Introduction

CSC 249/449
Sensory Motor Systems

Randal C. Nelson

What is Computer Vision

Classical definition: automatic extraction of a (universal) symbolic representation of an image. Some ideas for such a description might be

- Segmented image, parsed as labelled graph.
- 3D CAD-like model.
- Verbal description of a scene.

However none of the above have proved easy to define, feasible to produce or universally useful. It now seems that there is no single generic description whose extraction could be considered to be “solving the vision problem”.

More general definition needs to include the notion of functionality: Computer vision is the process of using images in a machine system in order to perform some task. For example,

- Video stabilization.
- Obstacle avoidance by a robot.
- Automated assembly.
- Talking with user about what is in a scene.

Task may be implicit, but since it determines what sort of a representation is produced from the imagery, it is important to keep it in focus.

Each task involves the generation of different representations. Some tasks probably need multiple representations.

Certain representations are likely to be useful for different tasks, so we can reasonably hope for reusability; just not complete generality.

Why Study Computer Vision?

A. Practical application areas

- Automated inspection.
- Optical character recognition
- Biomedical applications.
- Remote sensing.
- Automated assembly.
- Autonomous vehicles.
- Video surveillance.
- Image/video databases.
- Human-computer interfaces.
- Aids to handicapped.
- Population of virtual environments.
- And others...

B. Scientific interest

- Abstract problem: how is it possible to do what people/animals do using their eyes?
- Formalization: in fact, just what is it that people and animals do?
- Gain insight into possible biological mechanisms.
- Test models of biological mechanisms.
- What can be done with images that people/animals do not do?
- Gain insight into fundamental metaphor of human existence.

Why is Computer Vision Difficult

A. Better question is: why would anyone think it would be easy?

- Recognition seems entirely effortless.
- Compared to (say) chess.
- Huge amount of work done at subconscious level.

B. Only verbal interface is available introspectively

- Provides little insight into low-level mechanisms.
- Can be misleading as explanations tend to be post-hoc rationalization.
- Lots of visual functionality seems to bypass the language mechanisms almost entirely.

C. Neurophysiological evidence is intriguing, but still very sketchy.

- Behavior of first few levels of cells is interperable as low-level enhancement and edge/motion/feature detectors.
- Unclear if this description is complete or the best way of looking at what is going on.
- Higher levels are much less clear. Some evidence for cells that fire when specific, high-level objects, e.g. faces are seen.
- Details of connections between low-level and higher level function are very murky.

D. Most of what people do is probably very resource intensive by the computational standards of the 60s, 70s, 80s, 90s, 00s, ...

- In the late 90s, 10^9 OPS/bytes is a very big desktop.
- In the mid 00s, a few times 10^9 OPS/bytes is an average desktop.
- The largest existing machines have a few times 10^{13} OPS/bytes
- Estimates of effective human memory capacity range from 10^{12} to 10^{15} bytes.
- Estimates of effective human computational resources range from 10^{14} to 10^{17} OPS.

How to Study Machine Vision

A. Study specific tasks for which vision provides leverage. Solve them in an end to end manner, picking simpler problems until you get ones you can do something with.

B. Look for commonalities in solution methods. These are useful general principles. Attempt to relate problem types and solution methods. See if it can be done in a formal manner.

C. Think about issues abstractly, but don't expect too much from the exercise until you have a lot of working examples to ground the process.

D. Sources of Problems

- Biology
  - Obstacle avoidance.
  - Homing/navigation.
  - Using/dealing with motion.
  - Detection/interaction with food, predators, other individuals.
- People
  - Reading.
  - Face detection/recognition.
  - Expression/gesture interpretation.
  - Physical interaction with objects (manipulation/assembly).
  - Object Recognition.
  - Verbal description of pictures.
  - Interpretation of activity.
  - etc...

- Technology
  - Biomedical applications.
  - Interpretation of satellite/aerial data.
  - Indexing in image databases.
  - Shape recovery.
  - Object/scene modeling.
  - CAD/CAM
- New sensors: There are many sources of image-like data that have no direct analog in people or even animals. There may also be useful ways of processing data that correspond to no known biological function.
  - IR, UV, other bands.
  - Hyperspectral images.
  - Polarization.
  - Sonar, ultrasound.
  - Depth sensors.
  - Synthetic aperture radar (SAR).
  - Tomographic techniques.

Course Outline

A. Mathematical and physical preliminaries.

- Theory and practice of image representation.
- Theory and practice of image formation and capture.
- Physics of the interaction of light with matter.
- Notions of invariance and robustness.

B. Image processing basics

- Define classic techniques that map images to images.
- Describe what they do, and look forward to situations in which they will be useful.

C. Biological Vision

- Structure and capabilities of the eye
- Neural processing basics
- Visual pathways in the brain

D. The meat of the course

- Visual tasks.
- Successful and unsuccessful approaches.
- Generally useful techniques.
- Organized by technique, but we will pull in motivating applications at each stage.

Prototypical tasks

- Inspection
- Scene labeling
- Shape reconstruction
- Object recognition and detection
- Tracking
- Navigation
- Manipulation
- Model acquisition

General techniques (we may hit some others as well)

- Local feature detection
- Segmentation and extended feature detection
- Matching and classification methods
- Decision theory
- 2D and 3D shape representations
- Representations of appearance
- Color
- Texture
- Stereo
- Motion analysis