CS 534: Computer Vision

Ahmed Elgammal
Dept of Computer Science
Rutgers University

Outlines

• Vision What and Why?
• Human vision
• Computer vision
• General computer vision applications
• Course Outlines
• Administrative
What is vision?

• What does it mean to see?

“The plain man’s answer (and Aristotle’s too) would be, to know what is where by looking. In other words, vision is the process of discovering from images what is present in the world, and where it is.” David Marr, Vision 1982

What is vision?

• Recognize objects
  – people we know
  – things we own
• Locate objects in space
  – to pick them up
• Track objects in motion
  – catching a baseball
  – avoiding collisions with cars on the road
• Recognize actions
  – walking, running, pushing
Vision is

- Deceivingly easy
- Deceptive
- Computationally demanding
- Critical to many applications
Vision is deceivingly easy

- We see effortlessly
  - seeing seems simpler than “thinking”
  - we can all “see” but only select gifted people can solve “hard” problems like chess
  - we use nearly 70% of our brains for visual perception!

- All “creatures” see
  - frogs “see”
  - birds “see”
  - snakes “see”
  but they do not see alike

---

Vision is deceivingly easy

- The M.I.T. summer vision program
  - summer of 1965
  - point TV camera at stack of blocks
  - locate individual blocks
    - recognize them from small database of blocks
  - describe physical structure of the scene
    - support relationships
- Formally ended in 1985
Vision is deceptive

• Vision is an exceptionally strong sensation
  – vision is immediate
  – we perceive the visual world as external to ourselves, but it is a reconstruction within our brains
  – we regard how we see as reflecting the world “as it is;” but human vision is
    • subject to illusions
    • quantitatively imprecise
    • limited to a narrow range of frequencies of radiation
    • passive

Vision is deceptive

• Human vision is
  – subject to illusions
  – quantitatively imprecise
  – limited to a narrow range of frequencies of radiation
  – passive
Zollner's illusion - 1860

Delboef's illusion - Context Sensitivity
More illusion

- Subjective contours

More illusion

- Subjective contours
- Figure completion
Necker cube: The human visual system picks an interpretation of each part that makes the whole consistent.

More illusion

- Subjective contours
- Depth, reversibility, Figure completion
More illusion

• Depth, reversibility, Do the cubes shift independently or as a unit

More illusion

• The Hermann grid illusion: Illusory dark spots appear at all the intersections of the white stripes except the one on which you are currently fixated; lateral inhibition
More illusion

• We can see impossible figures

Spectral limitations of human vision

• We “see” only a small part of the energy spectrum of sunlight
  – we don’t see ultraviolet or lower frequencies of light
  – we don’t see infrared or higher frequencies of light
  – we see less than .1% of the energy that reaches our eyes
• But objects in the world reflect and emit energy in these and other parts of the spectrum
Non-human vision

- Infrared vision
- Polarization vision
  - navigation for birds and insects
- Ultrasound vision
- X-ray vision!
- RADAR vision
Infrared vision

- Vision systems exist that can see reflected and emitted infrared light
  - visual system of the pit viper
  - infrared cameras used for night vision
- Why don’t we see the infrared?
  - we would see the blood flow through the capillaries in the eye

Human vision is passive

- It relies on external energy sources (sunlight, light bulbs, fires) providing light that reflects off of objects to our eyes
- Vision systems can be “active” - carry their own energy sources
  - Radars
  - Bat acoustic imaging systems
According to Marr:

- Vision is an information-processing task
- But not just a process
- Our brain must somehow be capable of representing this information.

"vision study ... not only the study of how to extract from images the various aspects of the world that are useful to us, but also an inquiry into the nature of the internal representations by which we capture this information and thus make it available as a basis for decisions about our thoughts and actions"

We want to know how to program vision.
Computer Vision

- Understanding the content of images and videos

Azriel Rosenfeld: Picture Processing By Computer 1969
Related Fields: AI, pattern recognition, machine learning, signal processing, neural networks, cognitive vision.

Vision is deceivingly easy
= Computer Vision is hard

- The M.I.T. summer vision program
  - summer of 1966
  - point TV camera at stack of blocks
  - locate individual blocks
    - recognize them from small database of blocks
  - describe physical structure of the scene
    - support relationships
**Goals - General**

The primary goal of the project is to construct a system of programs which will divide a vidaesector picture into regions such as likely objects, likely background areas, chaos.

We shall call this part of its operation **FIGURE-GROUND analysis**.

It will be impossible to do this without considerable analysis of shape and surface properties, so **FIGURE-GROUND analysis** is really inseparable in practice from the second goal which is **REGION DESCRIPTION**.

The final goal is **OBJECT IDENTIFICATION** which will actually name objects by matching them with a vocabulary of known objects.

---

**Goals - Specific**

We plan to work by getting a simple form of the system going as soon as possible and then elaborating upon it. To keep the work reasonably coordinated there is a graduated scale of subgoals.

**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.

**Subgoal 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.
Vision is deceivingly easy
= Computer Vision is hard

• The M.I.T. summer vision program
  – summer of 1966
  – point TV camera at stack of blocks
  – locate individual blocks
    • recognize them from small database of blocks
  – describe physical structure of the scene
    • support relationships
• Formally ended in 1985

“ The first great revelation was that the problems are difficult. Of course, these days this fact is a commonplace. But in the 1960s almost no one realized that machine vision was difficult. The field had to go through the same experience as the machine translation field did in its fiascoes of the 1950’s before it was at least realized that here were some problems that had to be taken seriously.” D. Marr, Vision, 1982.

---

Early CV History (A. Rosenfeld, 1998)

since 1960 Digital image processing by computer
1968 1. Journal: Pattern Recognition (Pergamon, now Elsevier)
1969 1. Textbook: Picture Processing by Computer (A. Rosenfeld)
1970 1. International Conference on Pattern Recognition (ICPR)
1977 1. Computer Vision and Pattern Recognition (CVPR)
1978 International Association for Pattern Recognition (IAPR)
1979 IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)
1987 1. International Conference on Computer Vision (ICCV)
Understanding and Recognition

- People draw distinctions between what is seen
  - “Object recognition”
  - This could mean “is this a fish or a bicycle?”
  - It could mean “is this George Washington?”
  - It could mean “is this poisonous or not?”
  - It could mean “is this slippery or not?”
  - It could mean “will this support my weight?”
- Great mystery
  - How to build programs that can draw useful distinctions based on image properties

Generic Object Recognition

- Variations in scale, orientation and visibility
- Variability within Specificity
- Object of interest has to be recognized in context of multiple other objects and cluttered background
What are the problems in recognition?

- Which bits of image should be recognized together?
  - Segmentation.
- How can objects be recognized without focusing on detail?
  - Abstraction.
- How can objects with many free parameters be recognized?
  - No popular name, but it's a crucial problem anyhow.
- How do we structure very large model-bases?
  - again, no popular name; abstraction and learning come into this

Why to study Computer Vision?

- Cameras are everywhere now: in our pockets, watching over us at different scales, even inside our bodies.
  - Need tools for processing, annotation, archiving,…

- Images and movies are everywhere:
  - In 2009: 2.5 B photos uploaded to Facebook every month
  - That's 30 B photos per year on FB alone
  - In 2011: 6 B per month – 90 B total
  - Youtube: 48 hours of video are uploaded every minute, resulting in nearly 8 years of content uploaded every day

- We are at the golden age for computer vision
Why to study Computer Vision?

- Essential for real-time applications
  - Building 3D representations of the world from pictures and videos is essential for robotics
  - Recognizing objects around us

- Fast-growing collection of useful applications

- Various deep and attractive scientific mysteries
  - how does object recognition work?

- Greater understanding of human vision

Computer Vision Applications

- Communications: OCR, Virtual Reality, Human-machine interface (facial expression and gesture recognition).
- Automated surveillance (who’s doing what, where and how)
- Medicine: diagnosis, remote and telemedicine, surgical assistance
- Robot Navigation and object manipulation
- Transportation: autonomous driving: lane detection, pedestrian detection, …
- Entertainment: Video archival and retrieval, Motion capture, Augmented reality,
- Defense
- …
Manufacturing – Machine Vision

• Visual inspection for quality control
  – during the manufacture of parts in the automotive industry
  – inspection of semiconductors
• Visual control of robots
  – during assembly of parts from pieces
  – during calibration of robot control systems

Communications

• Smart document readers
  – character recognition
  – discrimination of text from graphics and images
  – reading cursive script
  – “language” recognition
• Virtual teleconferencing
• Virtual reality
Medicine

- Diagnosis
  - radiology - read X rays, CAT scans
  - pathology - read biopsies
- Remote and tele-medicine
- Virtual reality surgical assistance
  - project images onto head during brain surgery

Figures by kind permission of Eric Grimson; further information can be obtained from his web site http://www.ai.mit.edu/people/welg/welg.html.
Figures by kind permission of Eric Grimson; further information can be obtained from his web site http://www.ai.mit.edu/people/welg/welg.html.
Transportation

- Traffic safety and control
  - detection and ticketing of speeding vehicles
  - vehicle counting for flow control
- Robot drivers
  - convoys
- Advanced automobiles
  - autonomous parallel parking
  - road sign detectors and driver alerts
  - collision avoidance – detecting cars in blind spots
  - Pedestrian detection
  - smart cruise control (lane detection, car following)
  - Occupant detection

Pittsburgh to San Diego! – CMU 1996

Today: Google Car
Traffic sign results

Videos by: Dr. Vasanth Philomin - Dariu Gavrila, Vasanth Philomin: Real-Time Object Detection for "Smart" Vehicles. ICCV 1999

Pedestrian detection results

Videos by: Dr. Vasanth Philomin
Pedestrian detection - IR images

Videos by: Dr. Vasanth Philomin

- http://www.youtube.com/watch?v=XU43SjMIzW8&noredirect=1
**Entertainment**

- Acquisition of 3D computer models for graphical manipulation
- Augmented Reality
- Control of animation through vision – marker less motion capture
- Indexing tools for video databases

**Applications – Sport Broadcasting**

Tracking Baseball Pitches for Broadcast Television
- The system is used by ESPN for its Major League Baseball broadcast.
- The system draw a representation of the strike zone on the TV screen superimposed over the replayed broadcast video. The system would determine electronically whether the each pitch qualified as a strike or a ball.
- Detect ground plan in video and introduce pictures on them

Images and videos from: SYMAH VISION, Easily Virtual www.symah-vision.fr
CNN Hologram 2008

- http://www.youtube.com/watch?v=thOxW19vsTg
- http://www.youtube.com/watch?v=deoOTqT-SMI
Computational Photography

- Photo Editing tools
- Inpainting
- Super resolution
- Image matting, video matting
- Object Manipulation in videos

- Example Demos: check out vimeo.com
  - http://vimeo.com/5024379
  - http://vimeo.com/2345579
Agriculture

• Safety and quality inspection
  – sorting by size - peaches
  – sorting by shape - potatoes
  – identifying defects - blemishes on fruit, rot in potatoes
  – disease monitoring - chickens
• Robotic farming equipment
  – robotic harvesters - apple pickers, orange pickers

Looking at People

• Human detection and Tracking
• Human limbs tracking
• Human recognition and biometrics
  – Face recognition
  – Gait recognition
  – Iris, etc.
• Gesture recognition
• Facial expression recognition
• Activity recognition

Duchenne de Boulogne, C.-B. (1862) *The Mechanism of Human Facial Expression*
Human Motion Analysis
Humans are typically the most interesting object in images and videos
Why Human Motion is Challenging?
• Articulation
• Variability: different body styles, clothing
• Self occlusion
Many Applications: visual surveillance, human-machine interface, video archival and retrieval, computer graphics and animation, autonomous driving, virtual reality, games…

Applications
• Human Computer Interaction
  – Keyboard and mouse are restrictive
• Driver Assistance, Autonomous driving
• Motion Capture
• Video editing, archival and retrieval.
• Surveillance: security, safety, resource management
Visual Surveillance

Consider a visual surveillance system

State of the art: archive huge volumes of video for eventual off-line human inspection

Goal: Automatic understanding of events happening in the site.

– Efficient archiving
– Automatic Annotation
– Direct human attention
– Reduce bandwidth required for video transmission and storage.

Recognizing facial expressions

Videos by: Dr. Yaser Yacoob - Black, M.J., Yacoob, Y., Jepson, A.D., Fleet, D.J., Learning Parameterized Models of Image Motion, CVPR97
Recognizing facial expressions

Videos by: Dr. Yaser Yacoob - Black, M.J., Yacoob, Y., Jepson, A.D., Fleet, D.J., Learning Parameterized Models of Image Motion, CVPR97

Recognizing facial expressions

Videos by: Dr. Yaser Yacoob - Black, M.J., Yacoob, Y., Jepson, A.D., Fleet, D.J., Learning Parameterized Models of Image Motion, CVPR97
Motion Capture

Videos by: Dr. Thanarat Horprasert
Defense

- Automatic target recognition systems
  - cruise missiles
  - air to surface “smart” missiles
- Reconnaissance
  - monitoring strategic sites
- Simulation
  - acquisition of terrain models from imagery
  - model acquisition of buildings, roads, etc.
Structure from Motion

- Camera Stabilization
- 3D scene reconstruction
- Localization from Video
- Visual Odometry
- SLAM: Simultaneous localization and mapping

- Examples Demo Systems: Photo tourism, finding paths through the world’s photos
Computer Vision – Course Outlines

**Image Formation**
- Human vision
- Cameras
- Geometric Camera models
- Camera Calibration
- Radiometry
- Color

**Early Vision (one image)**
- Linear Filters
- Edge Detection
- Local Features
- Texture
- Motion

**Early Vision (Multiple images)**
- Geometry of Multiple images
- Stereo

**Mid-Level Vision:**
- Segmentation
  - By clustering
  - By model fitting
  - Probabilistic
- Tracking

**High-Level Vision:**
- Model-based vision
- Appearance-based vision
- Generic object recognition

Course Outline

- **Part I: The Physics of Imaging**
  Image formation and image models: Cameras, light, color
- **Part II: Early Vision in One Image**
  Edges and texture
- **Part III: Early Vision in Multiple Images**
  Stereopsis, structure from motion
- **Part IV: Mid-Level Vision**
  Finding coherent structure in images and movies: Segmentation, Tracking
- **Part V: High Level Vision (Geometry)**
  The relations between object geometry and image geometry: Model-based vision
- **Part VI: High Level Vision (Probabilistic)**
  Using classifiers and probability to recognize objects
Resources:
• Visual illusion http://dragon.uml.edu/psych/illusion.html