Computer Vision: Looking Back to Look Forward

Svetlana Lazebnik
IRIM Short Course
Spring 2020
Or, this is me asking, “What does it all mean?”
Outline of lectures

**Tues. 1/28:** Where are we going, where have we been?
**Wed. 1/29:** Stepping back: What is visual perception?
**Thurs. 1/30:** History of ideas in recognition – Part I

**Tues. 2/4:** History of ideas in recognition – Part II – *TSRB banquet hall*
**Wed. 2/5:** Emerging trends – or, what should we be working on?
**Thurs. 2/6:** Ethical issues for computer vision researchers
Today: Where are we going, where have we been?
Fear not: CVPR 2020 received 6655 submissions, “only” 30% over last year’s 5160

56% yearly growth with 26% acceleration → 10.8B submitted papers in 2028

Source: CVPR 2019 opening slides
CVPR Attendance Trend

Source: CVPR 2019 opening slides
Deep learning “revolution”

**AlexNet** (2012)

Photo source

---

*I wanna be in the room where it happens.*

---

Yann LeCun → Public

+Alex Krizhevsky’s talk at the ImageNet ECCV workshop yesterday made a bit of a splash. The room was overflowing with people standing and sitting on the floor. There was a lively series of comments afterwards, with +Alyosha Efros, Jitendra Malik, and I doing much of the talking.

---

Photo source
Total submissions: 4289

Topic distribution of submissions for ICCV 2019

Not deep learning: <10% of submissions?
Amusing how some computer vision researchers jokingly refer to work done before 2012 as "prehistoric".
So, what can today’s researchers learn from “prehistoric” computer vision?
Let’s dig into the history!
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
- 1959: Hubel & Wiesel, “Receptive fields of single neurones in the cat’s striate cortex”
**Origins of AI hype?**

**NEW NAVY DEVICE LEARNS BY DOING**

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau’s $2,000,000 “704” computer—learned to differentiate between right and left after fifty attempts in the Navy’s demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

**Without Human Controls**

The Navy said the perceptron would be the first non-living mechanism “capable of receiving, recognizing and identifying its surroundings without any human training or control.”

The “brain” is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be conscious of their existence.

---

1958 New York Times...

In today’s demonstration, the “704” was fed two cards, one with squares marked on the left side and the other with squares on the right side.

**Learns by Doing**

In the first fifty trials, the machine made no distinction between them. It then started registering a “Q” for the left squares and “O” for the right squares.

Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a “self-induced change in the wiring diagram.”

The first Perceptron will have about 1,000 electronic “association cells” receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.
Early applications of image analysis

• 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
Early applications of image analysis

• 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 (IEEE interview, 1998)
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.
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.
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

• 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
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)
1960s: the MIT-centric narrative wouldn’t be complete without...

- 1969: Minsky and Papert, *Perceptrons*

Meanwhile: Blocks world and pattern recognition on the West Coast

Shakey the Robot
SRI, 1966 - 1972
Video

Published in 1972
(table of contents)
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)

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
1980s: Progress on many fronts

- Optical flow and tracking

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

- Structure from motion


- Image pyramids

- Segmentation by energy minimization

- Active vision
1980s: The dead ends

• Alignment-based recognition

• Aspect graphs

• Invariants: Mundy & Zisserman (1992)
1980s: Milestones

• 1983: First CVPR
• 1987: First ICCV, IJCV
1980s: Meanwhile...

  - Video (short version)
- Parallel Distributed Processing: Rumelhart et al. (1987)
- Neural networks for digit recognition: LeCun et al. (1989)
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)
• Projective structure from motion: Faugeras and Luong (2001)
1990s: Data enters the scene

- Keypoint-based image indexing
- Constellation models for object categories
- First sustained use of classifiers and negative data
  - 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
- Tracking of complicated shapes
Late 1990’s debates

• See the last chapter of *Vision Algorithms: Theory and Practice* (1999)
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”
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
What did I omit?

• Image filtering
  • Wavelets, steerable filters, bilateral filtering...
  • Biologically inspired low-level representations (Olhausen & Field, etc.)

• Generation methods
  • E.g., texture generation (Heeger & Bergen, Efros & Leung, etc.)

• Image-based modeling and computational photography

• Dense reconstruction

• Probabilistic and graphical models (on purpose)

• Perceptual organization (on purpose)
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 MVP’s (Most Valuable Papers)

2. Pictorial structures – Fischler & Elschlager, 1973
3. RANSAC – Fischler & Bolles, 1981
4. Edge detection – Canny, 1986
5. Corner detection – Harris & Stephens, 1988
7. Graph cuts – Boykov et al., 2001
9. SIFT – Lowe, 2004
10. Deformable part models – Felzenszwalb et al., 2010
Further reading

• Azriel Rosenfeld interview (1998)
• Rosenfeld survey (1998)
• Computer vision: the last 50 years by Linda Shapiro
• Rick Szeliski book (history in sec. 1.2)
• Computer vision awards
• Serge Belongie’s Top 10 PAMI articles