Lecture 1: Computer Vision and Image Processing

Introduction

Saad J Bedros, PhD
Office: 105D Walter Library
Phone: (612) 626-3421
Email: sbedros@umn.edu
Office Hours W,F 1:15 PM : 2:00 PM
2. E. Trucco and A. Verri: *Introductory Techniques for 3-D Computer Vision*
General Comments about the Class

- Prerequisites—these are essential!
  - Vector calculus
  - Statistics
  - Linear algebra
- Course does not assume prior imaging experience
  - computer vision, image processing, graphics, etc.
- Emphasis on programming projects!
- Collaboration Policy
  - Assignments may be discussed, but all written work and coding must be done individually. Mini project may not be discussed. Individuals found submitting duplicate or substantially similar materials due to inappropriate collaboration may get an “F” in this class and there may be other serious consequences.
- Matlab will be used for Homeworks and Projects
# Course Syllabus: Vision Part

<table>
<thead>
<tr>
<th>Date</th>
<th>Month</th>
<th>Day</th>
<th>Lecture No</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>19 Mar</td>
<td>March</td>
<td>29</td>
<td>W</td>
<td>Intro, syllabus, robot overview.</td>
</tr>
<tr>
<td>20 Mar</td>
<td>March</td>
<td>31</td>
<td>F</td>
<td><strong>Guest Lecture on Robotics</strong></td>
</tr>
<tr>
<td>21 Apr</td>
<td>Apr</td>
<td>05</td>
<td>W</td>
<td>1- Introduction to computer vision and image processing: a review.</td>
</tr>
<tr>
<td>22 Apr</td>
<td>Apr</td>
<td>07</td>
<td>F</td>
<td>2- Image formation, Camera fundamentals.</td>
</tr>
<tr>
<td>23 Apr</td>
<td>Apr</td>
<td>12</td>
<td>W</td>
<td>3- Digital Image Representation and Color Fundamentals</td>
</tr>
<tr>
<td>24 Apr</td>
<td>Apr</td>
<td>14</td>
<td>W</td>
<td>QUIZ – NO Lecture</td>
</tr>
<tr>
<td>25 Apr</td>
<td>Apr</td>
<td>19</td>
<td>F</td>
<td>4- Spatial Domain Transformations: Point, Area/Mask, Geometric and Frame processing Transformations. Image coordinate transformations</td>
</tr>
<tr>
<td>26 Apr</td>
<td>Apr</td>
<td>21</td>
<td>W</td>
<td>5- Frequency Domain Transformations</td>
</tr>
<tr>
<td>27 Apr</td>
<td>Apr</td>
<td>26</td>
<td>F</td>
<td>6- Edge detection continued</td>
</tr>
<tr>
<td>28 Apr</td>
<td>Apr</td>
<td>28</td>
<td>W</td>
<td>7- Most Commong Edge detection – Canny and Sobel Edge Detectors</td>
</tr>
<tr>
<td>29 May</td>
<td>May</td>
<td>03</td>
<td>F</td>
<td>8- Corner/line detection and Hough transforms</td>
</tr>
<tr>
<td>30 May</td>
<td>May</td>
<td>05</td>
<td>W</td>
<td>9- Generalized Hough transform. Image Thresholding methods</td>
</tr>
</tbody>
</table>

**ME5286 – Lecture 1 (Theory)**
About Computer Vision

• Deals with the formation, analysis and Interpretation Images
• Is an integral part of Artificial Intelligence (AI)
• Is interdisciplinary subject area
• Is Practical and useful
• Is modern, challenging and continuously evolving
Computer Vision Difficulties

- Images are ambiguous: Projective
- Images are affected by many factors
  - Sensor model
  - Illumination
  - Shape of object(s)
  - Color of object(s)
  - Texture of object(s)
- There is no “Universal Solution” to the Vision problem
- There are many theories
- We do not understand how the human Visual System works
- Computer technology changes continuously, Hence Computer Vision is evolving.
Why is Computer Vision Difficult?

• It is a many-to-one mapping
  – A variety of surfaces with different material and geometrical properties, possibly under different lighting conditions, could lead to identical images
  – Inverse mapping is under-constrained – non-unique solution (a lot of information is lost in the transformation from the 3D world to the 2D image)

• It is computationally intensive

• We still do not understand the recognition problem
Illusions
What do humans see?
Nomenclature

Somewhat interchangeable names, with somewhat different implications:

- **Computer Vision**
  - Most general term

- **Computational Vision**
  - Includes modeling of biological vision

- **Image Understanding or Scene Understanding**
  - Automated scene analysis (e.g., satellite images, robot navigation)

- **Machine Vision**
  - Industrial, factory-floor systems for inspection, measurements, part placement, etc.
Related Fields

• Largely built upon
  – Image Processing
  – Statistical Pattern Recognition, Machine Learning
  – Artificial Intelligence

• Related areas
  – Robotics
  – Biological vision
  – Medical imaging
  – Computer graphics
  – Human-computer interaction
Brief History of Computer Vision

• B.C. (Before Computers)
  – Philosophy
  – Optics
  – Psychophysics
  – Neurophysiology

• Psychophysics and neurophysiology describe the behavior of cells or subjects, but do not explain such behavior
  – Black box – how can we figure it out?
Brief History of Computer Vision

• 1960s – First computer vision programs/systems
• Minsky at MIT – Solve vision in a summer
  – Not really…
• Empirical approaches
  – Ad hoc, “bag of tricks”
  – Image processing plus…
• Simplified worlds
  – “Blocks world” then generalize
  – AI approach
Brief History of Computer Vision

• Computational
  – Attempt at a rigorous study based on information processing model (David Marr)

• Task-oriented
  – Practical systems to accomplish specific tasks

• Currently
  – All of this (and more)
  – Limited progress
    • However, we have a much deeper understanding of the issues (and of the difficulty)
  – Still no “killer app” but getting closer
Progress in Computer Vision

• First generation: Military/Early Research
  – Few systems, each custom-built, cost $Ms
  – “Users” have PhDs
  – 1 hour per frame

• Second generation: Industrial/Medical
  – Numerous systems, 1000s of each, cost $10Ks
  – Users have college degree
  – Special hardware

• Third generation: Consumer
  – 100000(00) systems, cost $100s
  – Users have little or no training
  – Emphasis on software
The Three Processing Levels

• Low-level processing
  – Standard procedures are applied to improve image quality
  – No “intelligent” capabilities
The Three Processing Levels

- Intermediate-level processing
  - Extract and characterize components in the image
  - Some intelligent capabilities are required
The Three Processing Levels

- High-level processing
  - Recognition and interpretation
  - Procedures require high intelligent capabilities
Computer Vision

• Some applications
  – **Robotics**
    • Navigation, object manipulation, interaction with humans…
  – Inspection, measurement
  – **Medical imaging**
  – Graphics and animation, special effects
  – Multimedia database indexing and retrieval
  – Human-computer interaction
  – **Surveillance and security**
    • Face and Iris Recognition
Computer Vision

Output:

Model

Real Scene

Cameras

Images

ME5286 – Lecture 1 (Theory)
Computer Graphics

Output:

Image

Synthetic Camera

Model

Projection, shading, lighting models
Another View

Interesting

Computer vision

Actually works

Computational vision

Machine vision
Why study Computer Vision?

Interpretation of Visual Information

• Images and movies are everywhere
  – Mobile phone, cheaper cameras enabling the imaging capture
• Fast-growing collection of useful applications
  – Face finding, recognition, analysis
  – building representations of the 3D world from pictures
  – automated surveillance (who’s doing what)
  – movie post-processing
• Various deep and attractive scientific mysteries
  – how does object recognition work?
• Greater understanding of human vision
Every picture tells a story

- Goal of computer vision is to enable machines analyze and interpret images
Part I: The Physics of Imaging

• How images are formed
  – Cameras
    • How a camera creates an image
    • How to tell where the camera was
  – Light
    • How to measure light
    • What light does at surfaces
    • How the brightness values we see in cameras are determined
  – Color
    • The underlying mechanisms of color
    • How to describe it and measure it
Part II: Early Vision in One Image

• Representing small patches of image
  – For three reasons
  • We wish to establish correspondence between (say) points in different images, so we need to describe the neighborhood of the points
  • Sharp changes are important in practice --- known as “edges”
  • Representing texture by giving some statistics of the different kinds of small patch present in the texture.
    – Tigers have lots of bars, few spots
    – Leopards are the other way
Representing an image patch

• Filter outputs
  – essentially form a dot-product between a pattern and an image, while shifting the pattern across the image
  – strong response -> image locally looks like the pattern
  – e.g. derivatives measured by filtering with a kernel that looks like a big derivative (bright bar next to dark bar)
Texture

• Many objects are distinguished by their texture
  – Tigers, cheetahs, grass, trees
• We represent texture with statistics of filter outputs
  – For tigers, bar filters at a coarse scale respond strongly
  – For cheetahs, spots at the same scale
  – For grass, long narrow bars
  – For the leaves of trees, extended spots
• Objects with different textures can be segmented
• The variation in textures is a cue to shape
Part III: Early Vision in Multiple Images

• The geometry of multiple views
  – Where could it appear in camera 2 (3, etc.) given it was here in 1 (1 and 2, etc.)?

• Stereopsis
  – What we know about the world from having 2 eyes

• Structure from motion
  – What we know about the world from having many eyes
    • or, more commonly, our eyes moving.
Part IV: Mid-Level Vision

• Finding coherent structure so as to break the image or movie into big units
  – Segmentation:
    • Breaking images and videos into useful pieces
    • E.g. finding image components that are coherent in internal appearance
  – Tracking:
    • Keeping track of a moving object through a long sequence of views
Part V: High Level Vision (Geometry)

• The relations between object geometry and image geometry
  – Model based vision
    • find the position and orientation of known objects
  – Smooth surfaces and outlines
    • how the outline of a curved object is formed, and what it looks like
  – Aspect graphs
    • how the outline of a curved object moves around as you view it from different directions
  – Range data
Part VI: High Level Vision (Probabilistic)

• Using classifiers and probability to match and recognize objects
  – Templates and classifiers
    • how to find objects that look the same from view to view with a classifier
  – Relations
    • break up objects into big, simple parts, find the parts with a classifier, and then reason about the relationships between the parts to find the object.
  – Geometric templates from spatial relations
    • extend this trick so that templates are formed from relations between much smaller parts
Examples of Commercial Vision Systems

High Precision Inspection System

Par Systems: Vision Guided Robotics for manufacturing and assembly lines

NASA Rover

There are many products that have an element of computer vision, and the number is increasing every day
Security Applications

Face Recognition:
  • Face Detection
  • Face Labeling
  • Face Finding

Can you Name any Company Using Face Applications?
Face Detection

Challenges: Pose and Illumination variations
Face Recognition
Multi Biometrics
Vision-Based Interfaces

Intel and Microsoft
New 3D Sensors

Gesture Recognition

3D Face Analysis and Recognition
Visual Inspection
Indexing into Databases

• Shape content

A query object leads to intuitive results
The instance found, corresponding parts
Application: Optical Character Recognition

Select conversion options
Only converted files can be edited online. File types and size limits

Tick box: Convert documents, presentations and spreadsheets to the corresponding Google Docs formats

Tick box: Convert text from PDF or image files to Google Docs documents

Instead of letting the concussion heal by staying in bed, Paddy got up the next morning and began digging potatoes. In stooping over the potatoes, he hit his head and suddenly became blind. As his blindness continued to worsen, he was taken ill with "Bronchiitis," and eventually passed away while awaiting a death certificate. Seven days after his death, the priest who had called the priest to hear the last rites of the deceased came to Paddy's funeral. He asked Paddy's widow whether she had written Paddy's death in the ledger. She corrected him. Paddy had been dead for seven days, she said, but the priest had only heard about it the day after Paddy's funeral.

One month later on February 28, Paddy's widow received a letter from the priest. It read:

"Beloved, I am writing to inform you that Paddy has been declared dead. He was last seen alive on January 28, 1945. The priest who performed the last rites has been arrested. He is charged with murder."

License Plate Recognition
Indexing into Databases

- Color, texture

\[ T = 33.6\text{s}, \text{found 2 of 2} \]
Target Recognition

- Department of Defense (Army, Airforce, Navy)
Interpretation of Aerial Photography

ME5286 – Lecture 1 (Theory)
Traffic Monitoring
Applications: Robotics
Types of images

- infra-red
- ultra-violet
- radio-waves (radio astronomy)
- visible light
- micro-waves (radar)
- roentgen (tomography)
- sound-waves (echoscopy, sonar)
- electrons (microscopy)
- positron emission (PET-scan)
- magnetic resonance (NMR)
Electromagnetic (EM) Spectrum
Electromagnetic (EM) Spectrum

- Gamma rays: $10^{-3}\text{nm}$
- X-rays: $10^{-1}\text{nm}$
- Ultraviolet: $10^{1}\text{nm}$
- Visible Spectrum: $400\text{nm} - 700\text{nm}$
- Infrared: $10^{3}\text{nm}$
- Radio waves: $10^{9}-10^{13}\text{nm}$

Shorter wavelength on the left, Longer wavelength on the right.

Colors:
- Blue: $436\text{nm}$
- Green: $546\text{nm}$
- Red: $700\text{nm}$
Imaging in Different Wavelengths
Face and Face Feature Localization

ME5286 – Lecture 1 (Theory)
Thermal Imaging
Camera Network Physical Design

- color cameras
- medium resolution
- smaller Field of Views

---

camera symbol

\[60\]  Field of View symbol
Statistical Modeling of Image Pixels
Continuous Learning of Foreground/Background pair
Outdoor Motion Detection
Video Surveillance: Tracking
Video Surveillance: Tracking
Human Tracking and Pan/Tilt Control
Video Tracking: Multiple PTZ Cameras
People Detection and Tracking

Online Pedestrian Tracking with Motion Agreement
Human Activity Recognition

Person 0: Periodic Motion, Carry Object
Person 1: Periodic Motion, Does Not Carry Object

Person 0: Interaction with Person 1
Person 1: Interaction with Person 0

Person 0: Periodic Motion, Does Not Carry Object
Person 1: Periodic Motion, Carry Object

ME5286 – Lecture 1 (Theory)
Effective Command & Control

Security Today:
- Coverage areas are enormous
- Too many images - it is difficult to interpret activities

ME5286 – Lecture 1 (Theory)
Command & Control

- Integrate hundreds of sensors in a unified view
- Move seamlessly in space - god’s eye view to up close
- Move seamlessly in time - replay and rehearse
- Alert operators to key events
- ID individuals
- Track and overlay
- Record, compress and store
- Archive and data mine
Next Generation Browser “Fly-Through” with Real-Time Video

Project video on a graphics model – with ability to render from any view
3D Tracking and Registration
Smart City
Can computers match human perception?

• Not yet
  – computer vision is still no match for human perception
  – but catching up, particularly in certain areas
Computer Vision Representations

Image(s)
- Color
- Shape
- Texture
- Shading
- Motion
- Stereo

Surfaces
"2 1/2-D Sketch"

Objects

Iconic → Segmented → Geometric → Relational
Autonomous Vehicles

- Land, Underwater, Space
Grand Challenge 2005: 195 Teams

Not shown: Autonomy, West Coast Robotics and UBC-CERMS Team Thunderbird – Canada, Grand Challenge New Zealand, Dobot Team – France.
Urban Darpa Challenge 2007

Autonomous Driving in Traffic
The DRC is a competition of robot systems and software teams vying to develop robots capable of assisting humans in responding to natural and man-made
Autonomous Navigation

Auto X
Conclusion

• Computer Vision is an exciting and challenging field
• Applied in many real world solutions
• Tremendous progress in the last decades
• Still have long way to go …