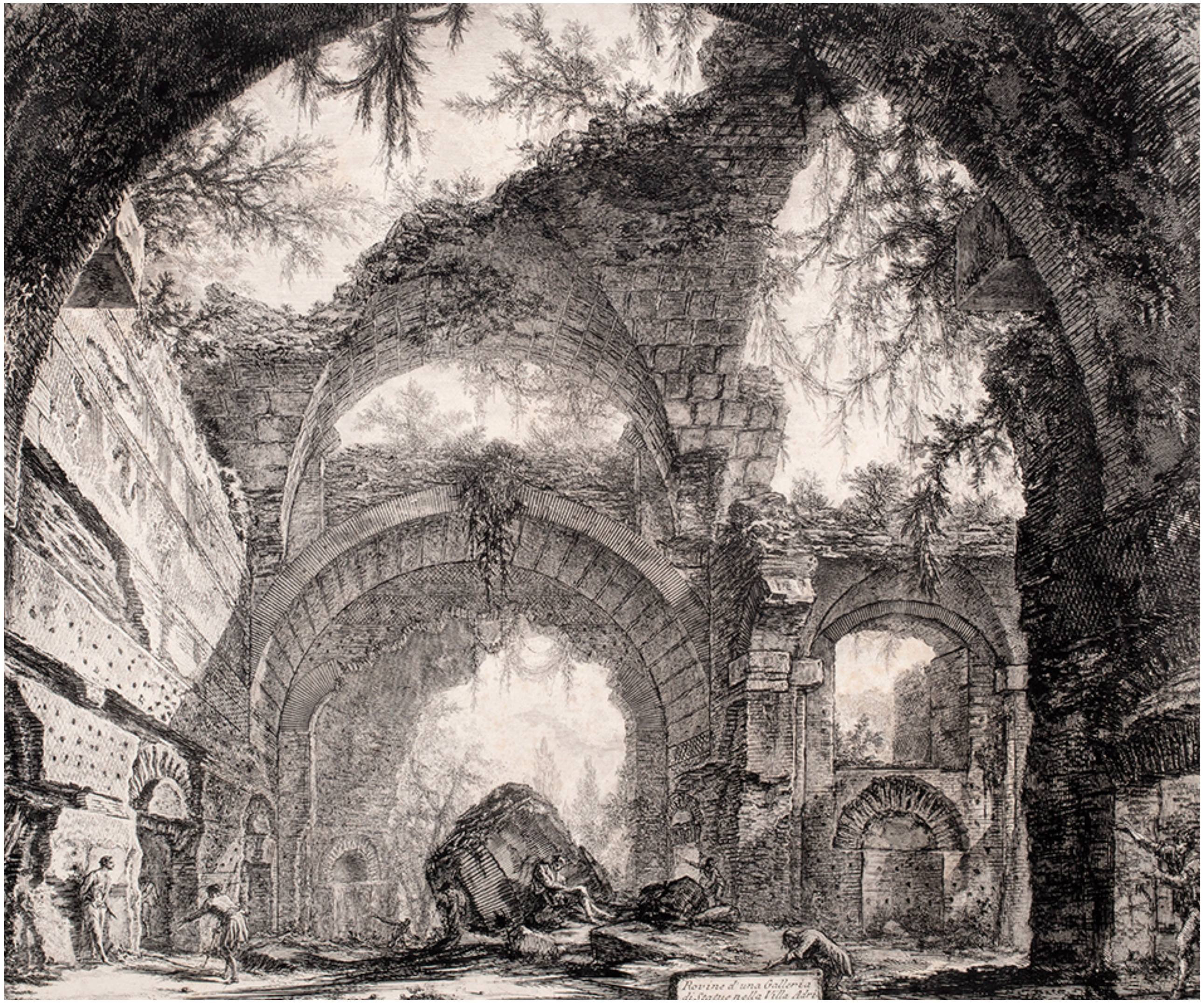
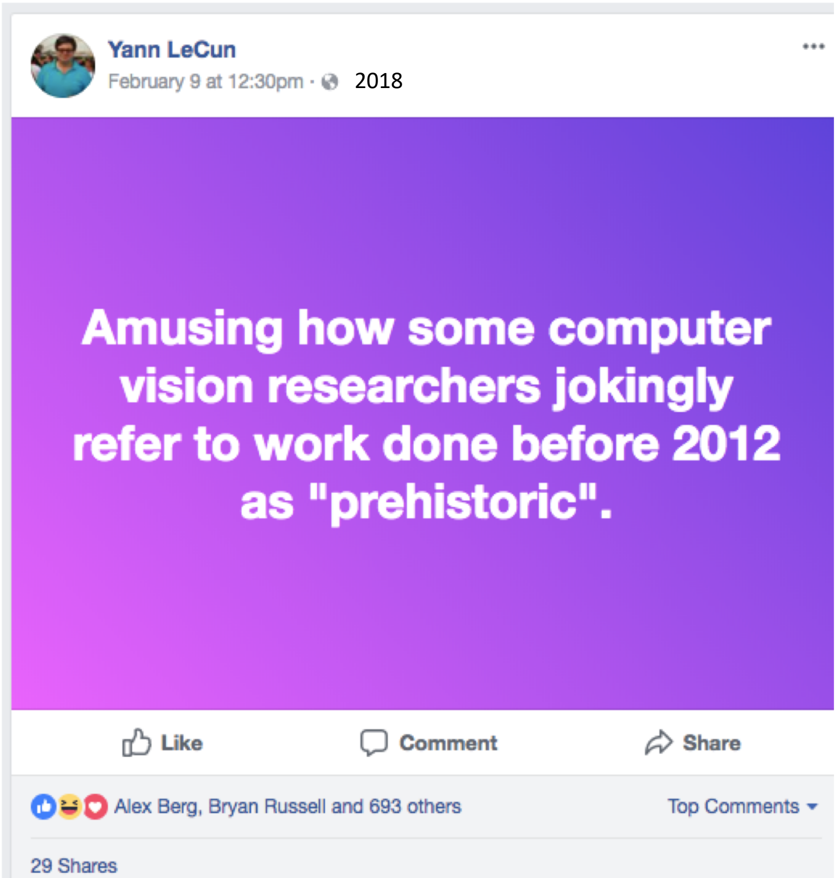


Computer vision:
What will stand
the test of time?

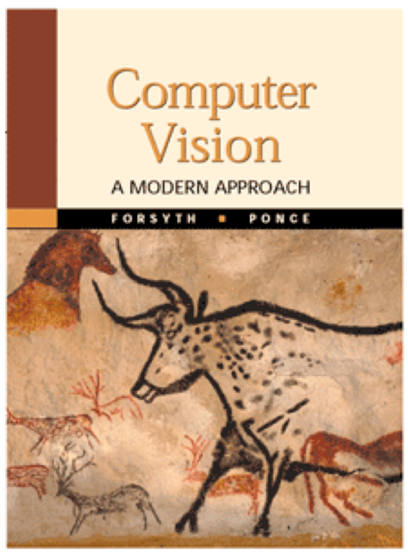
A brief history of computer vision

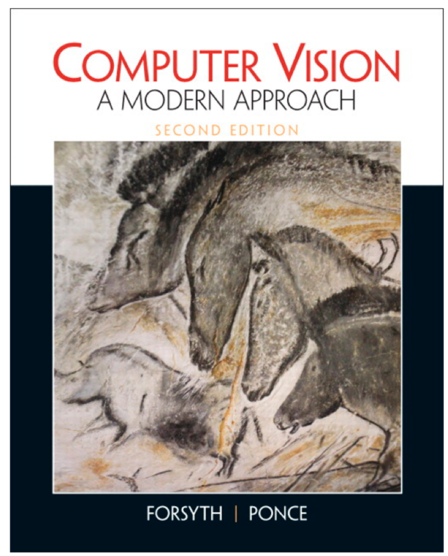
https://www.clevelandart.org/art/1972.119


Yann LeCun
February 9 at 12:30pm · 2018
Amusing how some computer vision researchers jokingly refer to work done before 2012 as "prehistoric".
Like
Comment
Share
Alex Berg, Bryan Russell and 693 others
Top Comments
29 Shares

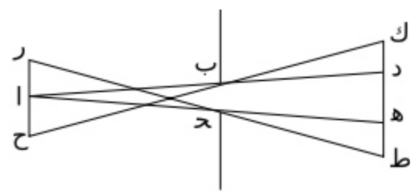

Serge Belongie I call it "antedeepluvian" (before the flood of deep learning papers)
46
Like · Reply · 5d
View 1 more reply
Yann LeCun Yeah, I like this porte manteau.
3
Like · Reply · 4d
Alex Berg "some"... 🙂
Like · Reply · 4d
Yann LeCun Yes, some. And jokingly.
1
Like · Reply · 4d
Aaron Hertzmann I'm just thrilled when I get a paper to review that even acknowledges the existence of related work published prior to 2012.
9
Like · Reply · 4d
Alyosha Efros Around 2005, I was at a vision workshop in MSRI where one of the Gemans said: "physics before Newton is now called 'miscellaneous early efforts'. Vision is at this same stage." Jury still out if we've seen our Newton yet. But looking hopeful.
9
Like · Reply · 5d · Edited

So, what can today’s researchers learn from “prehistoric” computer vision?

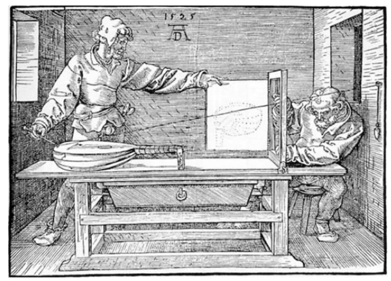




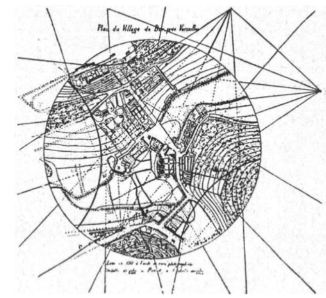
(Actually, “prehistoric” goes even farther back…)



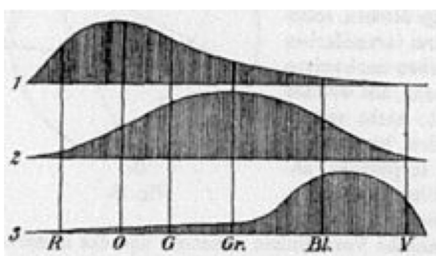
Pinhole projection, optics



Projective geometry



Surveying, photogrammetry



Models of color vision (trichromacy)

Early theories of visual perception: Helmholtz, others

Development of cameras, TV, computers, digital imaging

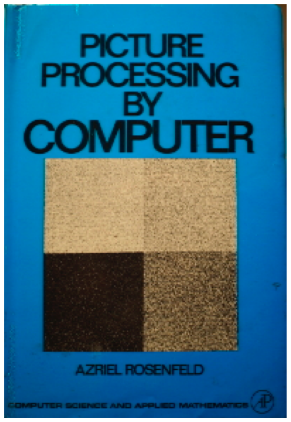
Decade by decade

- **1960s**: Image processing and pattern recognition, blocks world
- **1970s**: Key recovery problems defined: structure from motion, stereo, shape from shading, color constancy. Attempts at knowledge-based recognition
- **1980s**: Fundamental and essential matrix, multi-scale analysis, corner and edge detection, optical flow, geometric recognition as alignment
- **1990s**: Multi-view geometry, statistical and appearance-based models for recognition, first approaches for (class-specific) object detection
- **2000s**: Local features, generic object recognition and detection
- **2010s**: Deep learning, big data

Adapted from J. Malik

1960s (and earlier): A wealth of applications

- Character and digit recognition
  - First OCR conference in 1962
- Microscopy, cytology
- Interpretation of aerial images
  - Even before satellites!
- Particle physics
  - Hough transform for analysis of bubble chamber photos published in 1959
- Face recognition
  - Article about W. Bledsoe
- Fingerprint recognition



Azriel Rosenfeld (1931-2004)

"Father of computer vision"

- Ph.D. in mathematics, Columbia, 1957
- Professor at UMD and ordained rabbi
- Wrote first textbook in the field in 1969
- Oral history, survey (1998)

1960s (and earlier): A wealth of applications

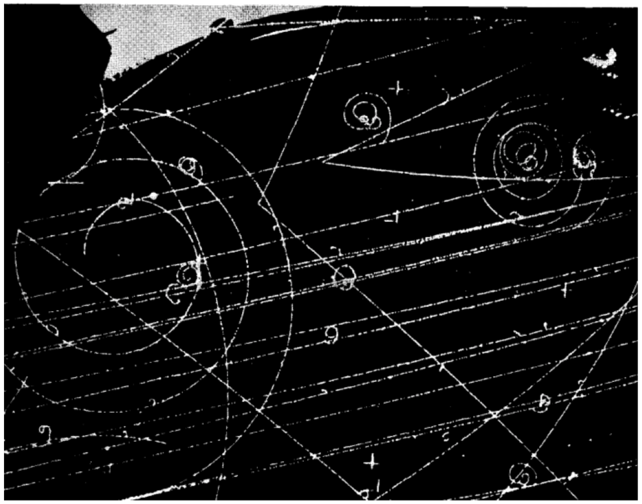
- Character and digit recognition
  - First OCR conference in 1962
- Microscopy, cytology
- Interpretation of aerial images
  - Even before satellites!
- Particle physics
  - Hough transform for analysis of bubble chamber photos published in 1959
- Face recognition
  - Article about W. Bledsoe
- Fingerprint recognition

Aerial mosaic from 1919

Piecing together aerial mosaics at Chanute Field training school in Illinois in the 1920s

1960s (and earlier): A wealth of applications

- Character and digit recognition
  - First OCR conference in 1962
- Microscopy, cytology
- Interpretation of aerial images
  - Even before satellites!
- Particle physics
  - Hough transform for analysis of bubble chamber photos published in 1959
- Face recognition
  - Article about W. Bledsoe
- Fingerprint recognition



Bubble chamber photo

1960s: the MIT-centric narrative

- 1963: Roberts Ph.D. thesis at MIT

MACHINE PERCEPTION OF THREE-DIMENSIONAL SOLIDS

by

LAWRENCE GILMAN ROBERTS

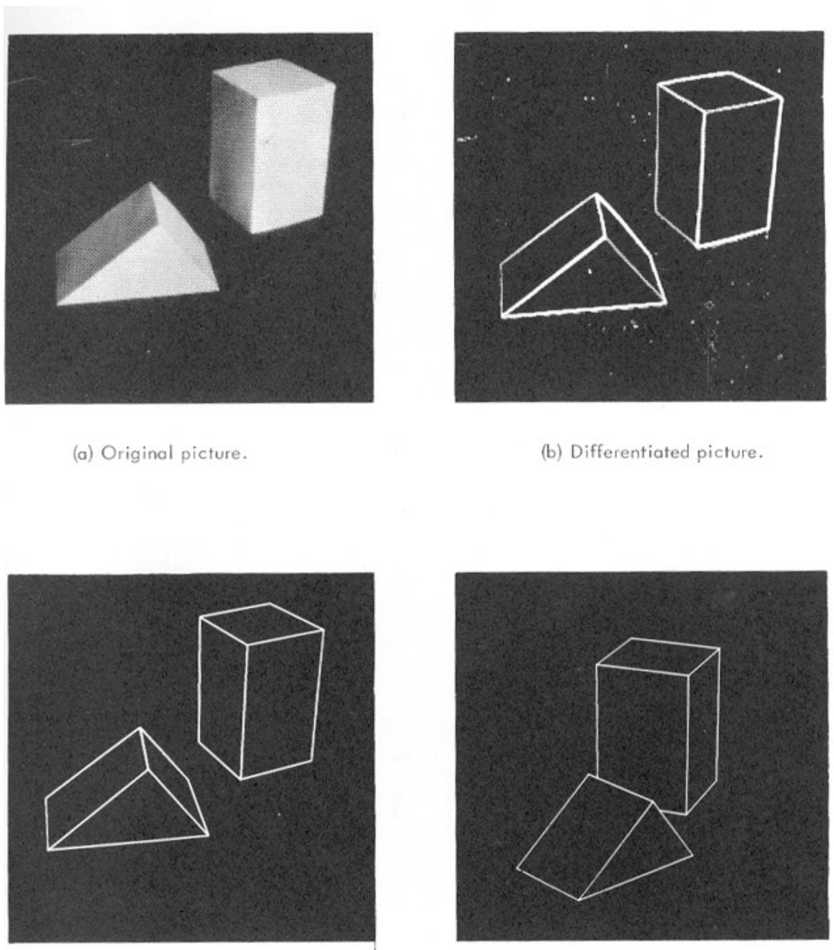
Submitted to the Department of Electrical Engineering on May 10, 1963, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

1937– 2018
Wikipedia bio

CHAPTER I

INTRODUCTION

The problem of machine recognition of pictorial data has long been a challenging goal, but has seldom been attempted with anything more complex than alphabetic characters. Many people have felt that research on character recognition would be a first step, leading the way to a more general pattern recognition system. However, the multitudinous attempts at character recognition, including my own, have not led very far. The reason, I feel, is that the study of abstract, two-dimensional forms leads us away from, not toward, the techniques necessary for the recognition of three-dimensional objects. The perception of solid objects is a process which can be based on the properties of three-dimensional transformations and the laws of nature. By carefully utilizing these properties, a procedure has been developed which can not only identify objects, but also determines their orientation and position in space.



(a) Original picture. (b) Differentiated picture. (c) Line drawing. (d) Rotated view.

From the abstract:

"It is assumed that a photograph is a projection of… **known three-dimensional models**… These assumptions enable a computer to obtain a reasonable, three-dimensional description from the edge information in a photograph by means of a topological, mathematical process."

BIBLIOGRAPHY

1. Somerville, D.M.Y., Analytical Geometry of Three Dimensions, Cambridge University Press, 1959.

2. Roberts, L.G., "Pattern Recognition With An Adaptive Network," IRE International Convention Record, Pt. 2, pp. 66-70, 1960.

3. Selfridge, O.G., and U. Neisser, "Pattern Recognition by Machine," Scientific American, Vol. 203, No. 3, pp 60-68, August, 1960.

4. Hodes, L., Machine Processing of Line Drawings," Lincoln Laboratory, MIT, Lexington, Mass., Group Report 54G0028, March, 1961.

5. Julesz, B., "Toward the Automation of Binocular Depth Perception," Proceedings of the I.F.I.P. Congress, Munich, 1962.

6. Roberts, L.G., "Picture Coding Using Pseudo-Random Noise," IRE Trans. on Information Theory, Vol. IT-8, No. 2, pp 145-154, February, 1962.

7. Gibson, J.J., The Perception of the Visual World, H. Mifflin Company, Boston, Mass., 1950.

8. Ittelson, W.H., "Size As a Cue to Distance," American J. Psychology, Vol. 64, pp 54-67, 1951.

9. Attneave, F. and Arnoult, "The Quantitative Study of Shape and Pattern Perception," Psychological Bull., Vol. 53, p 452, 1956.

10. Langdon, J., "The Perception of 3-D Solids," Quar. J. Exp. Psychology, Vol. 7, 1955.

11. Stevens, S.S., "The Psychophysiology of Vision," in Sensory Communication, W. Rosenblith, Editor, MIT Press, Cambridge, and John Wiley and Sons, New York, N.Y., p. 13, 1961.

12. Sutherland, I.E., Sketchpad, A Man-Machine Graphical Communication System, Ph.D. Thesis, Massachusetts Institute of Technology, Electrical Engineering Department, Cambridge, Mass., February, 1963.

13. Johnson, T., Sketchpad III, 3-D, Graphical, Communication with a Digital Computer, Masters Thesis, Massachusetts Institute of Technology, Mechanical Engineering Department, Cambridge, Mass., June, 1963.

1960s: the MIT-centric narrative

- 1963: Roberts Ph.D. thesis at MIT
  - “Computer vision” explicitly defined in opposition to “pattern recognition” – the key is interpreting images as projections of 3D scenes, not flat 2D “patterns”
- 1966: MIT Summer Vision Project led by Seymour Papert

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".

Subgoal for July

Analysis of scenes consisting of non-overlapping objects from the following set:

balls

bricks with faces of the same or different colors or textures

cylinders.

Each face will be of uniform and distinct color and/or texture. Background will be homogeneous.

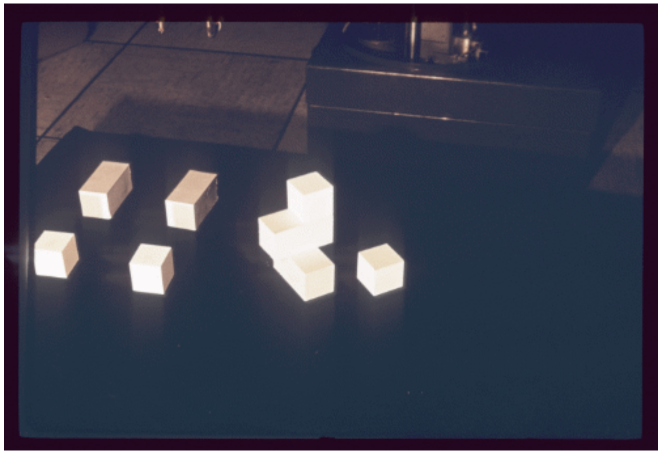
Extensions for August

The first priority will be to handle objects of the same sort but with complex surfaces and backgrounds, e.g. cigarette pack with writing and bands of different color, or a cylindrical battery.

Then extend class of objects to objects like tools, cups, etc.

1960s: the MIT-centric narrative

- 1963: Roberts Ph.D. thesis at MIT
  - “Computer vision” explicitly defined in opposition to “pattern recognition” – the key is interpreting images as projections of 3D scenes, not flat 2D “patterns”
- 1966: MIT Summer Vision Project led by Seymour Papert
  - Underestimated the challenge of computer vision, committed to “blocks world”
- 1970: MIT copy demo (video)
  - An attempt at a “closed loop” robotics system that encompasses sensing, planning, and actuation that affects the environment



Making blocks world cool again?

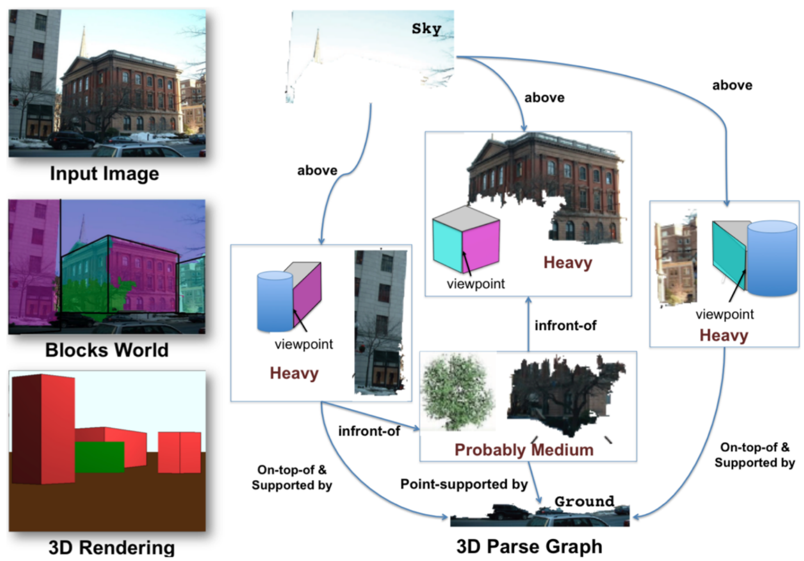
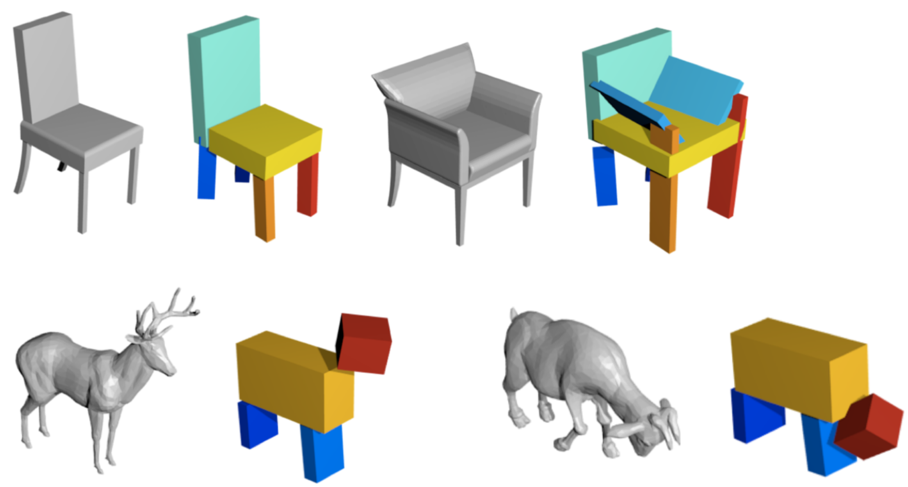


Fig. 1. Example output of our automatic scene understanding system. The 3D parse graph summarizes the inferred object properties (physical boundaries, geometric type, and mechanical properties) and relationships between objects within the scene.

A. Gupta et al. Blocks World Revisited: Image Understanding Using Qualitative Geometry and Mechanics. ECCV 2010



S. Tulsiani et al. Learning Shape Abstractions by Assembling Volumetric Primitives. CVPR 2017

1960s: the MIT-centric narrative

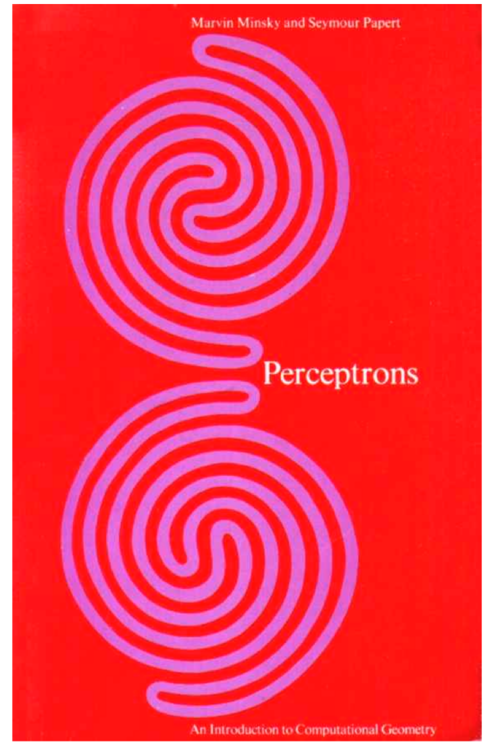
wouldn't be complete without…

- 1969: Minsky and Papert, Perceptrons

0.9 Seductive Aspects of Perceptrons

The purest vision of the perceptron as a pattern-recognizing device is the following:

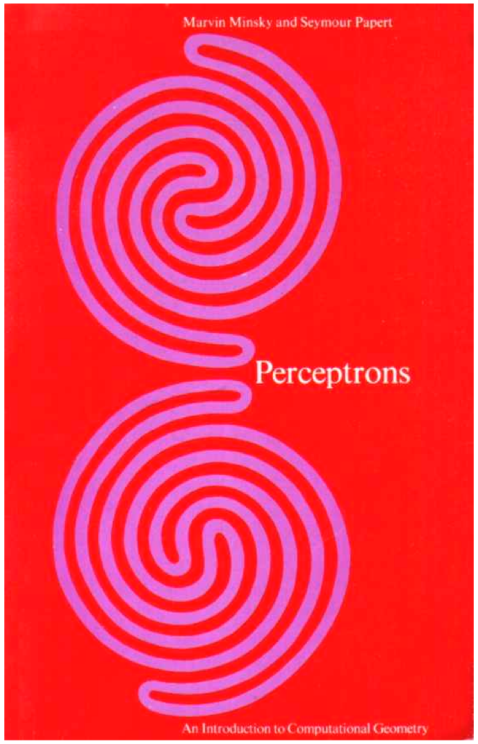
The machine is built with a fixed set of computing elements for the partial functions φ, usually obtained by a random process. To make it recognize a particular pattern (set of input figures) one merely has to set the coefficients α_φ to suitable values. Thus "programming" takes on a pleasingly homogeneous form. Moreover since "programs" are representable as points $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ in an n-dimensional space, they inherit a metric which makes it easy to imagine a kind of automatic programming which people have been tempted to call *learning*: by attaching feedback devices to the parameter controls they propose to "program" the machine by providing it with a sequence of input patterns and an "error signal" which will cause the coefficients to change in the right direction when the machine makes an inappropriate decision. The *perceptron convergence theorems* (see Chapter 11) define conditions under which this procedure is guaranteed to find, eventually, a correct set of values.



1960s: the MIT-centric narrative

wouldn't be complete without…

- 1969: Minsky and Papert, Perceptrons
  - Fascinating reading: M. Olazaran, A Sociological Study of the Official History of the Perceptrons Controversy, *Social Studies of Science*, 1996



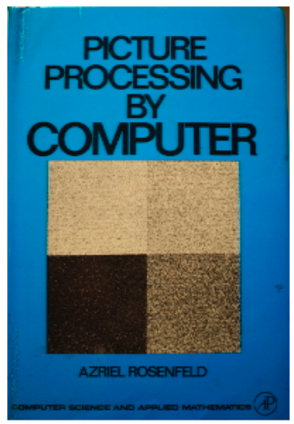
Last time: Computer vision history, decade by decade

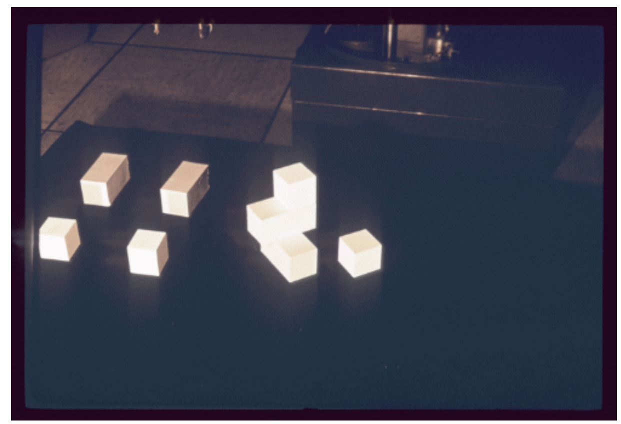
- **1960s**: Image processing and pattern recognition, blocks world
- **1970s**: Key recovery problems defined: structure from motion, stereo, shape from shading, color constancy. Attempts at knowledge-based recognition
- **1980s**: Fundamental and essential matrix, multi-scale analysis, corner and edge detection, optical flow, geometric recognition as alignment
- **1990s**: Multi-view geometry, statistical and appearance-based models for recognition, first approaches for (class-specific) object detection
- **2000s**: Local features, generic object recognition and detection
- **2010s**: Deep learning, big data

Adapted from J. Malik

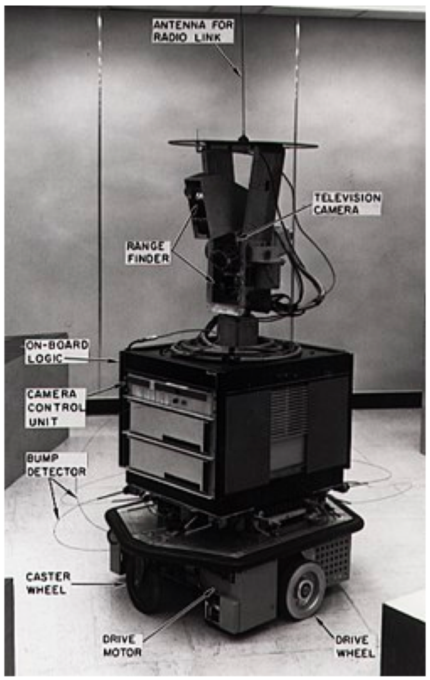
Last time: Computer vision history, decade by decade

- **1960s**: Image processing and pattern recognition, blocks world

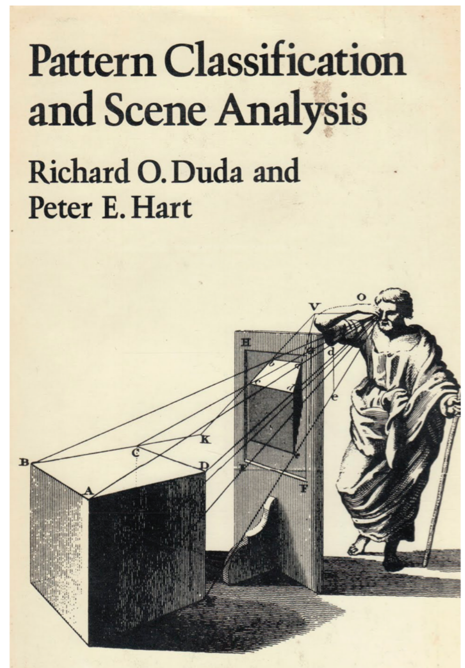




(Not all the action was at MIT, by the way…)



Shakey the Robot
SRI, 1966 - 1972
Video

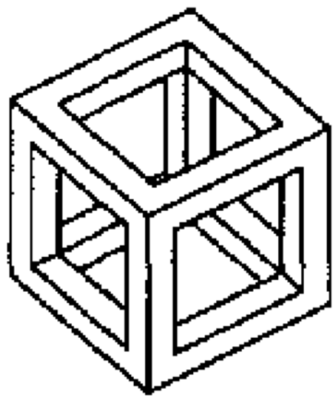


Published in 1972
(table of contents, 2nd edition)

1970s: Recovery

- Shape-from-X

Shading



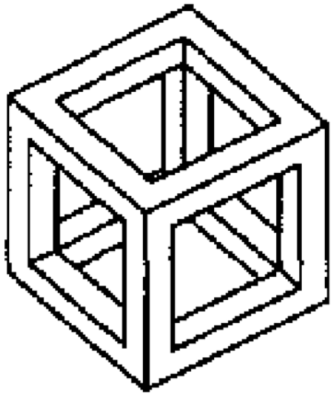
Contour

Texture

1970s: Recovery

- Shape-from-X

Shading



Contour

Texture

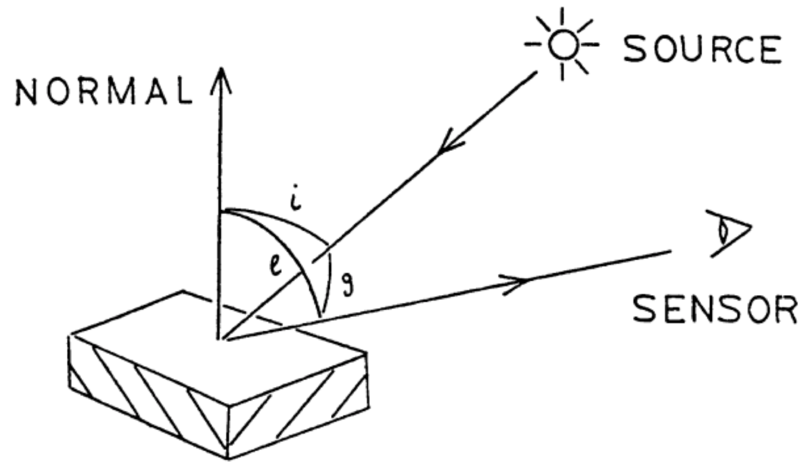
Stereo

1970s: Recovery

- Shape-from-X
  - Shading: Horn (1970)

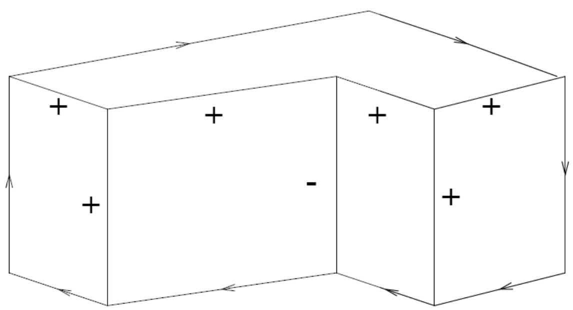
Figure 1: Pictures of a nose with superimposed characteristic solutions and contours. Shape determined from the shading (not intensity contours). See section 4.3 for details.

Horn (1970)

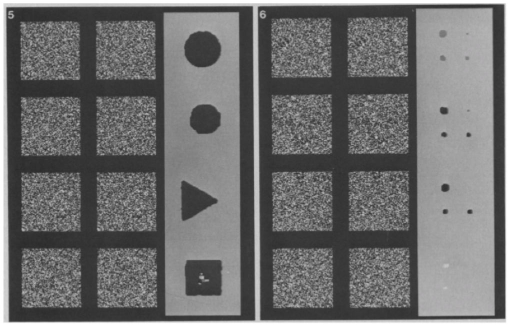


1970s: Recovery

- Shape-from-X
  - Shading: Horn (1970)
  - Contour: Guzman (1971), Waltz (1975), etc.
  - Texture: Bajczy & Lieberman (1976)
  - Stereo: Marr & Poggio (1976)



Waltz (1975)



Marr & Poggio (1976)

1970s: Recovery

- Shape-from-X
  - Shading: Horn (1970)
  - Contour: Guzman (1971), Waltz (1975), etc.
  - Texture: Bajczy & Lieberman (1976)
  - Stereo: Marr & Poggio (1976)
- Color constancy: Land & McCann (1971)
- Intrinsic images: Barrow & Tenenbaum (1978)
- Range images
- Time-varying images
- Optical flow, structure from motion
  - Koenderink & Van Doorn (1975), Ullman (1977)

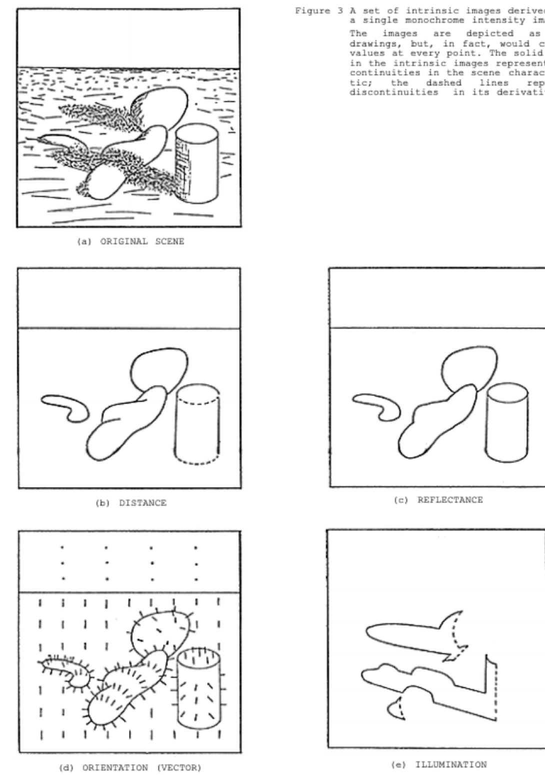


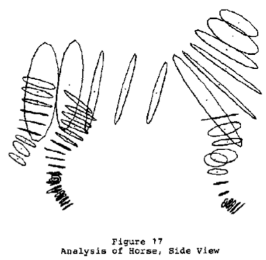
Figure 3 A set of intrinsic images derived from a single monochrome intensity image

The images are depicted as line drawings, but, in fact, would contain values at every point. The solid lines in the intrinsic images represent discontinuities in the scene characteristic; the dashed lines represent discontinuities in its derivative.

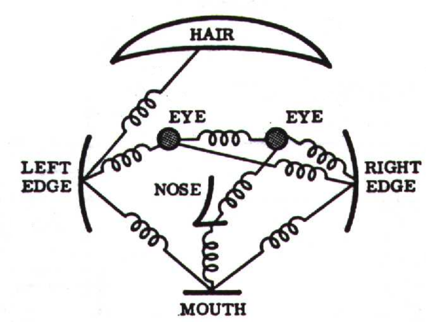
[Barrow & Tenenbaum (1978)](#)

1970s: Representation and recognition

- 3D shape representation
  - Generalized cylinders: Binford et al. (1971, etc.)
- Deformable templates: Fischler & Elschlager (1973)



Binford et al.

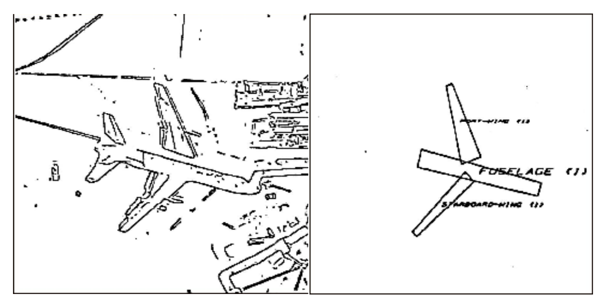


Fischler & Elschlager (1973)

1970s: Representation and recognition

- 3D shape representation
  - Generalized cylinders: Binford et al. (1971, etc.)
- Deformable templates: Fischler & Elschlager (1973)
- Syntactic/procedural recognition systems
  - Faces: Kanade (1973)
  - Scenes: Yakimovsky & Feldman (1973), Hanson & Riseman (1978), Ohta & Kanade (1978)
  - Objects: Brooks (1979)

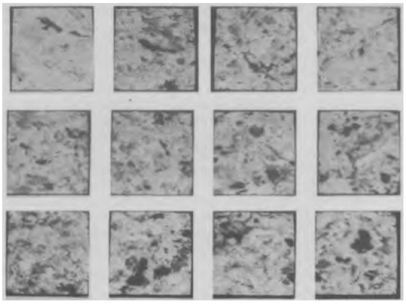
Hanson & Riseman (1978)



Brooks (1981)

1970s: Representation and recognition

- 3D shape representation
  - Generalized cylinders: Binford et al. (1971, etc.)
- Deformable templates: Fischler & Elschlager (1973)
- Syntactic/procedural recognition systems
  - Faces: Kanade (1973)
  - Scenes: Yakimovsky & Feldman (1973), Hanson & Riseman (1978), Ohta & Kanade (1978)
  - Objects: Brooks (1979)
- Relaxation labeling: Rosenfeld et al. (1976)
- Texture recognition: Julesz (1960-1981), Haralick (1979), etc.



Haralick (1979)

Last time: Computer vision history, decade by decade

- **1960s**: Image processing and pattern recognition, blocks world
- **1970s**: Key recovery problems defined: structure from motion, stereo, shape from shading, color constancy. Attempts at knowledge-based recognition
- **1980s**: Fundamental and essential matrix, multi-scale analysis, corner and edge detection, optical flow, geometric recognition as alignment
- **1990s**: Multi-view geometry, statistical and appearance-based models for recognition, first approaches for (class-specific) object detection
- **2000s**: Local features, generic object recognition and detection
- **2010s**: Deep learning, big data

Adapted from J. Malik

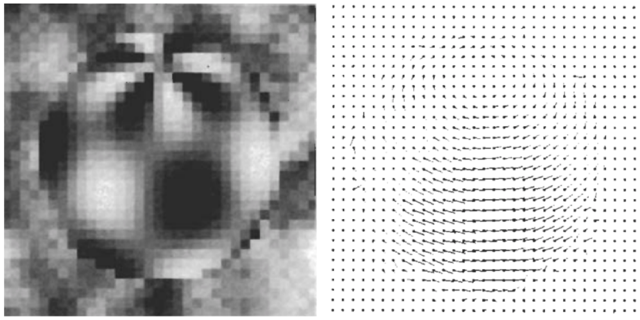
Last time: Computer vision history, decade by decade

- **1960s**: Image processing and pattern recognition, blocks world
- **1970s**: Key recovery problems defined: structure from motion, stereo, shape from shading, color constancy. Attempts at knowledge-based recognition
- **1980s**: Fundamental and essential matrix, multi-scale analysis, corner and edge detection, optical flow, geometric recognition as alignment

Adapted from J. Malik

1980s: 3D vision

- Optical flow and tracking
  - Horn & Schunck (1981), Lucas & Kanade (1981)



Horn & Schunck (1981)

1980s: 3D vision

- Optical flow and tracking
  - Horn & Schunck (1981), Lucas & Kanade (1981)
- Structure from motion
  - RANSAC: Fischler & Bolles (1981)
  - Essential matrix: Longuet-Higgins (1981)

E. A "Real" Location Determination Problem

Cross-correlation was used to locate 25 landmarks in an aerial image taken from approximately 4,000 feet with a 6-inch lens. The image was digitized on a grid of 2,000 by 2,000 pixels, which implies a ground resolution of approximately two feet per pixel. Three gross errors were made by the correlation feature detector. When RANSAC was applied to this problem, it located a consensus set of 17 on the first triple selected and then extended that set to include all 22 good correspondences after the initial least-squares fit. The final standard deviations about the camera parameters were as follows:

X: 0.1 feet	Heading: .01 degrees
Y: 6.4 feet	Pitch: .10 degrees
Z: 2.1 feet	Roll: .12 degrees

Fischler & Bolles (1981)

Nature Vol. 293 10 September 1981

A computer algorithm for reconstructing a scene from two projections

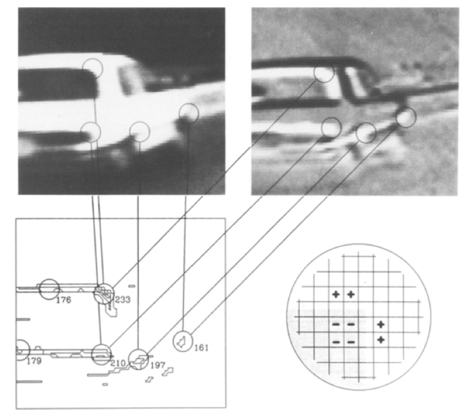
H. C. Longuet-Higgins

Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, UK

Longuet-Higgins (1981)

1980s: 3D vision

- Optical flow and tracking
  - Horn & Schunck (1981), Lucas & Kanade (1981)
- Structure from motion
  - RANSAC: Fischler & Bolles (1981)
  - Essential matrix: Longuet-Higgins (1981)
- Active vision
  - Bajczy (1985, 1988), Dickmanns (1988), Ballard (1989), etc.
    - Interesting read: [The man who invented the self-driving car (in 1986)]



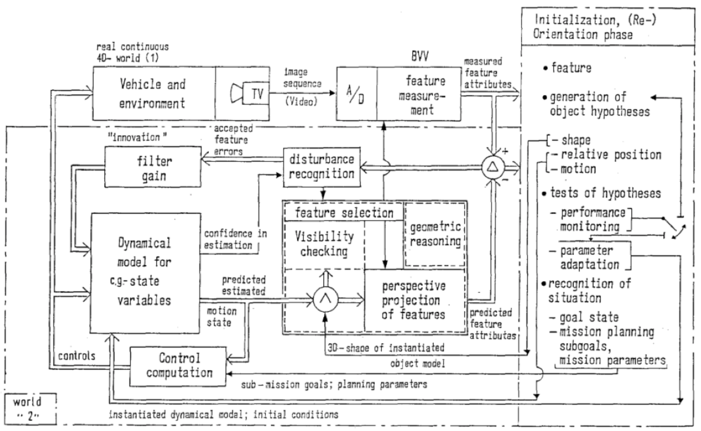
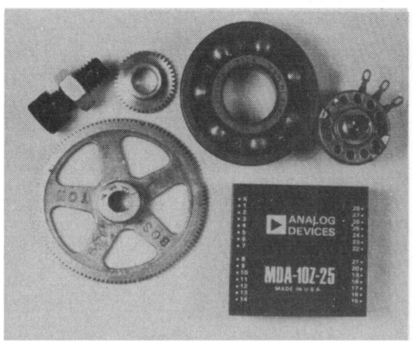


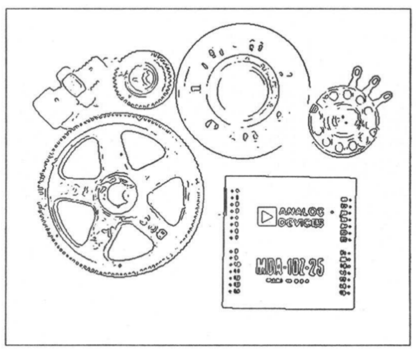
Figure 6. Block diagram showing the information flow in 4-D recursive state estimation for dynamic machine vision.

[Dickmanns & Graefe (1988)]

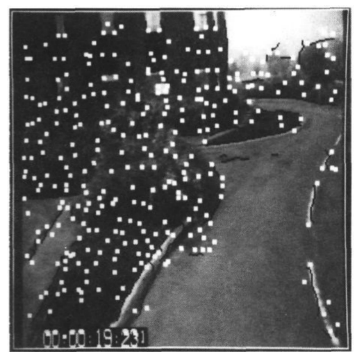
1980s: Image description and inference

- “Definitive” detectors
  - Edges: Canny (1986); corners: Harris & Stephens (1988)





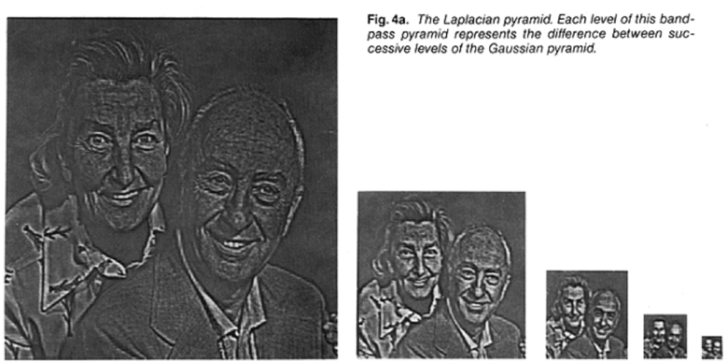
Canny (1986)



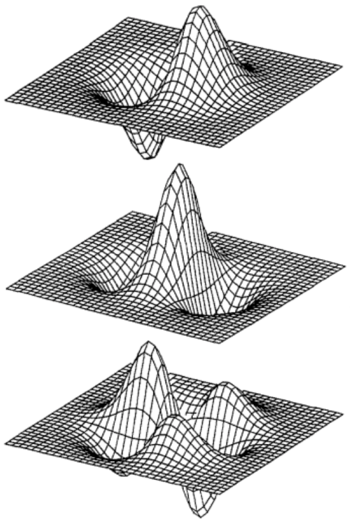
Harris & Stephens (1988)

1980s: Image description and inference

- "Definitive" detectors
  - Edges: Canny (1986); corners: Harris & Stephens (1988)
- Multiscale image representations
  - Witkin (1983), Burt & Adelson (1984), Koenderink (1984, 1987), etc.



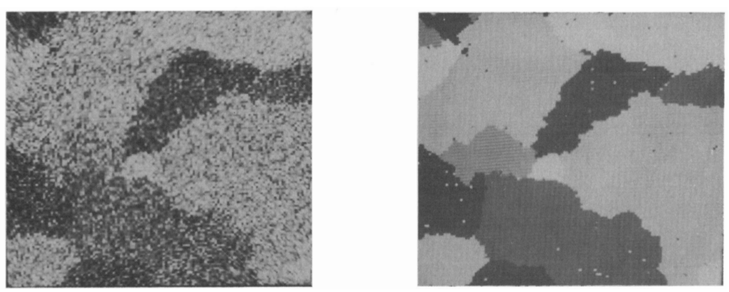
Adelson et al. (1984)



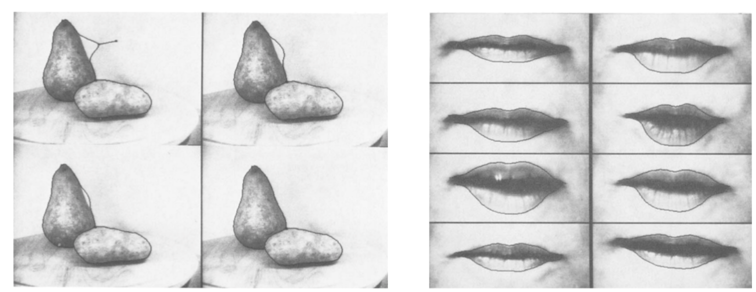
Koenderink (1987)

1980s: Image description and inference

- "Definitive" detectors
  - Edges: Canny (1986); corners: Harris & Stephens (1988)
- Multiscale image representations
  - Witkin (1983), Burt & Adelson (1984), Koenderink (1984, 1987), etc.
- Markov Random Field models: Geman & Geman (1984)
- Segmentation by energy minimization
  - Kass, Witkin & Terzopoulos (1987), Mumford & Shah (1989)



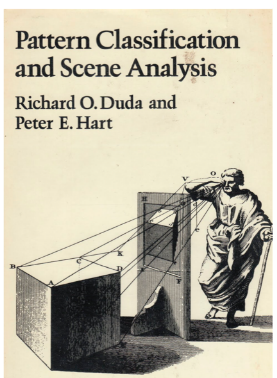
Geman & Geman (1984)

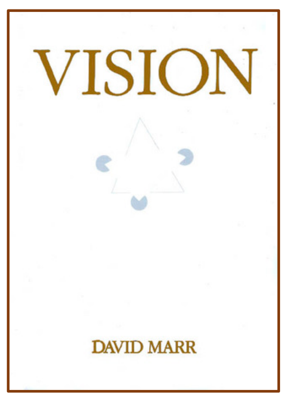


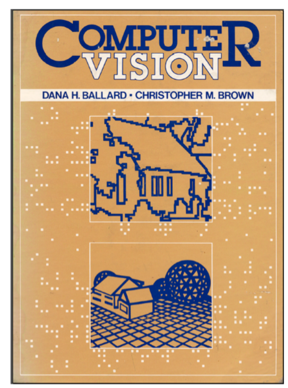
Kass, Witkin & Terzopoulos (1987)

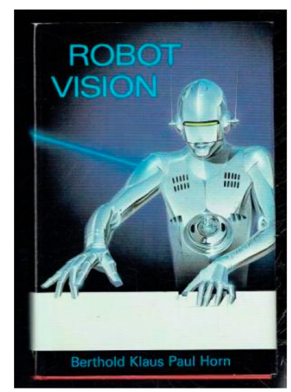
Conferences, journals, books

- Conferences: ICPR (1973), CVPR (1983), ICCV (1987)
- Journals: TPAMI (1979), IJCV (1987)
- Books: Duda & Hart (1972), Marr (1982), Ballard & Brown (1982), Horn (1986)



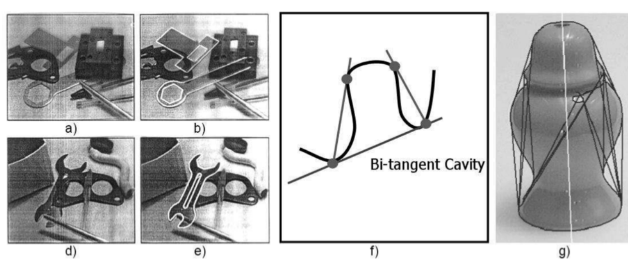


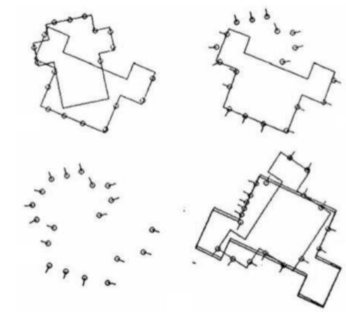




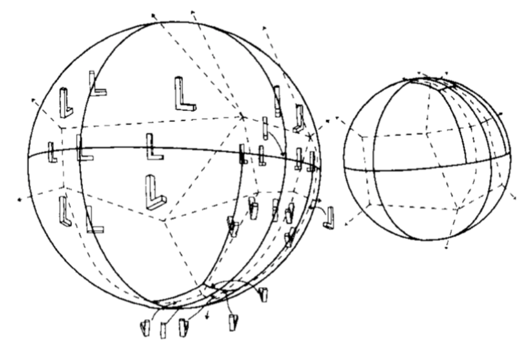
1980s: The dead ends

- Alignment-based recognition
  - Faugeras & Hebert (1983), Grimson & Lozano-Perez (1984), Lowe (1985), Huttenlocher & Ullman (1987), etc.
- Aspect graphs
  - Koenderink & Van Doorn (1979), Plantinga & Dyer (1986), Hebert & Kanade (1985), Ikeuchi & Kanade (1988), Gigus & Malik (1990)
- Invariants: Mundy & Zisserman (1992)





Grimson & Lozano-Perez (1984)



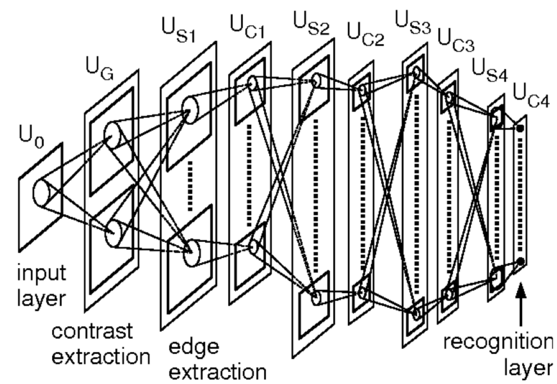
Gigus & Malik (1990)

Fig. 14. A meeting of researchers central to the geometric invariance movement at Schenectady, New York during the month of July, 1992. Top row, left to right: Andrew Zisserman, Charles Rothwell, Luc VanGool, Joseph Mundy, Stephen Maybank and Daniel Huttenlocher. Bottom row, left to right: Thomas Binford, Richard Hartley, David Forsyth and Jon Kleinberg.

Source

1980s: Meanwhile…

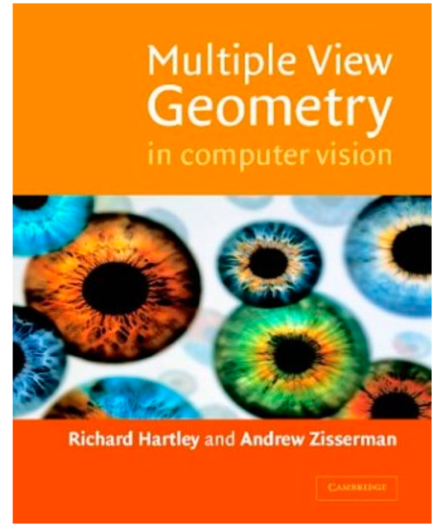
- Neocognitron: Fukushima (1980)
  - Video (short version)
- Back-propagation: Rumelhart, Hinton & Williams (1986)
  - Origins in control theory and optimization: Kelley (1960), Dreyfus (1962), Bryson & Ho (1969), Linnainmaa (1970)
  - Application to neural networks: Werbos (1974)
  - Interesting blog post: Backpropagating through time Or, How come BP hasn't been invented earlier?
- Parallel Distributed Processing: Rumelhart et al. (1987)
- Neural networks for digit recognition: LeCun et al. (1989)

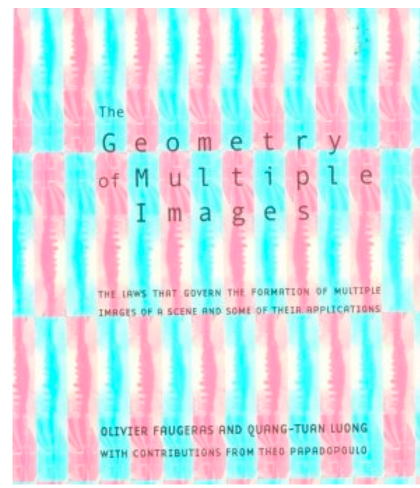


Fukushima (1980)

1990s: Geometry reigns

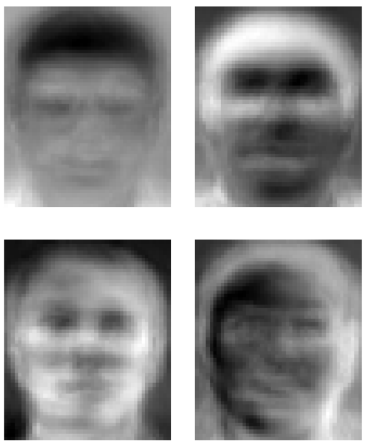
- Fundamental matrix: Faugeras (1992)
- Normalized 8-point algorithm: Hartley (1997)
- RANSAC for robust fundamental matrix estimation: Torr & Murray (1997)
- Bundle adjustment: Triggs et al. (1999)
- Hartley & Zisserman book (2000)
- Projective structure from motion: Faugeras and Luong (2001)



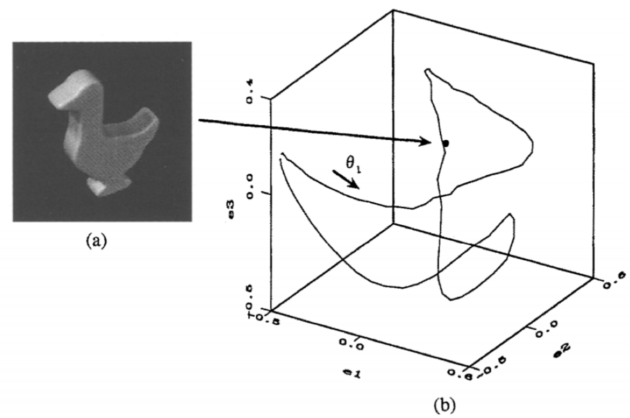


1990s: Data enters the scene

- Appearance-based models: Turk & Pentland (1991), Murase & Nayar (1995)



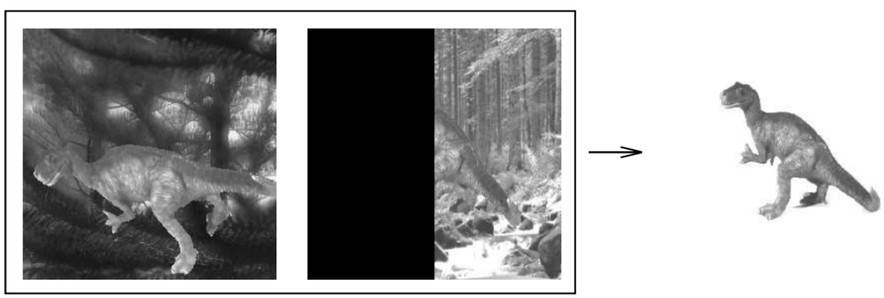
Turk & Pentland (1991)



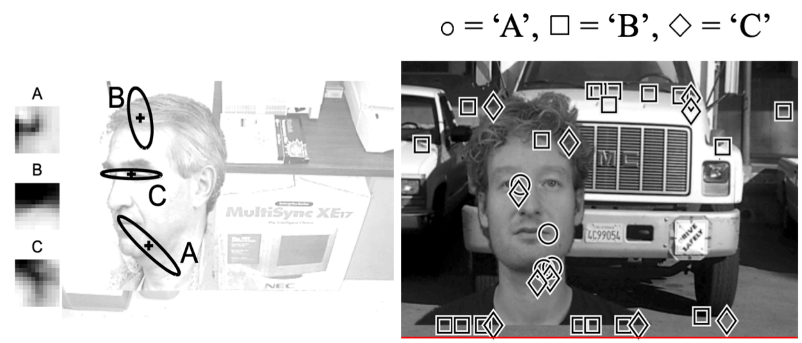
Murase & Nayar (1995)

1990s: Data enters the scene

- Appearance-based models: Turk & Pentland (1991), Murase & Nayar (1995)
- Keypoint-based image indexing
  - Schmid & Mohr (1996), Lowe (1999)
- Constellation models for object categories
  - Burl, Weber & Perona (1998), Weber, Welling & Perona (2000)



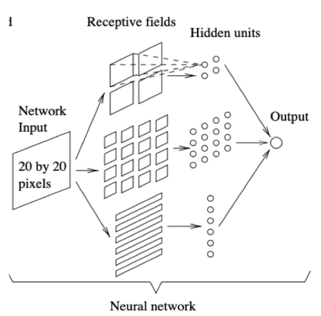
Schmid & Mohr (1996)



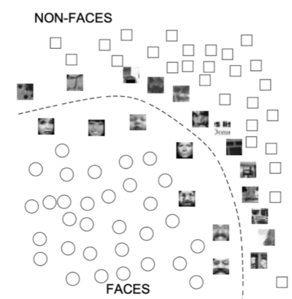
Weber, Welling & Perona (2000)

1990s: Data enters the scene

- Appearance-based models: Turk & Pentland (1991), Murase & Nayar (1995)
- Keypoint-based image indexing
  - Schmid & Mohr (1996), Lowe (1999)
- Constellation models for object categories
  - Burl, Weber & Perona (1998), Weber, Welling & Perona (2000)
- First sustained use of classifiers and negative data
  - Face detectors: Rowley, Baluja & Kanade (1996), Osuna, Freund & Girosi (1997), Schneiderman & Kanade (1998), Viola & Jones (2001)
  - Convolutional nets: LeCun et al. (1998)



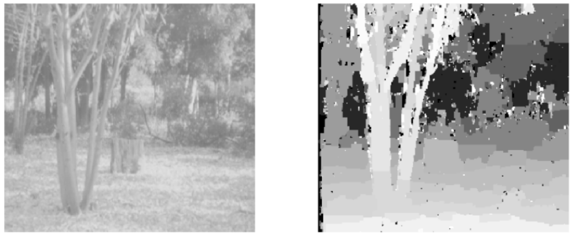
Rowley, Baluja, Kanade (1998)



Osuna, Freund, Girosi (1997)

1990s: Data enters the scene

- Appearance-based models: Turk & Pentland (1991), Murase & Nayar (1995)
- Keypoint-based image indexing
  - Schmid & Mohr (1996), Lowe (1999)
- Constellation models for object categories
  - Burl, Weber & Perona (1998), Weber, Welling & Perona (2000)
- First sustained use of classifiers and negative data
  - Face detectors: Rowley, Baluja & Kanade (1996), Osuna, Freund & Girosi (1997), Schneiderman & Kanade (1998), Viola & Jones (2001)
  - Convolutional nets: LeCun et al. (1998)
- Graph cut image inference
  - Boykov, Veksler & Zabih (1998)
- Segmentation
  - Normalized cuts: Shi & Malik (2000)
  - Berkeley segmentation dataset: Martin et al. (2001)
- Optical flow, tracking
  - Adelson & Wang (1993), Black & Anandan (1993), Isard & Blake (1998)



Boykov, Veksler & Zabih (1998)

Last two weeks: Computer vision history, decade by decade

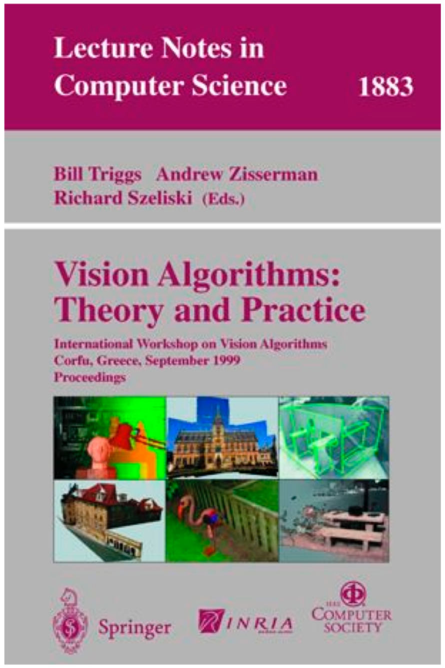
- **1960s**: Image processing and pattern recognition, blocks world
- **1970s**: Key recovery problems defined: structure from motion, stereo, shape from shading, color constancy. Attempts at knowledge-based recognition
- **1980s**: Fundamental and essential matrix, multi-scale analysis, corner and edge detection, optical flow, geometric recognition as alignment
- **1990s**: Multi-view geometry, statistical and appearance-based models for recognition, first approaches for (class-specific) object detection
- **2000s**: Local features, generic object recognition and detection
- **2010s**: Deep learning, big data

Last two weeks: Computer vision history, decade by decade

- **1960s**: Image processing and pattern recognition, blocks world
- **1970s**: Key recovery problems defined: structure from motion, stereo, shape from shading, color constancy. Attempts at knowledge-based recognition
- **1980s**: Fundamental and essential matrix, multi-scale analysis, corner and edge detection, optical flow, geometric recognition as alignment
- **1990s**: Multi-view geometry, statistical and appearance-based models for recognition, first approaches for (class-specific) object detection

Late 1990's debates

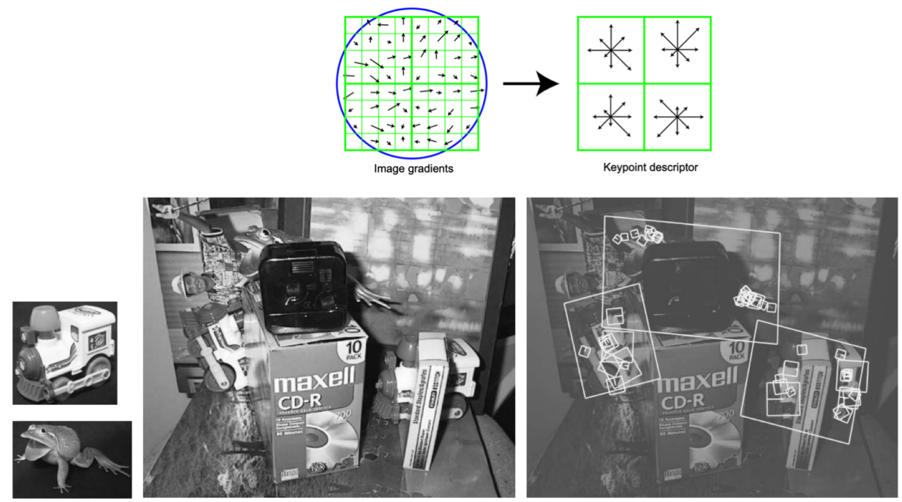
- See the last chapter of Vision Algorithms: Theory and Practice (1999)



2000s: Keypoints and reconstruction

- Keypoints craze
  - Kadir & Brady (2001), Mikolajczyk & Schmid (2002), Matas et al. (2004), Lowe (2004), Bay et al. (2006), etc.

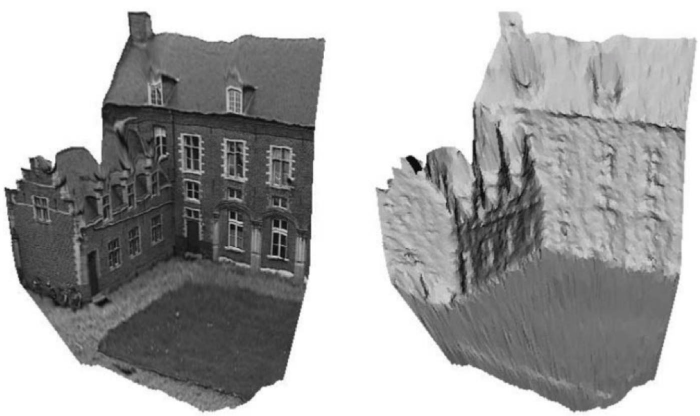
Mikolajczyk & Schmid (2002)



Lowe (2004)

2000s: Keypoints and reconstruction

- Keypoints craze
  - Kadir & Brady (2001), Mikolajczyk & Schmid (2002), Matas et al. (2004), Lowe (2004), Bay et al. (2006), etc.
- 3D reconstruction “in the wild”
  - SFM in the wild
  - Multi-view stereo, stereo on GPU’s

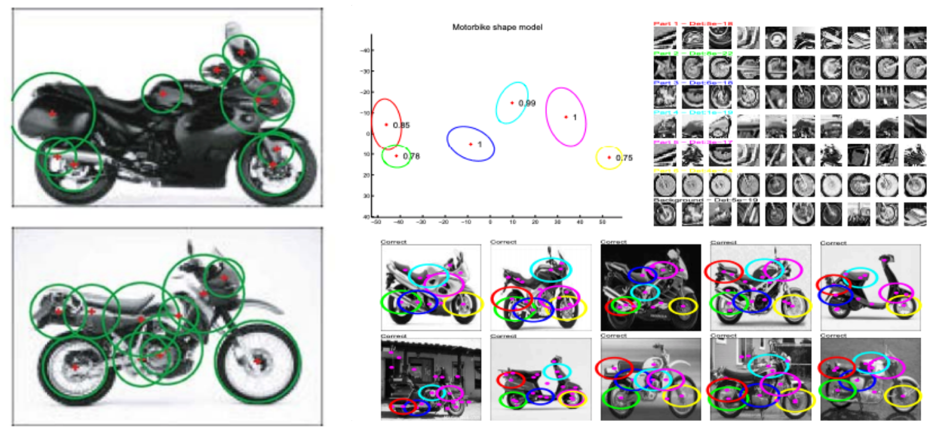


Pollefeys et al. (2004)

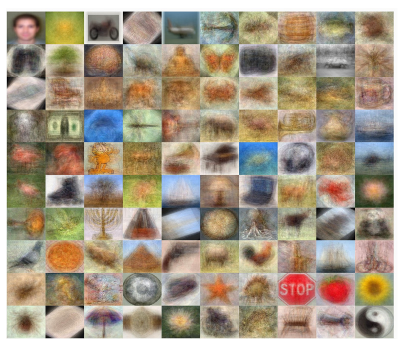
Snavely et al. (2006)

2000s: Recognition

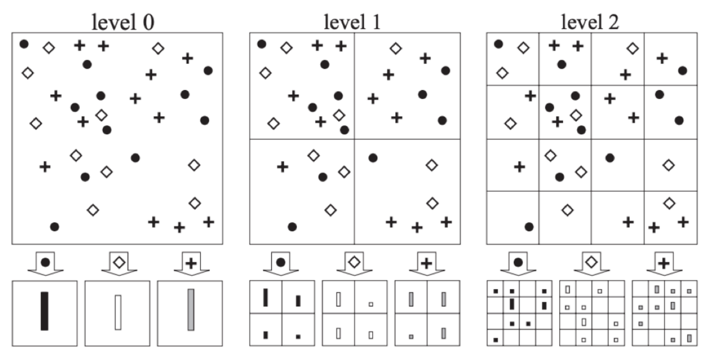
- Generic object recognition
  - Constellation models
  - Bags of features
  - Datasets: Caltech-101 -> ImageNet



Fergus, Perona & Zisserman (2007)



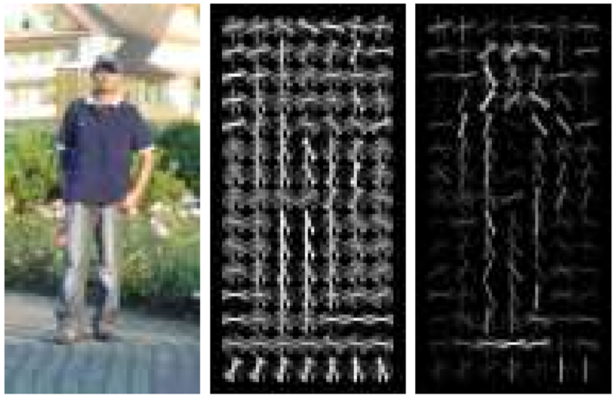
Caltech-101 (2005)



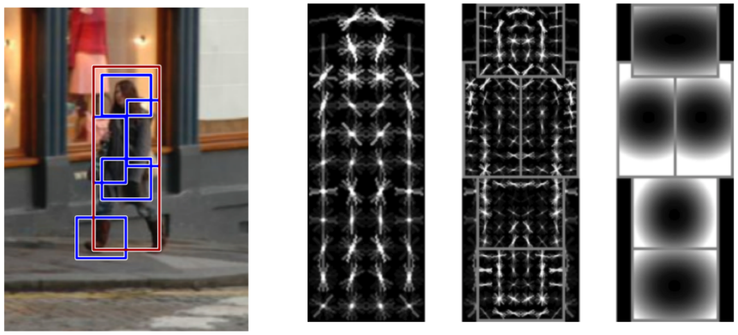
Lazebnik et al. (2006)

2000s: Recognition

- Generic object recognition
  - Constellation models
  - Bags of features
  - Datasets: Caltech-101 -> ImageNet
- Generic object detection
  - PASCAL dataset
  - HOG, Deformable part models
- Action and activity recognition: “misc. early efforts”



Dalal & Triggs (2005)



Felzenszwalb et al. (2010)

1990s-2000s: Dead ends (?)

- Probabilistic graphical models

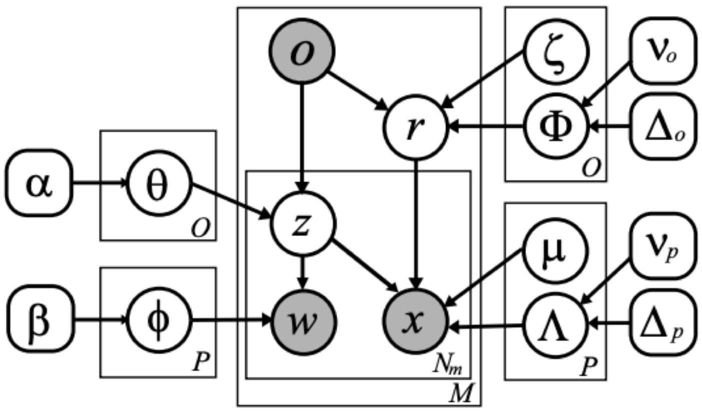
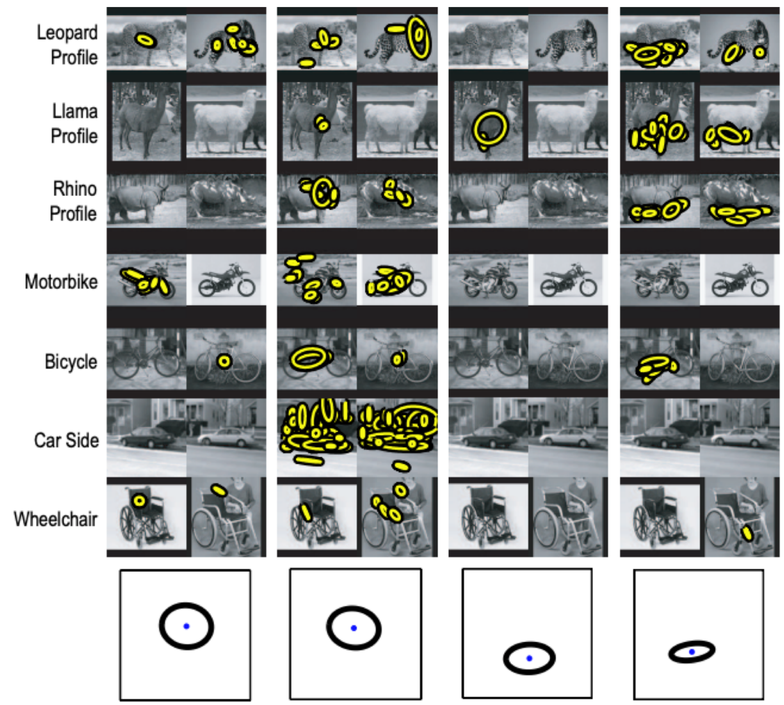


Figure 1. Graphical model describing how latent parts $\mathbf{z}$ generate the appearance $\mathbf{w}$ and position $\mathbf{x}$, relative to an image–specific reference location $\mathbf{r}$, of the features detected in an image of object $\mathbf{o}$. Boxes denote replication of the corresponding random variables: there are M images, with N_m observed features in image m.

E. Sudderth et al. Learning Hierarchical Models of Scenes, Objects, and Parts. ICCV 2005



Shared parts

1990s-2000s: Dead ends (?)

- Perceptual organization

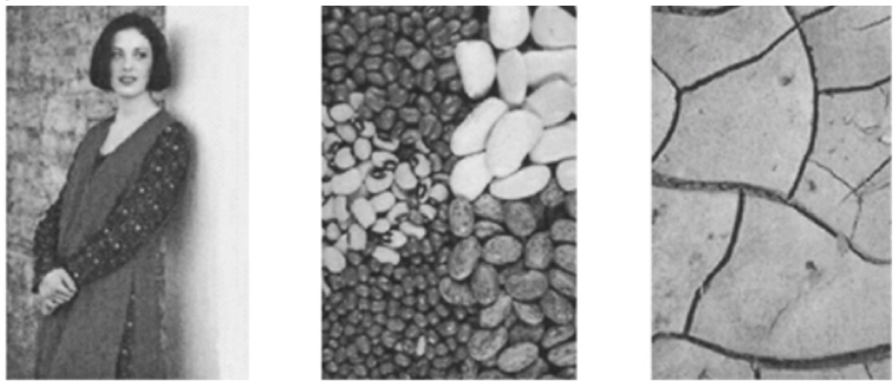


Figure 1. Some challenging images for a segmentation algorithm. Our goal is to develop a single grouping procedure which can deal with all these types of images.

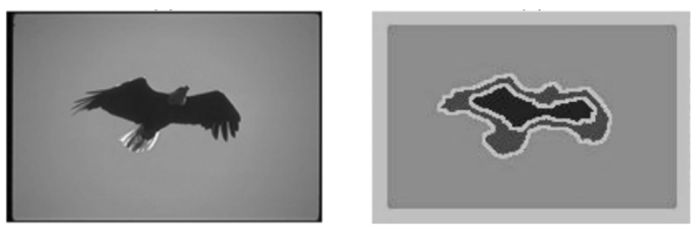


Figure 3. Demonstration of the "contour-as-a-texture" problem using a real image. (a) Original image of a bald eagle. (b) The groups found by an EM-based algorithm (Belongie et al., 1998).

J. Malik et al. [Contour and Texture Analysis for Image Segmentation](). IJCV 2001

1990s-2000s: Dead ends (?)

- Perceptual organization

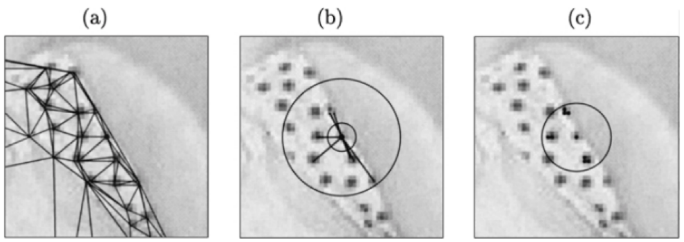


Figure 7. Illustration of scale selection. (a) Closeup of Delaunay triangulation of pixels in a particular texton channel for polka dot image. (b) Neighbors of thickened point for pixel at center. The thickened point lies within inner circle. Neighbors are restricted to lie within outer circle. (c) Selected scale based on median of neighbor edge lengths, shown by circle, with all pixels falling inside circle marked with dots.

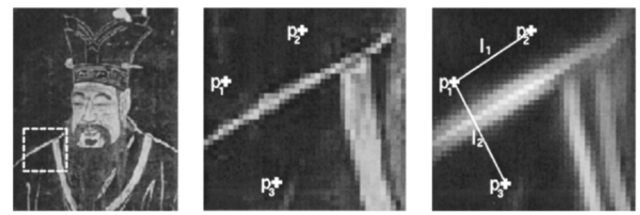


Figure 8. Left: the original image. Middle: part of the image marked by the box. The intensity values at pixels p_1, p_2 and p_3 are similar. However, there is a contour in the middle, which suggests that p_1 and p_2 belong to one group while p_3 belongs to another. Just comparing intensity values at these three locations will mistakenly suggest that they belong to the same group. Right: orientation energy. Somewhere along l_2, the orientation energy is strong which correctly proposes that p_1 and p_3 belong to two different partitions, while orientation energy along l_1 is weak throughout, which will support the hypothesis that p_1 and p_2 belong to the same group.

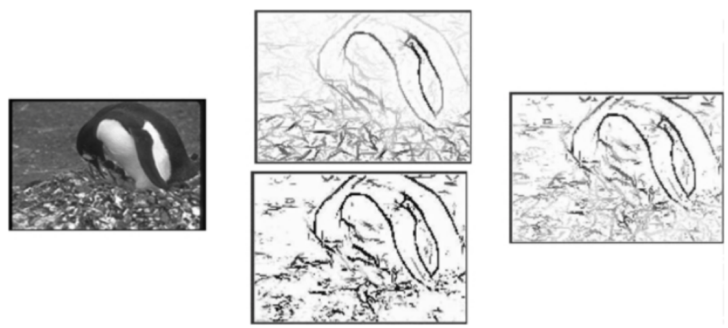


Figure 10. Gating the contour cue. Left: original image. Top: oriented energy after nonmaximal suppression, OE^*. Bottom: $1 - p_{texture}$. Right: p_B, the product of $1 - p_{texture}$ and $p_{con} = 1 - \exp(-OE^*/\sigma_{IC})$. Note that this can be thought of as a "soft" edge detector which has been modified to no longer fire on texture regions.

J. Malik et al. Contour and Texture Analysis for Image Segmentation. IJCV 2001

1990s-2000s: Dead ends (?)

- Perceptual organization

Segmentation results

J. Malik et al. [Contour and Texture Analysis for Image Segmentation](). IJCV 2001

Six decades of computer vision: Reductive summary

- **1960s and 70s: “Antiquity”**
  - Community goes through its blocks world phase
  - Canonical recovery problems are defined and initial approaches are proposed
  - Ambitious scene understanding approaches flower briefly and prematurely
  - Marr’s book sums up progress to date
- **1980s and 90s: “Middle ages”**
  - The field goes through its geometric recognition phase and gets over irrelevant geometric obsessions
  - Multi-view geometry matures and becomes useful, as summarized in the Hartley & Zisserman book
  - The field stops being afraid of pixels, probability, and statistical learning
- **2000s and 2010s: “Early modern era”**
  - Local features “solve” structure from motion and instance recognition
  - Generic category recognition and detection become central problems
  - The field becomes driven by datasets and benchmarks

What did I omit?

- Image filtering
  - Wavelets, steerable filters, bilateral filtering…
  - Biologically inspired low-level representations (Olhausen & Field, etc.)
- History of image generation
  - Starting with texture generation (Heeger & Bergen, Efros & Leung, etc.)
  - Stylization, inpainting, colorization, etc.
  - Image-based modeling and rendering
- Video
  - Optical flow estimation, action/activity recognition, etc.

Are there any themes?

- Optimization – not procedural reasoning
- Data and learning – not rules
- Rich appearance descriptors – not simple point and line features
- Deformable templates for recognition
- In the long run, data and computation win over cleverness – Sutton's "bitter lesson"

Where did we go wrong?

- In retrospect, computer vision has had several periods of "spinning its wheels"
  - We've always prioritized methods that could already do interesting things over potentially more promising methods that could not yet deliver
  - We've undervalued simple methods, data, and learning
  - When nothing worked, we distracted ourselves with fancy math
  - On a few occasions, we unaccountably ignored methods that later proved to be "game changers" (RANSAC, SIFT)
  - We've had some problems with bandwagon jumping and intellectual snobbery
- But it's not clear whether any of it mattered in the end…

Lana’s Top Ten “classical” vision papers

1. Hough transform – Duda & Hart, 1972
2. Pictorial structures – Fischler & Elschlager, 1973
3. RANSAC – Fischler & Bolles, 1981
4. Edge detection – Canny, 1986
5. Corner detection – Harris & Stephens, 1988
6. Normalized 8-point algorithm – Hartley, 1997
7. Graph cuts – Boykov et al., 2001
8. Face detection with boosting – Viola & Jones, 2001
9. SIFT – Lowe, 2004
10. Deformable part models – Felzenszwalb et al., 2010

Inspirational quote

“The past is never dead. It's not even past.”

– William Faulkner