Computer Vision

Introduction: Overview

Dr. Mrinmoy Ghorai

Indian Institute of Information Technology

Sri City, Chittoor

Today's Agenda

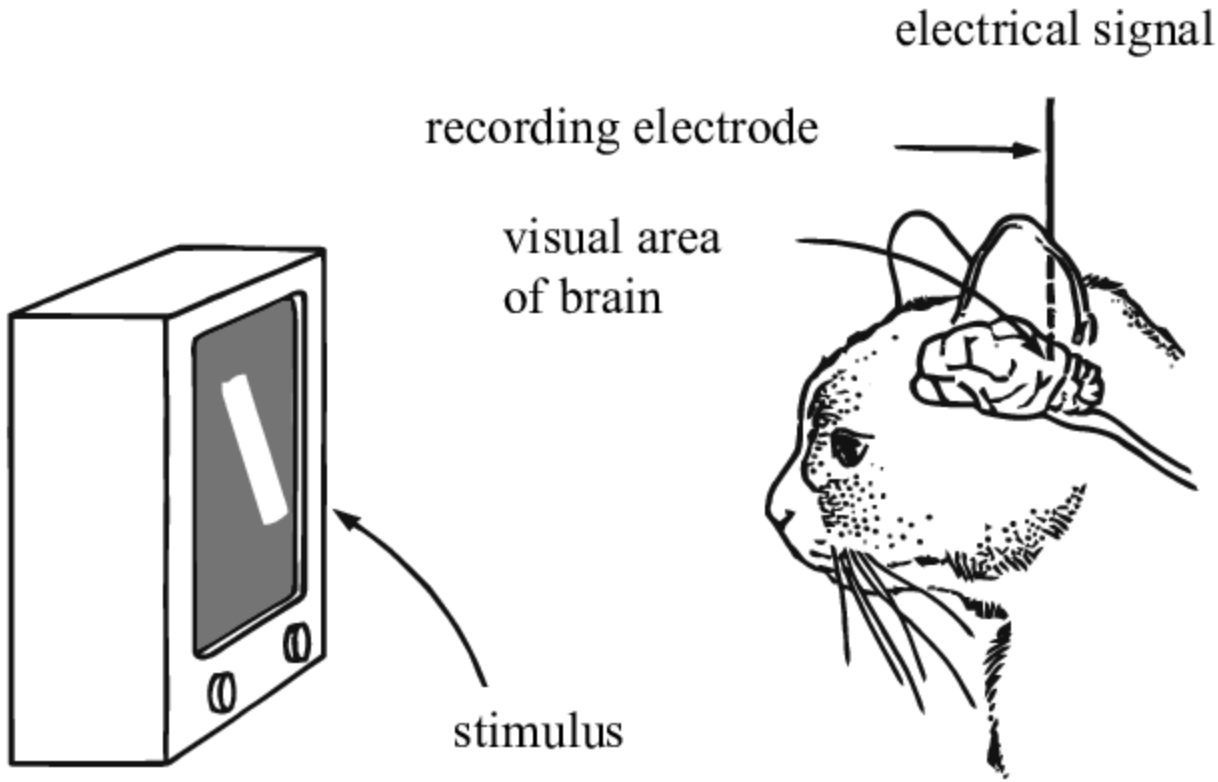
- History of computer vision
- Course overview

1940s – 1950s: Prelude

- 1943: McCulloch & Pitts, A Logical Calculus of the Ideas Immanent in Nervous Activity
- 1948: Wiener, *Cybernetics: or Control and Communication in the Animal and the Machine*
- 1949: Hebb, *The Organization of Behavior*
- 1950: Turing test
- 1956: Dartmouth workshop on AI
- 1957: Digital scanner invented at NIST
- 1958: Rosenblatt, “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain”
- 1959: Hubel & Wiesel, “Receptive fields of single neurones in the cat’s striate cortex”

Neuroscience Experiment

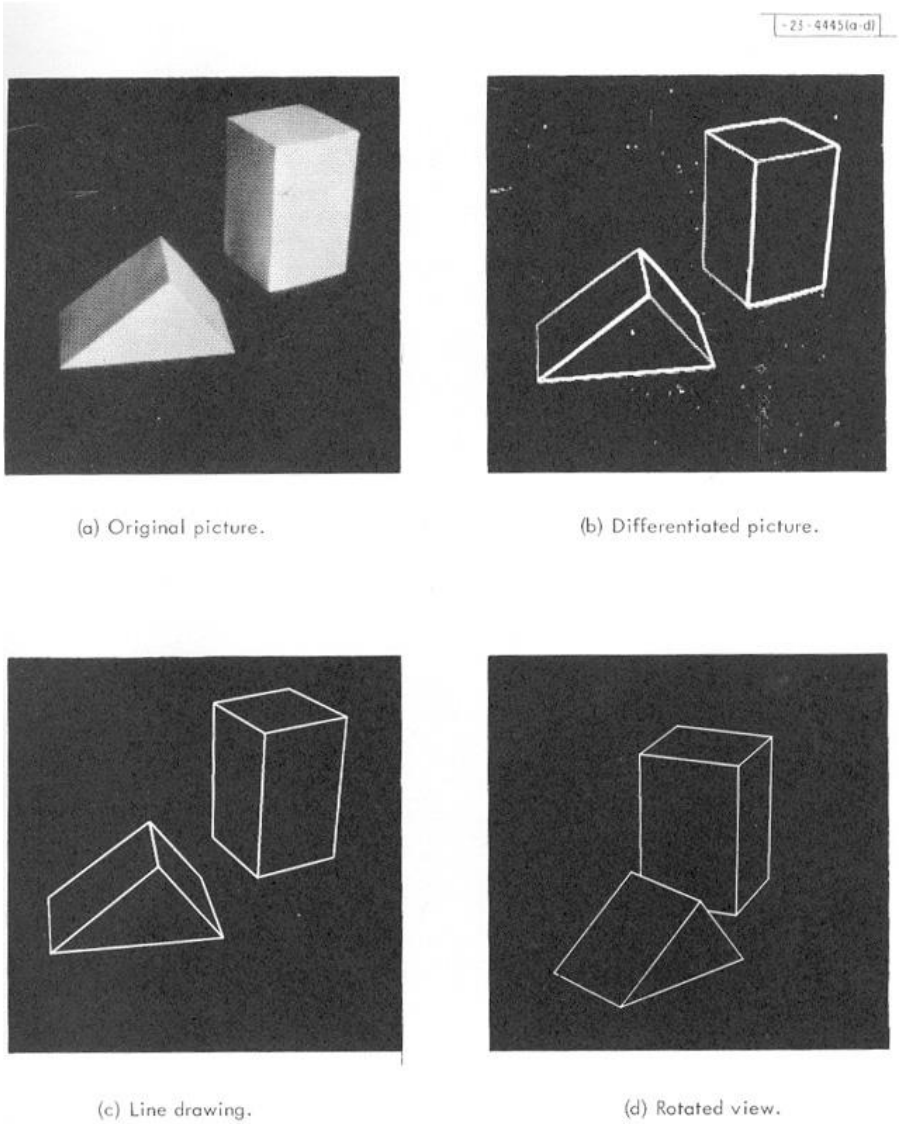
Hubel & Wiesel



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

1960s: the MIT-centric narrative



L. G. Roberts *Machine Perception of Three Dimensional Solids*

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."

1960s: the MIT-centric narrative

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".

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)

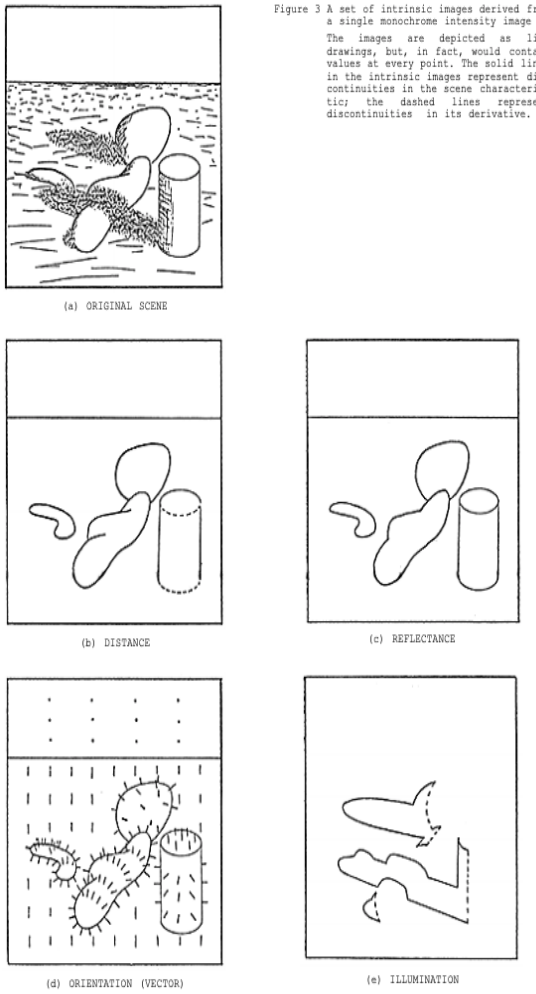


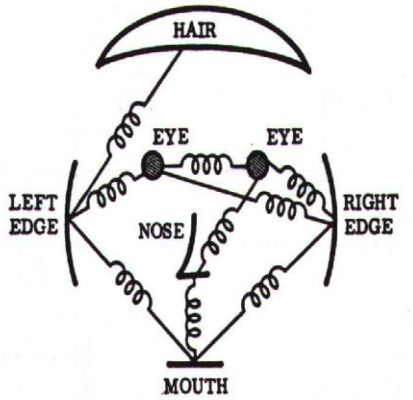
Figure 3 A set of intrinsic images derived from a single monochrome intensity image

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

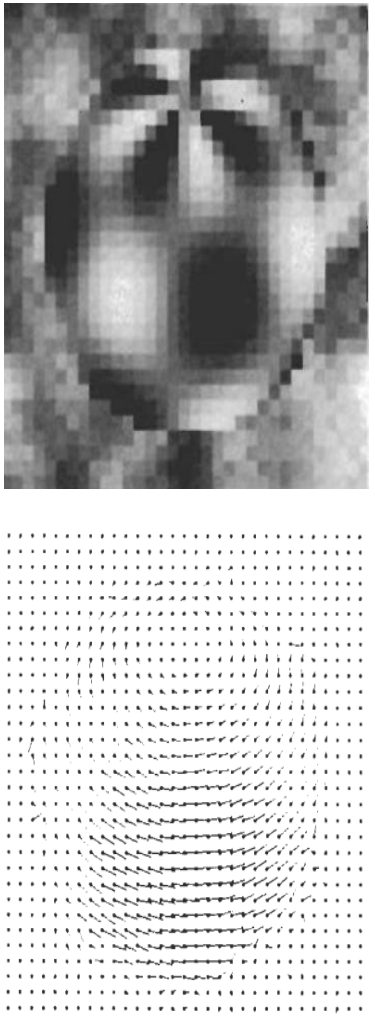
- 3D shape representation
  - Generalized cylinders: Binford et al. (1971, etc.)
- Deformable templates: Fischler & Elschlager (1973)
- Syntactic/procedural recognition systems
  - Faces: Kanade (1973)
  - Scenes: Yakimovsky & Feldman (1973), Hanson & Riseman (1978), Ohta & Kanade (1978)
  - Objects: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



Fischler & Elschlager (1973)

1980s: Progress on many fronts

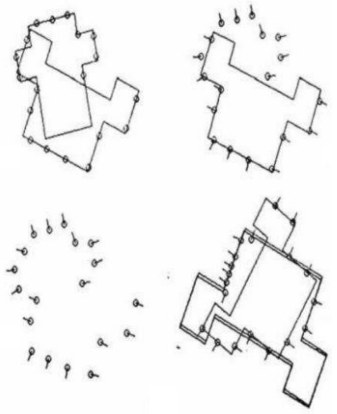
- Optical flow and tracking
  - Horn & Schunck (1981), Lucas & Kanade (1981), etc.
- “Definitive” detectors
  - Edges: Canny (1986)
  - Corners: Harris & Stephens (1988)
- Structure from motion
  - RANSAC: Fischler & Bolles (1981)
  - Essential matrix: Longuet-Higgins (1981)
- Markov Random Field models: Geman & Geman (1984)
- Image pyramids
  - Witkin (1983), Burt & Adelson (1984), Koenderink (1984), etc.
- Segmentation by energy minimization
  - Kass, Witkin & Terzopoulos (1987), Mumford & Shah (1989)
- Active vision
  - Bajczy (1985, 1988), Dickmanns (1988), Ballard (1989), etc.



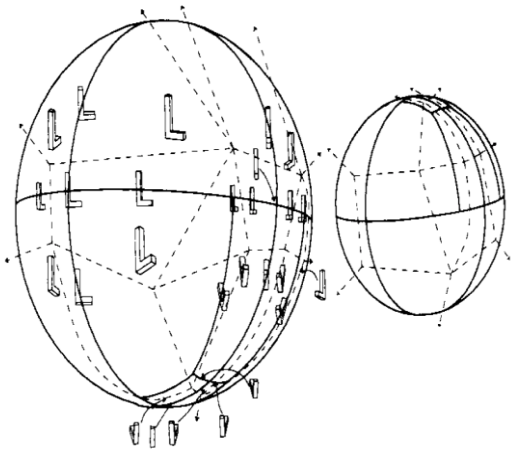
Horn & Schunck (1981)

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)



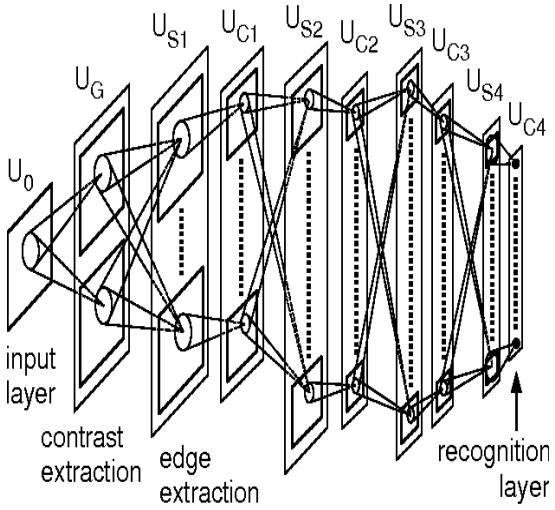
Grimson & Lozano-Perez (1984)



Gigus & Malik (1990)

1980s: Meanwhile…

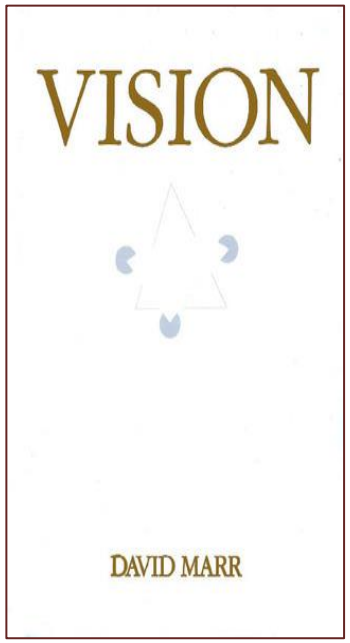
- Neocognitron: Fukushima (1980)
  - Video (short version)
- Back-propagation: Rumelhart, Hinton & Williams (1986)
  - Origins in control theory and optimization: Kelley (1960), Dreyfus (1962), Bryson & Ho (1969), Linnainmaa (1970)
  - Application to neural networks: Werbos (1974)
- Parallel Distributed Processing: Rumelhart et al. (1987)
- Neural networks for digit recognition: LeCun et al. (1989)

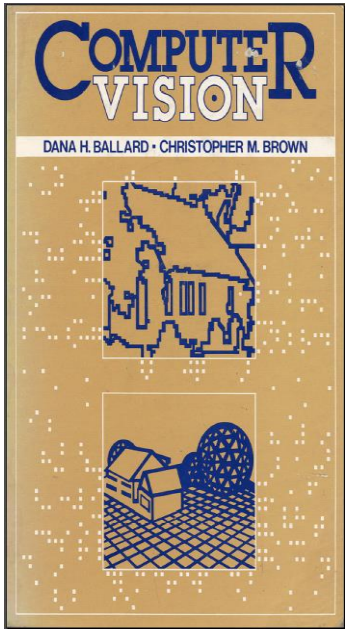


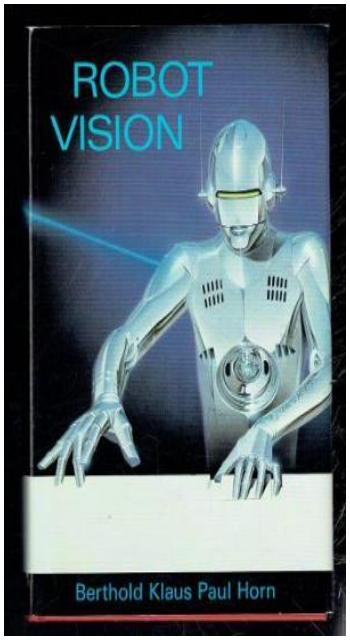
Fukushima (1980)

1980s: Milestones

- 1983: First CVPR
- 1987: First ICCV, IJCV
- Books: Marr (1982), Ballard & Brown (1982), Horn (1986)

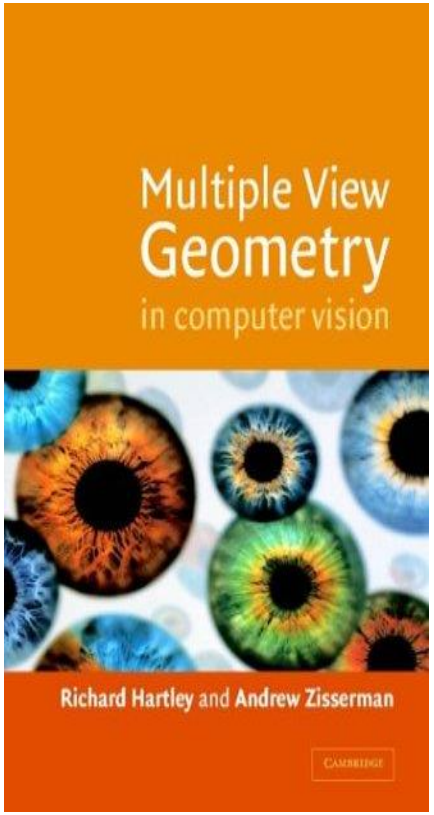






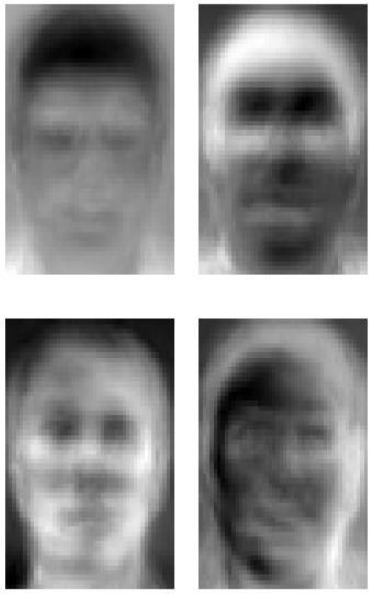
1990s: Geometry reigns

- 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

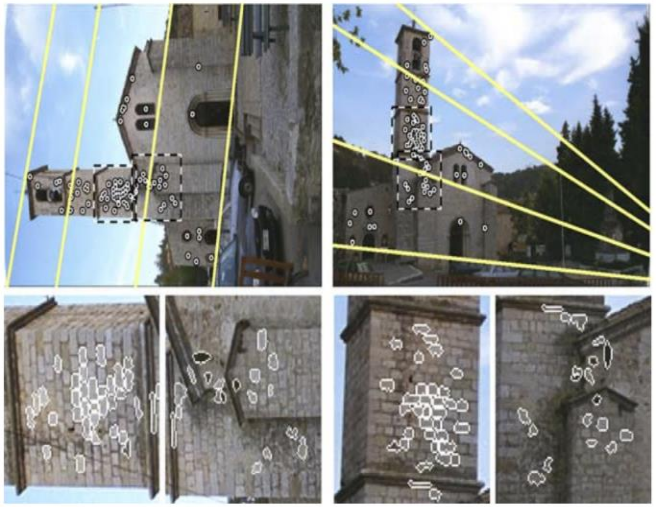
- Appearance-based models: Turk & Pentland (1991), Murase & Nayar (1995)
- Keypoint-based image indexing
  - Schmid & Mohr (1996), Lowe (1999)
- Constellation models for object categories
  - Burl, Weber & Perona (1998), Weber, Welling & Perona (2000)
- First sustained use of classifiers and negative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



Turk & Pentland (1991)

2000s: The era of features

- Keypoints craze
  - Kadir & Brady (2001), Mikolajczyk & Schmid (2002), Matas et al. (2002), Bay et al. (2006), etc.
- 3D reconstruction gets “solved”
  - SFM in the wild
  - Multi-view stereo, stereo on GPU’s
- Generic object recognition
  - Constellation models, graphical models craze
  - Bags of features
  - Datasets: Caltech-101, PASCAL, ImageNet
- Generic object detection
  - PASCAL dataset
  - HOG, Deformable part models
- Action and activity recognition – “misc. early efforts”



Matas et al. (2002)

Six decades of computer vision: Reductive summary

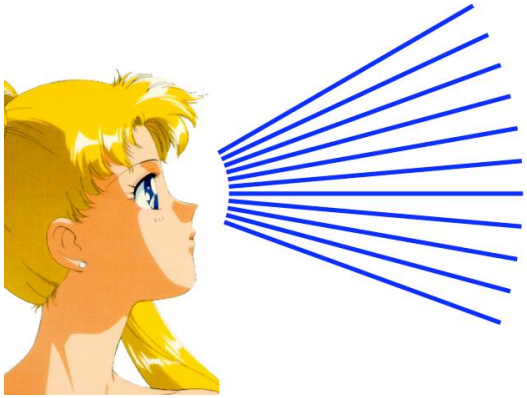
- **1960s and 70s: The "early years"**
  - Community gets over its blocks world phase
  - Canonical recovery problems are defined and initial approaches are proposed
  - Ambitious scene understanding approaches flower briefly and prematurely
  - Marr's book sums up progress to date
- **1980s and 90s: The "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 and discovers data and classifiers
- **2000s and 2010s: The 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

Course overview

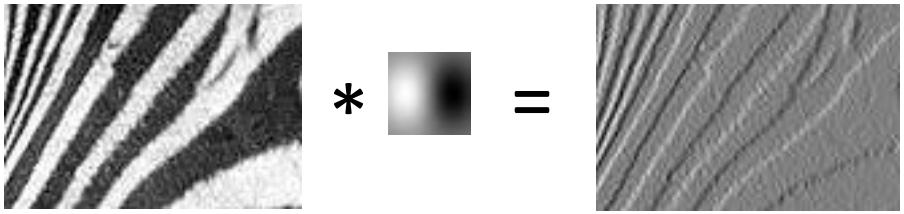
I. Early vision: Image formation and processing
II. Mid-level vision: Grouping and fitting
III. Multi-view geometry
IV. Recognition
V. Additional topics

Early vision

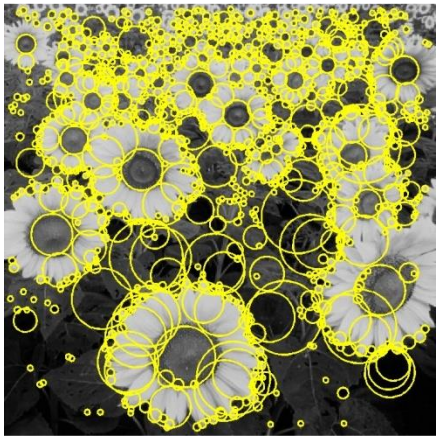
- Basic image formation and processing



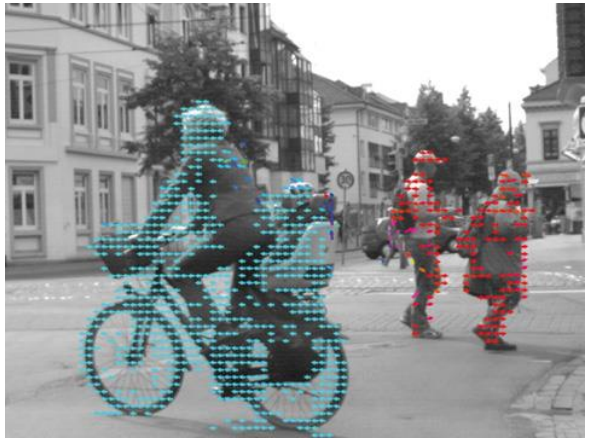
Cameras and sensors
Light and color



Linear filtering
Edge detection



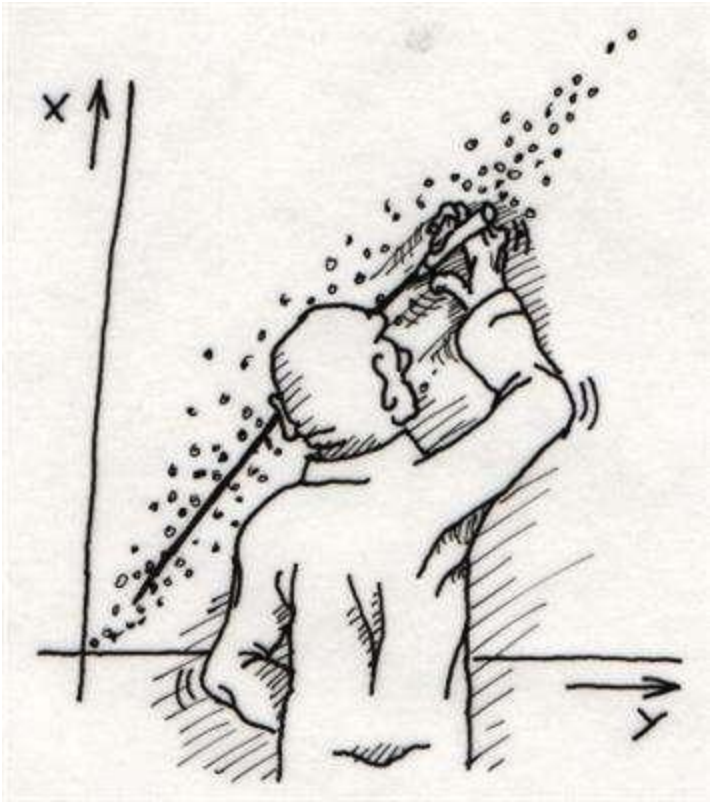
Feature extraction



Optical flow

“Mid-level vision”

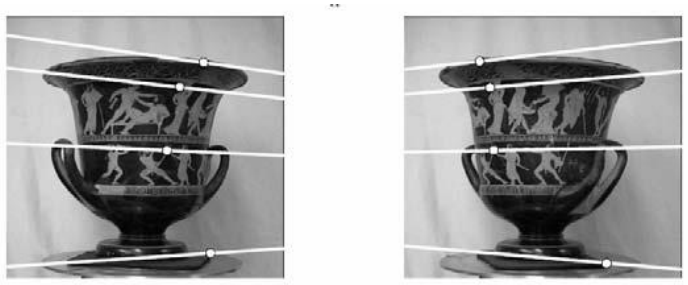
- Fitting and grouping



Fitting: Least squares
Voting methods

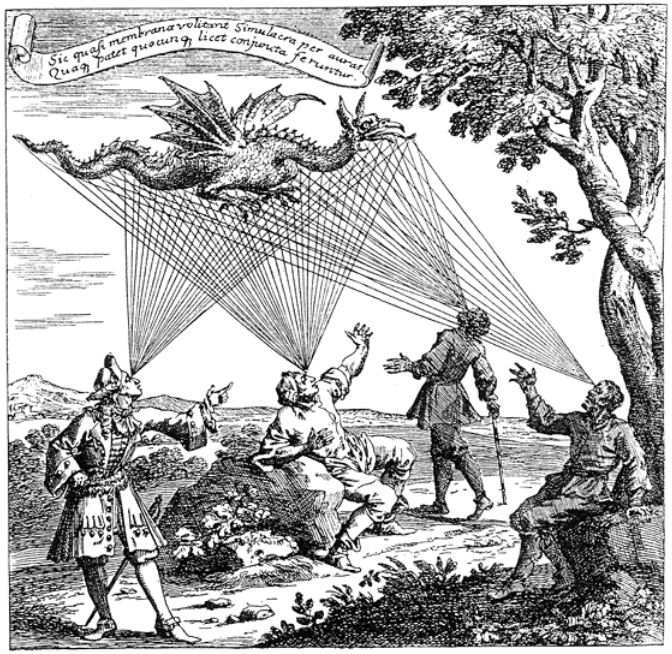
Alignment

Multi-view geometry



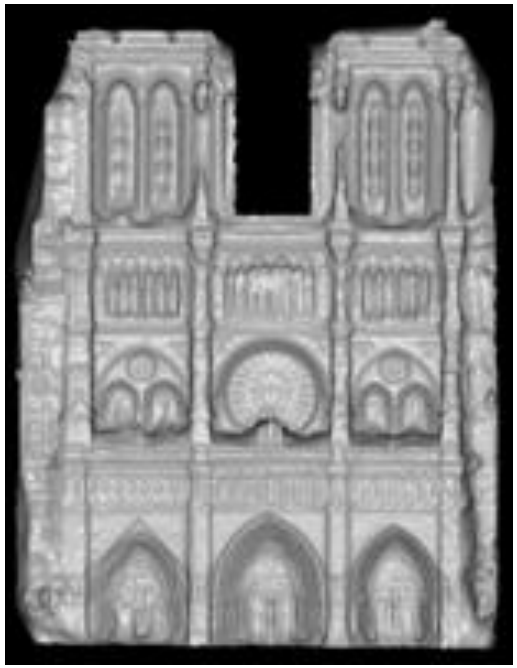
Epipolar geometry

Two-view stereo



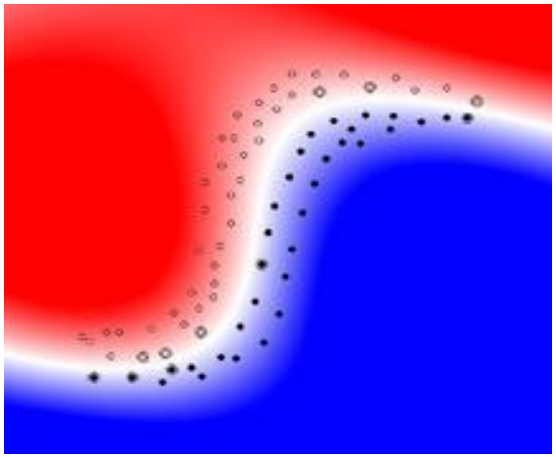
Драконъ, видимый подъ различными углами зрѣнія
По гравюрѣ на мѣди изъ „Oculus artificialis teledioptricus" Цана. 1702 года.

Structure from motion

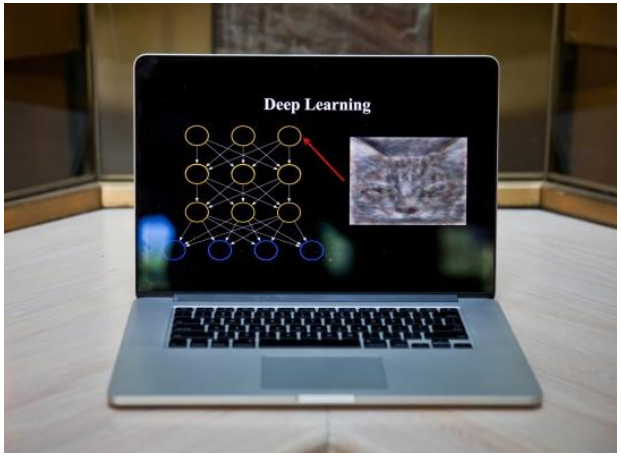


Multi-view stereo

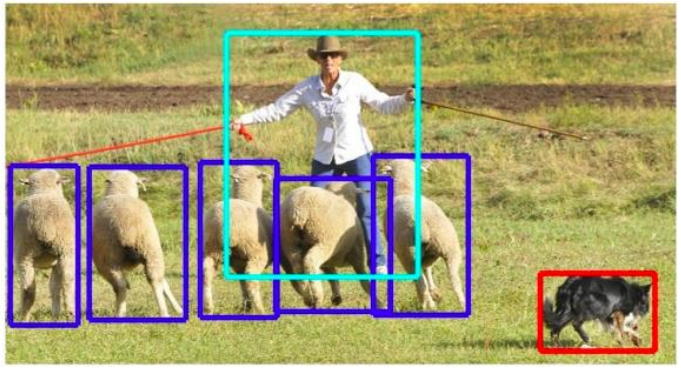
Recognition



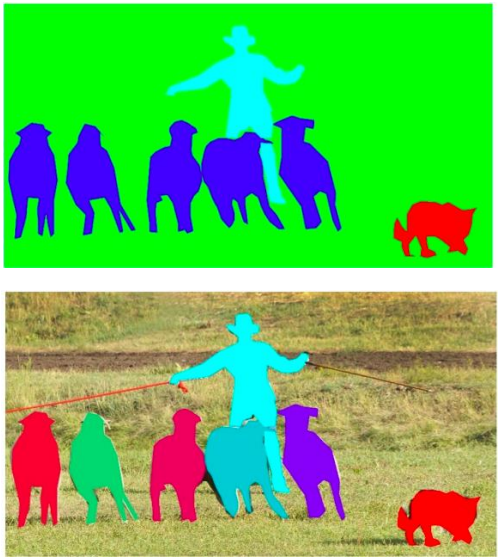
Basic classification



Deep learning



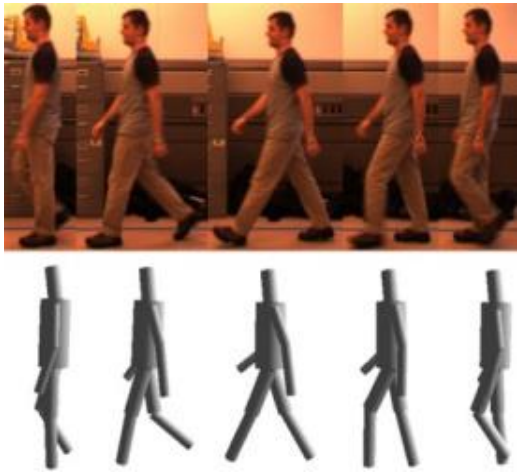
Object detection



Segmentation

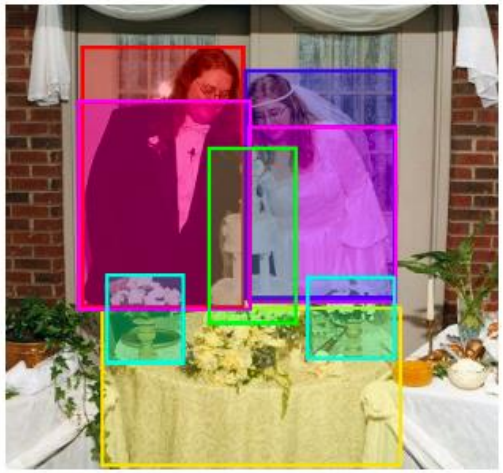
Additional Topics (time permitting)

Generation



Video

3D scene understanding



Images and text

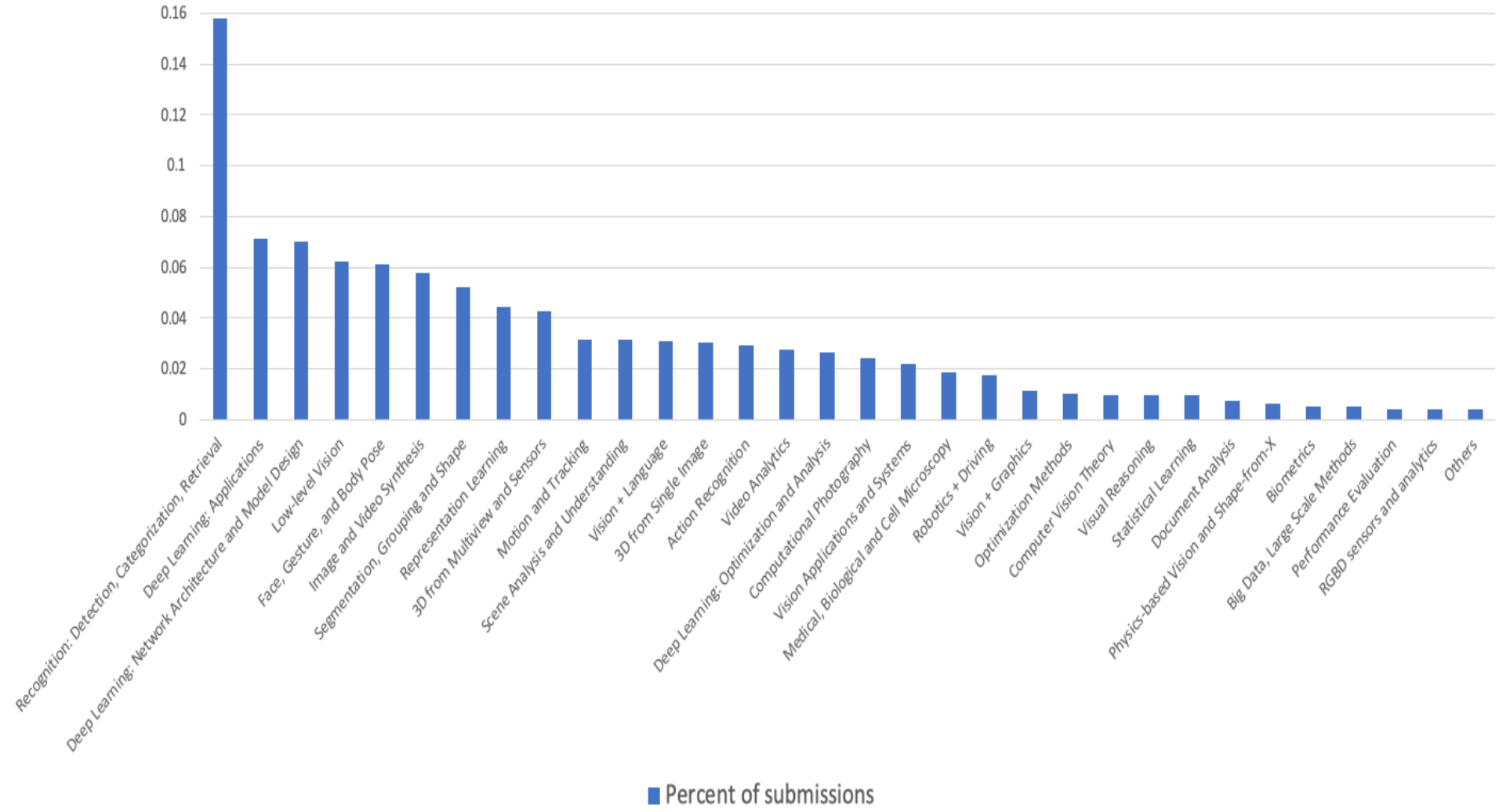
Resources: Journals

- IEEE Transactions on Pattern Analysis and Machine Intelligence
- International Journal of Computer Vision
- Pattern Recognition
- Computer Vision and Image Understanding
- Image and Vision Computing
- Machine Vision and Applications

Resources: Conferences

- IEEE International Conference on Computer Vision (ICCV)
- IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)
- Winter Conference on Applications of Computer Vision (WACV)
- ACM SIGGRAPH
- European Conference on Computer Vision (ECCV)

Topic distribution of submissions for ICCV 2019



Total submissions: 4289

Most Valuable 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

Thank you: Question?