aspect graph arise from transitions in the graph structure relating vertices, edges and faces of the projected object.
- pre-compiling 2-d views into an efficient recognition plan.
- Did not carry forward due to high complexity!!

Alignment

Alignment

- Alignment: fitting a model to a transformation between pairs of features (*matches*) in two images

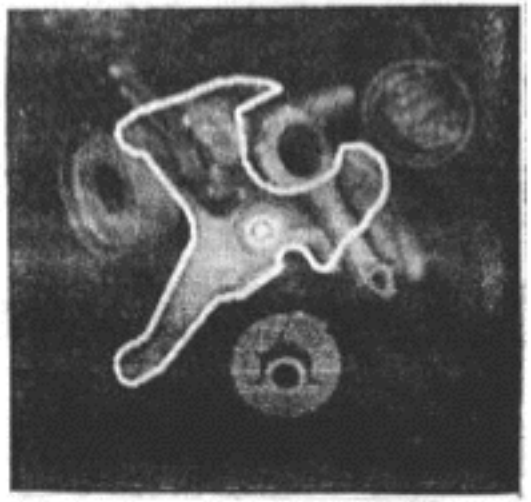


Find transformation T that minimizes

$$\sum_i \text{residual}(T(x_i), x_i')$$

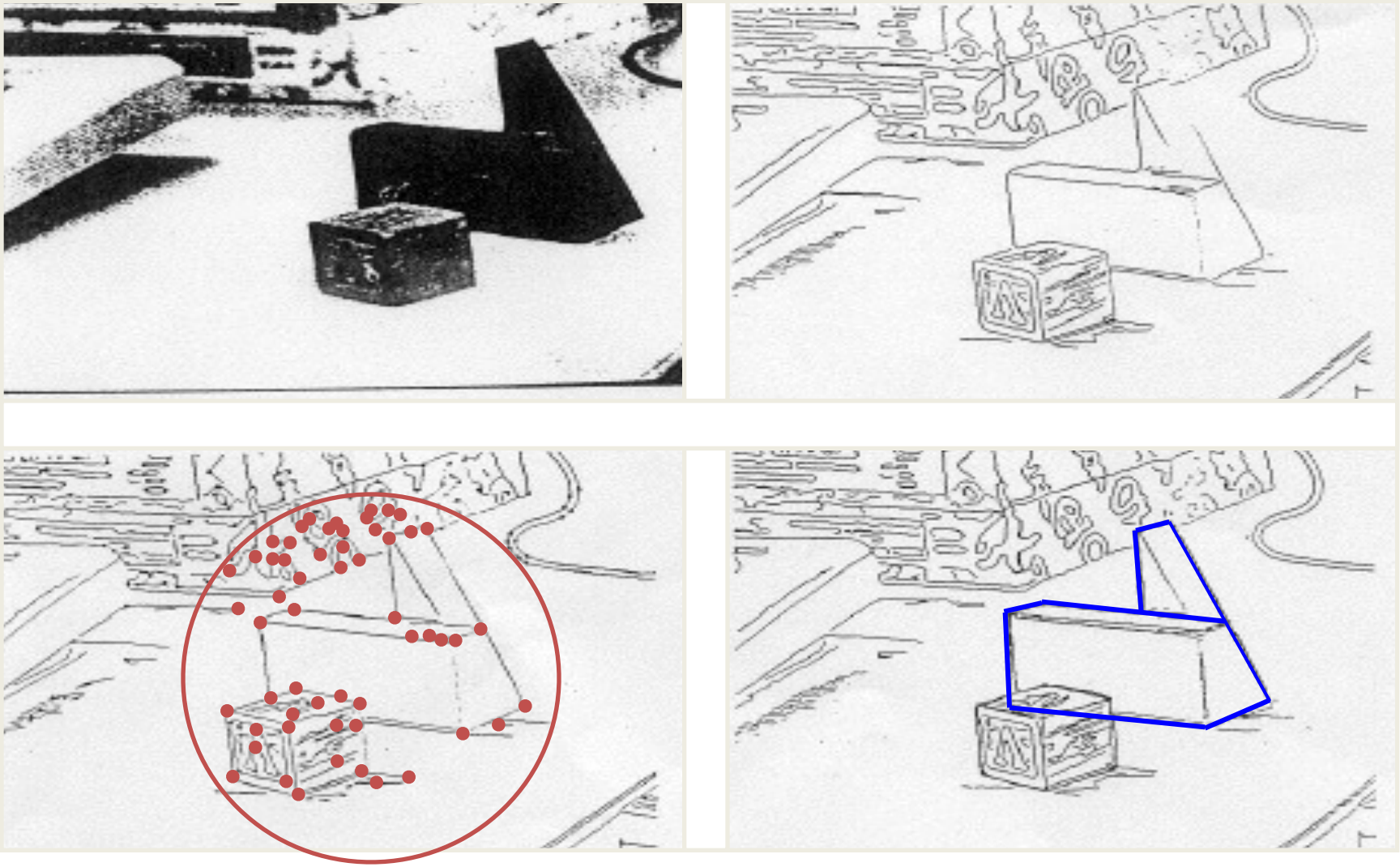
a)

b)



c)

Alignment: Huttenlocher & Ullman (1987)



Geometric Invariants such as cross-ratios and ratio of area ratios.

Projective invariants (Rothwell et al., 1992):

General 3D objects do not admit monocular viewpoint invariants (Burns et al., 1993)

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models

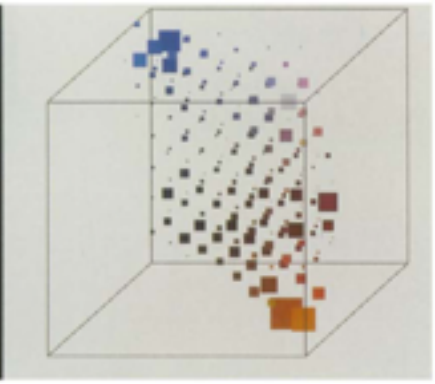
Appearance makes an Appearance

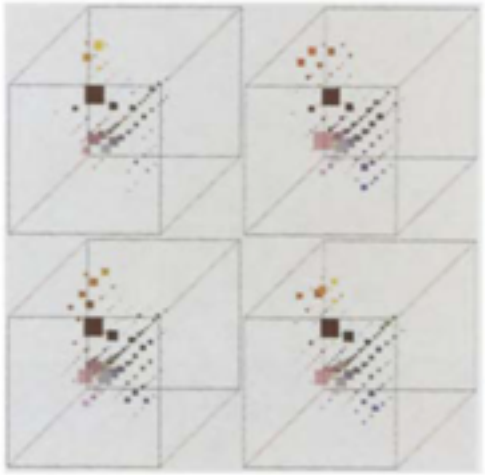
Eigenfaces (Turk & Pentland, 1991)

Experimental	Correct/Unknown Recognition Percentage		
Condition	Lighting	Orientation	Scale
Forced classification	96/0	85/0	64/0
Forced 100% accuracy	100/19	100/39	100/60
Forced 20% unknown rate	100/20	94/20	74/20

Svetlana Lazebnik

Color Histograms





Swain and Ballard, Color Indexing, IJCV 1991.

Limitations of global appearance models

- Requires global registration of patterns
- Not robust to clutter, occlusion, geometric transformations

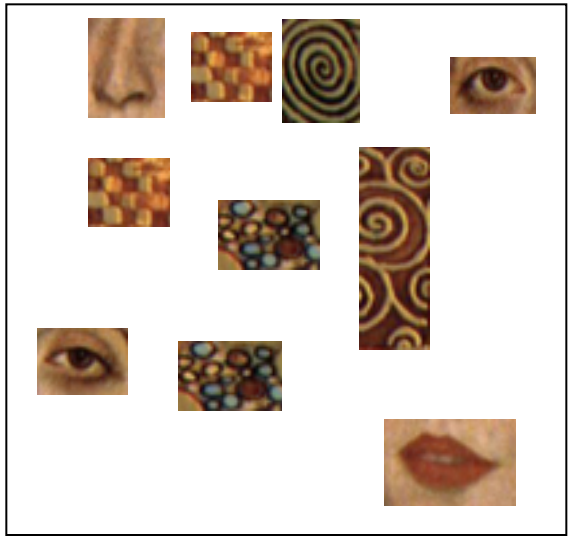
Modeling Local Appearances

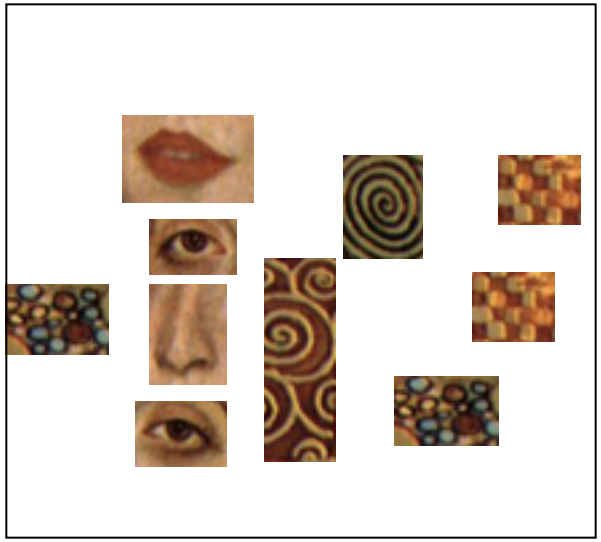
Bag-of-words models

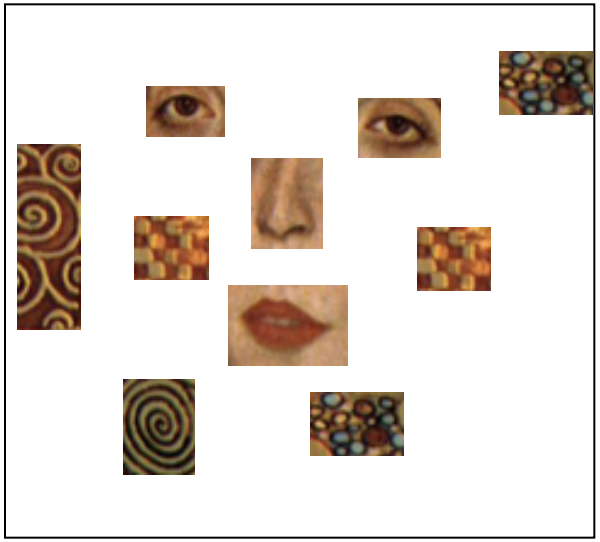
Object → **Bag of ‘words’**

Objects as texture

- All of these are treated as being the same







- No distinction between foreground and background: scene recognition?

Sliding-Window Paradigm

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- 1990s – 2013: sliding window approaches

Sliding window approaches

Sliding window approaches

- Turk and Pentland, 1991
- Belhumeur, Hespanha, & Kriegman, 1997
- Schneiderman & Kanade 2004
- Viola and Jones, 2000

- Schneiderman & Kanade, 2004
- Argawal and Roth, 2002
- Poggio et al. 1993

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Early 2000s: parts-and-shape models

Parts-and-shape models

- Model:
  - Object as a set of parts
  - Relative locations between parts
  - Appearance of part

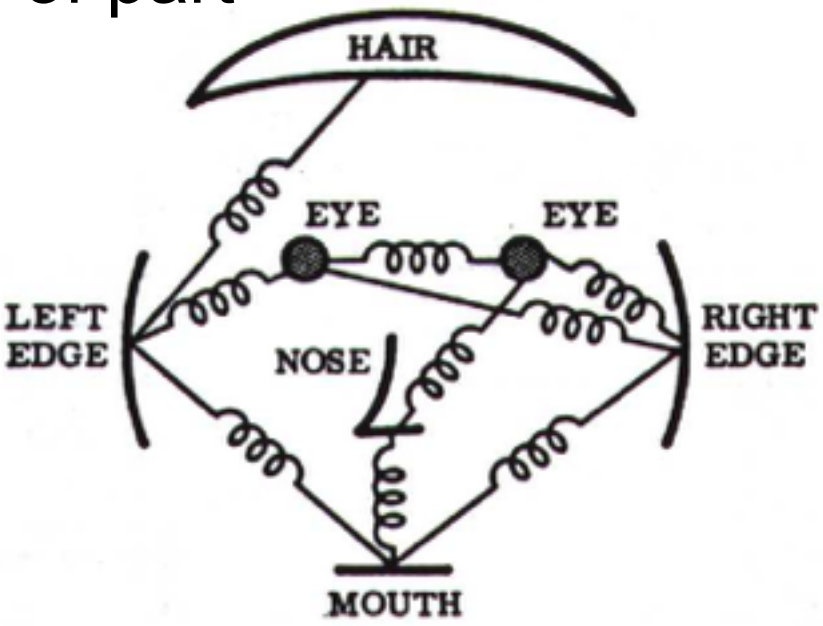
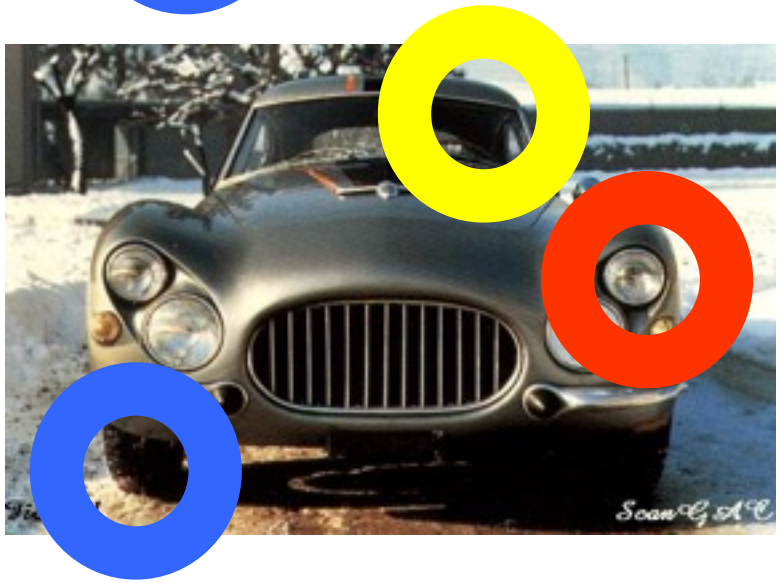
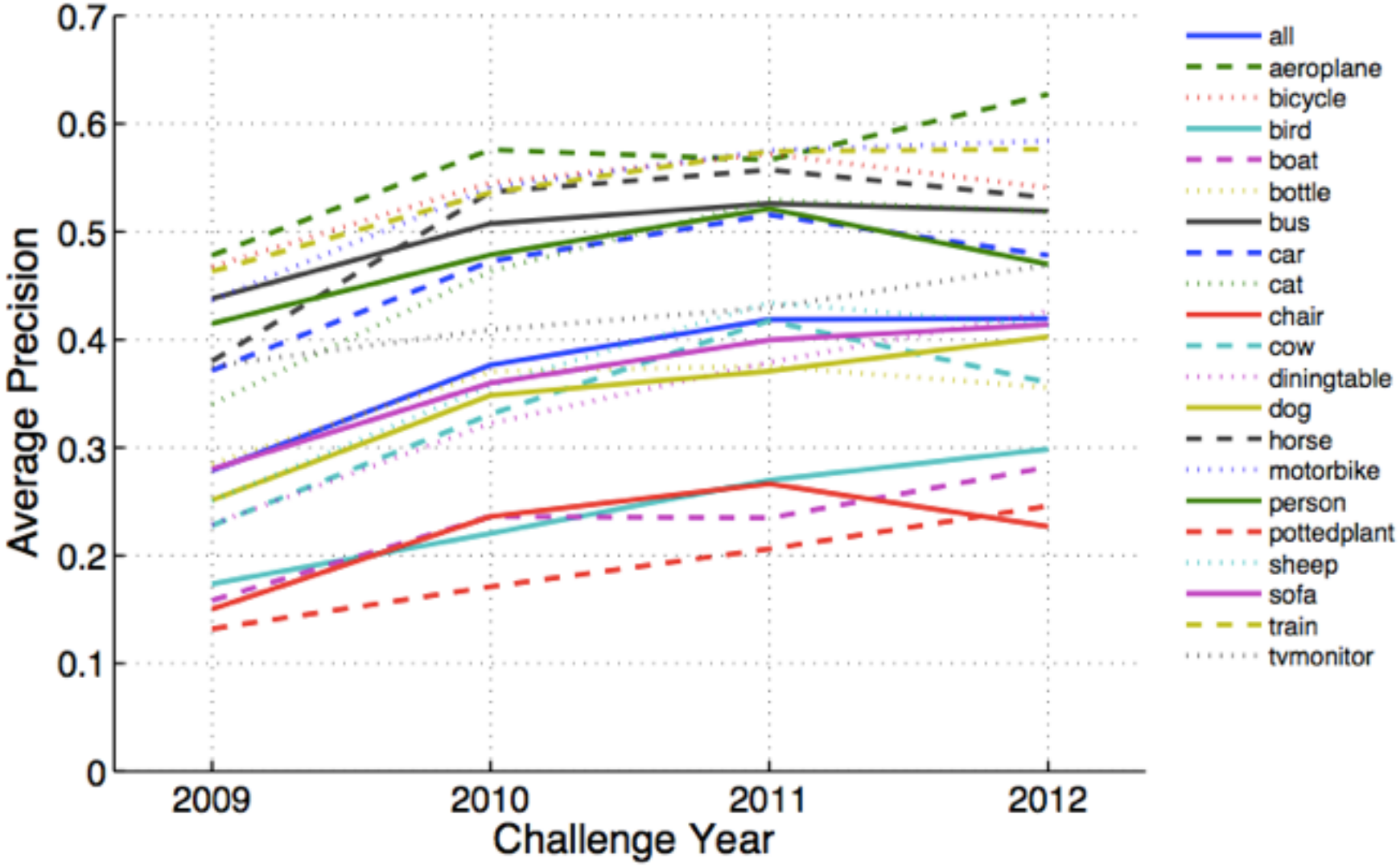


Figure from [Fischler & Elschlager 73]

Constellation models



Weber, Welling & Perona (2000), Fergus, Perona & Zisserman (2003)



VOC Benchmark Dataset

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- 1990s-Present: sliding window approaches
- Early 2000s-Present: parts-and-shape models
- Return of the Convolutional neural networks.

Convolutional Neural Networks

- Designed in 80's inspired by human neuron network in brain.
- Quite successful in the domain of digit recognition – Used in Postal Sorting Center
- But never competitive for generic recognition tasks.

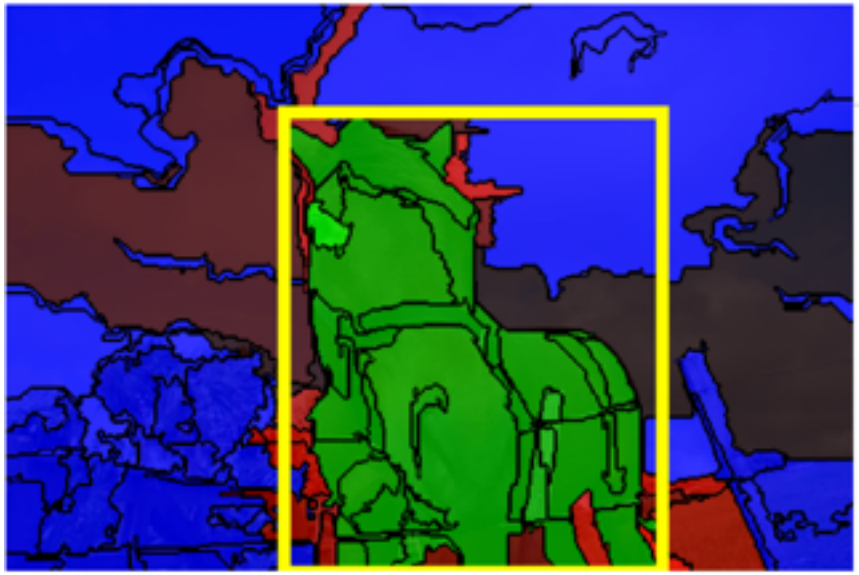
2012

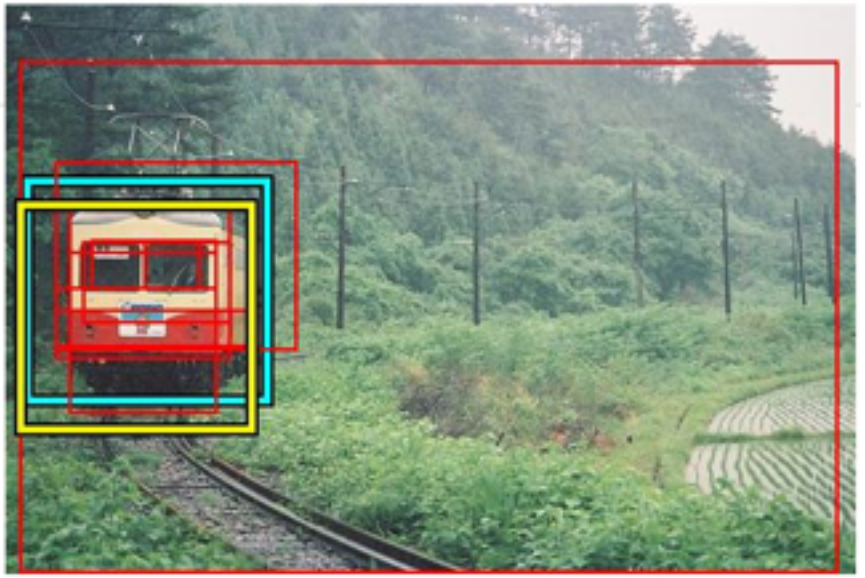
- Success with the advent of bigger (10s of millions of image datasets) and better computing capacity.
- Completely supervised domain.
- Broken state of art in Classification, Detection, 3D Scene Understanding, Human Pose Estimation.
  - Videos still remain a challenge!

History of ideas in recognition

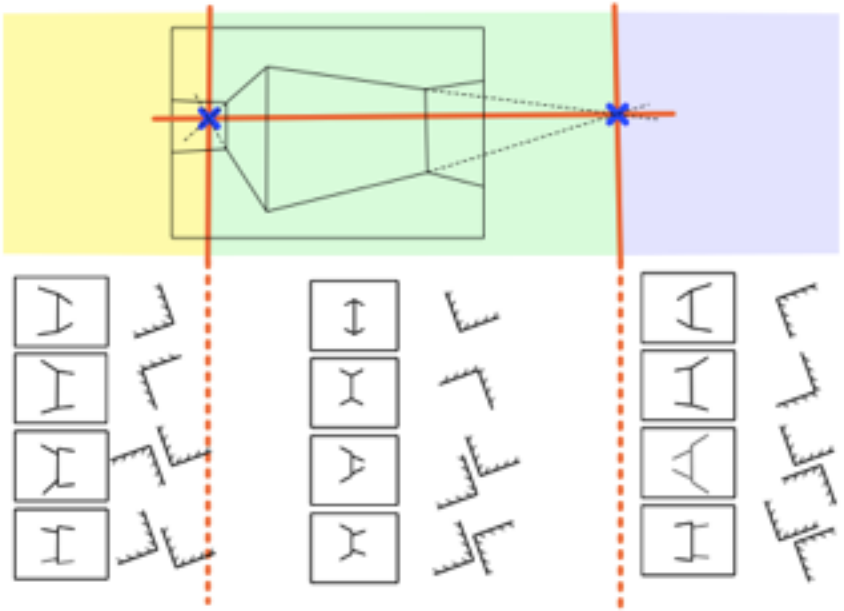
- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- 1990s-Present: sliding window approaches
- Early 2000s-Present: parts-and-shape models
- Return of the Convolutional neural networks.
- 2010s - Object Proposals + CNNs

Object Proposals

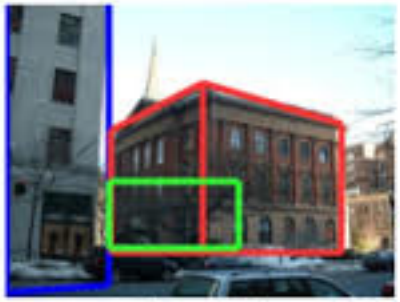




What's next?
Return of Geometric Models?

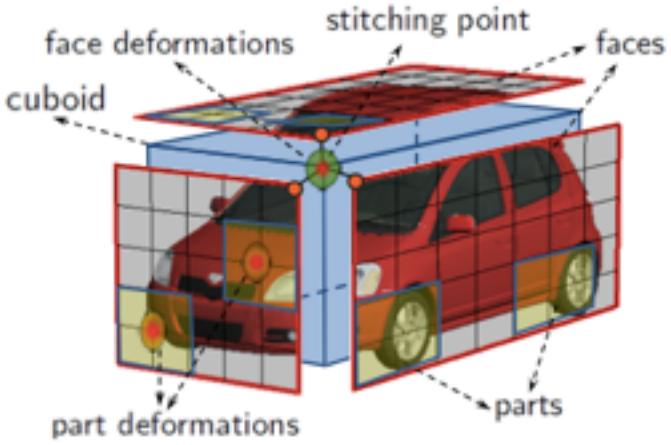


Geometric Layout (CVPR'09)

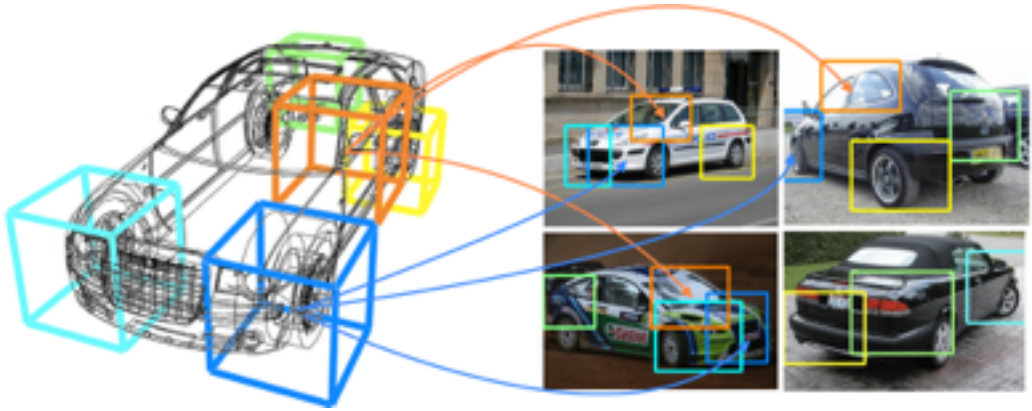


Blocks World Revisited

ECCV 2010



NIPS 2012



CVPR 2012

What “works” today

- Reading license plates, zip codes, checks

Biometrics

Fingerprint scanners on many new laptops, other devices

Face recognition systems now beginning to appear more widely
http://www.sensiblevision.com/

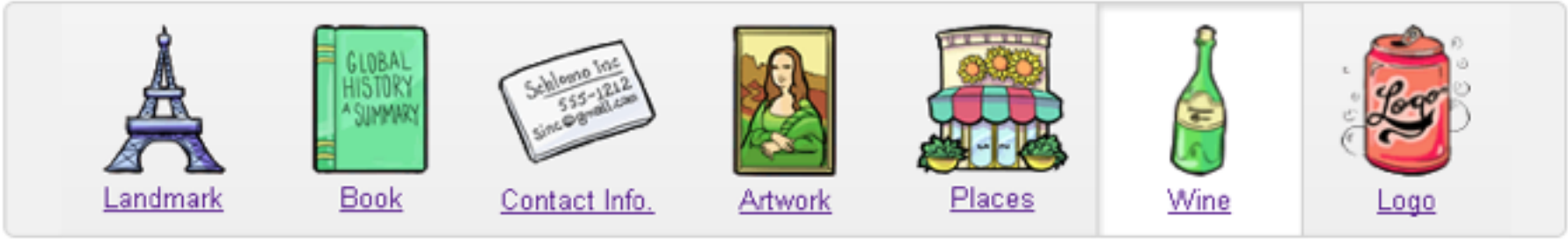
Source: S. Seitz

Mobile visual search: Google Goggles

Google Goggles in Action

Click the icons below to see the different ways Google Goggles can be used.

Google goggles labs

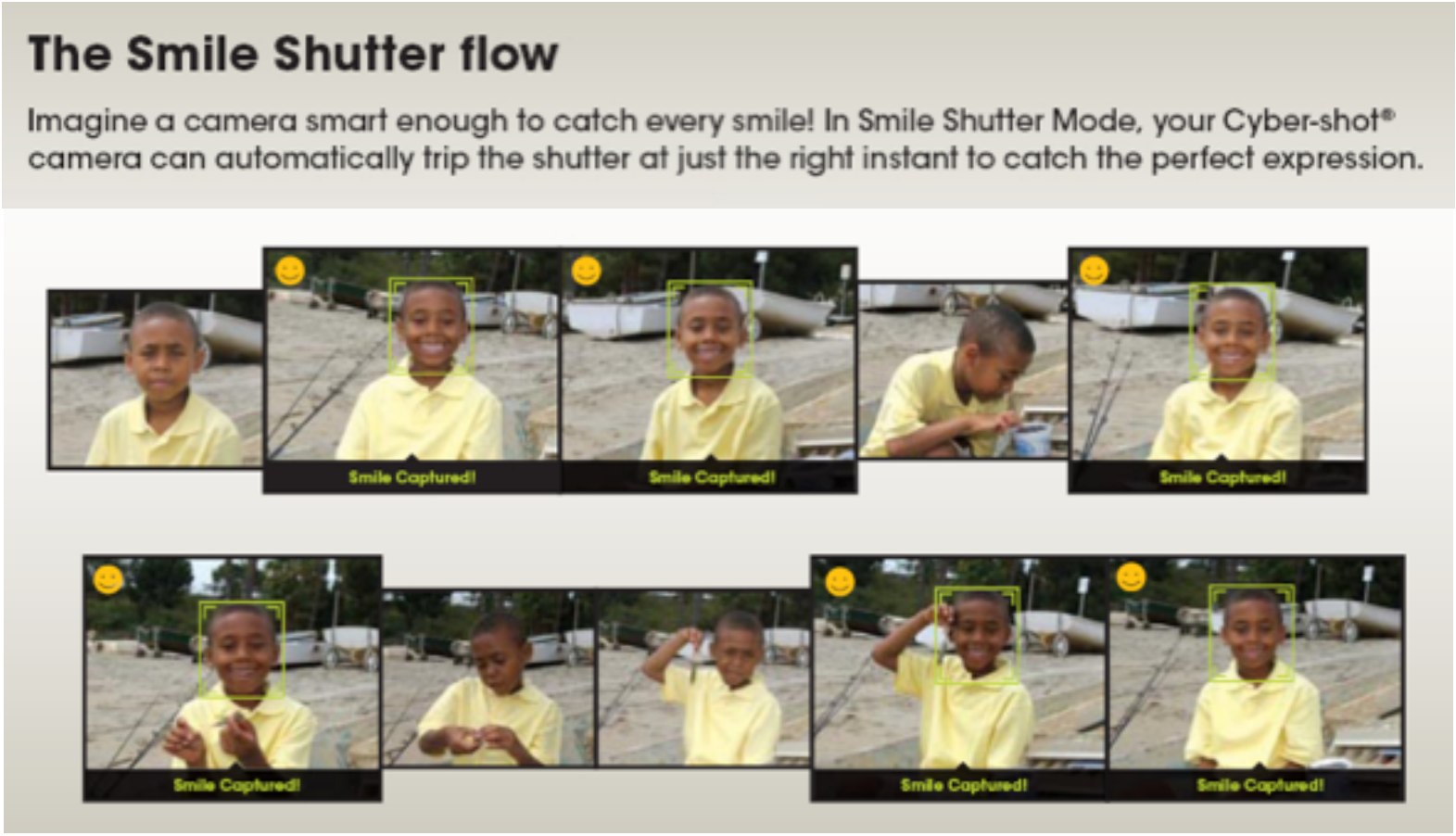


Face detection

- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, …

Source: S. Seitz

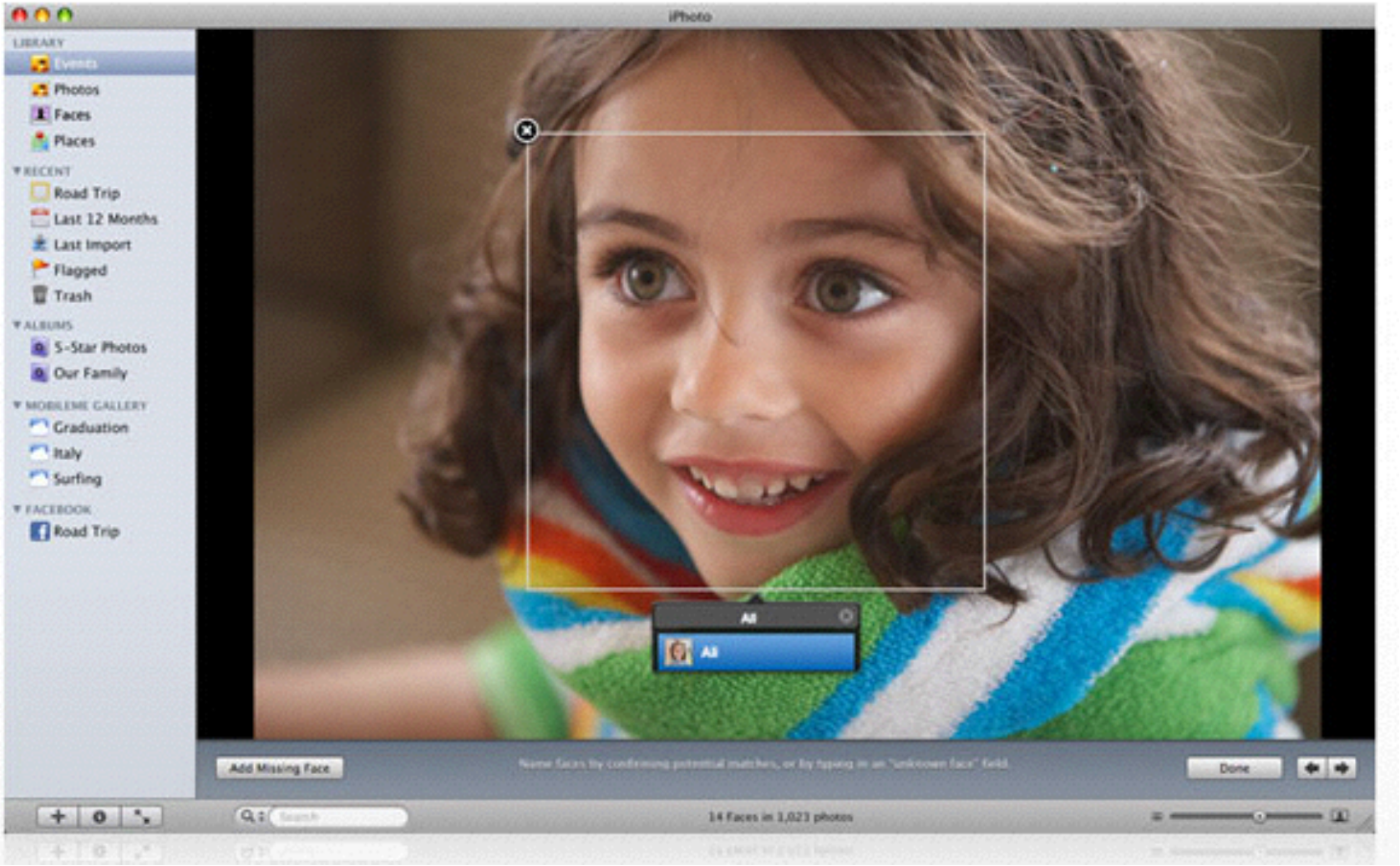
Smile detection



Sony Cyber-shot® T70 Digital Still Camera

Source: S. Seitz

Face recognition: Apple iPhoto, Facebook, Google, etc



Object recognition (in supermarkets)

LaneHawk by EvolutionRobotics

“A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you are assured to get paid for it… “

BBC NEWS

Watch One-Minute World News

News Front Page

Africa
Americas
Asia-Pacific
Europe
Middle East
South Asia
UK
England
Northern Ireland
Scotland
Wales
UK Politics
Education
Magazine
Business
Health
Science &

Last Updated: Wednesday, 31 August 2005, 05:44 GMT 06:44 UK

E-mail this to a friend Printable version

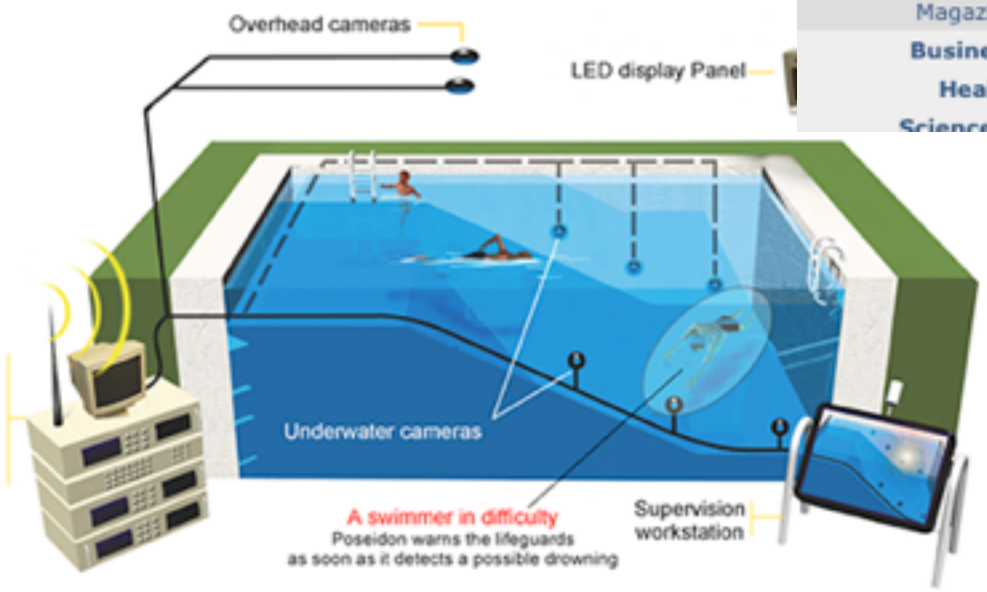
Computer alert for drowning girl

A 10-year-old girl has been saved from drowning by a computer system designed to raise the alarm when swimmers get into difficulties.

VIDEO Watch the rescue

The girl, from Rochdale, was at the deep end of the pool in Bangor, north Wales, when she sank to the bottom.

The £65,000 system, called Poseidon, detected her on the pool floor and sounded the alarm. A lifeguard pulled her out and she recovered in hospital.



Security

Local

Cameras help confirm Scott suicide ruling

Friday, December 04, 2009

TAGS: local, paul meincke

Comment Now Email Print Report a typo

Paul Meincke
More: Bio, News Team

December 4, 2009 (CHICAGO) (WLS) -- Chicago police have closed the case in the death of Chicago School Board President Michael Scott.

Police Supt. Jody Weis says investigators used police cameras in the city to trace Scott's last steps in the hours before his body was found in November.

Scott's death has been ruled a suicide. The medical examiner's office concluded --not long after Scott's body was found -- that he had committed suicide. Police did not dispute the finding but wanted to pursue all the investigative leads they could. They say they have done that and have now reached the same conclusion.

Share this Story

Recommend Be the first of your friends to recommend this.

Tweet 0 +1 Recommend this on Google

News Headlines Video abcNEWS

- 2 suspects arrested in volleyball star's murder 47 min ago
- BP Gas Recall: BP finds, fixes source of bad gas
- Teachers union, board resume negotiating
- Back to School
- 5 injured in South Side shooting 49 min ago
- Pastor: Stacy Peterson said she lied for Drew

Automotive safety

- Mobileye: Vision systems in high-end BMW, GM, Volvo models
  - Pedestrian collision warning
  - Forward collision warning
  - Lane departure warning
  - Headway monitoring and warning

Source: A. Shashua, S. Seitz

Google cars

Oct 9, 2010. "Google Cars Drive Themselves, in Traffic". *The New York Times*. John Markoff
June 24, 2011. "Nevada state law paves the way for driverless cars". *Financial Post*. Christine Dobby
Aug 9, 2011, "Human error blamed after Google's driverless car sparks five-vehicle crash". *The Star* (Toronto)

Vision-based interaction: Xbox Kinect

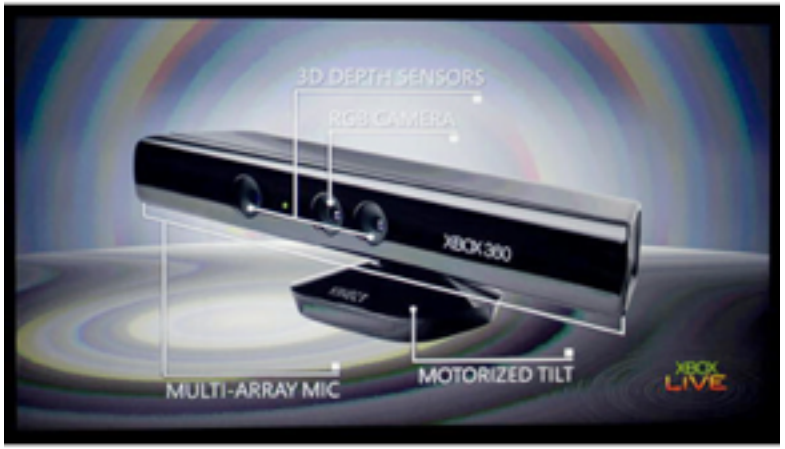


Image Search and Annotation

Jeremy McWilliams' photostream

265 uploads

browse

This photo also belongs to:

april 2008 (Set)

4 items

browse

Tags

stella

cocker spaniel

Hide machine tags (1)

dogs:breed=cocker spaniel

Would you like to comment?

Sign up for a free account, or sign in (if you're already a member).