CS595: Introduction to Computer Vision

Instructor: Qi Li

Course syllabus

- Instructor
  - E-mail: qi.li@cs.wku.edu
  - Office: TCCW 135
  - Office hours
    - MW: 9:00-10:00, 15:00-16:00
    - T: 9:00-12:00, 14:00-16:00
    - F: 9:00-10:00
    - By appointment
Course syllabus

- **Textbook**
  - *Computer Vision* by Linda Shapiro and George Stockman

- **Grading**
  - 15%: Midterm exam (Mar, 5)
  - 30%: Final exam (May, 6)
  - 50%: Projects
  - 5%: Attendance

Course syllabus

- **Background:**
  - Linear algebra
  - Calculus

- **Programming languages:**
  - Mathematica
  - Additional: C/C++, Java
Course description

An introduction to the analysis of images and video in order to recognize, reconstruct and model objects in the three-dimensional world.

We will study the geometry of image formation; basic concepts in image processing such as smoothing, edge and feature detection, color, and texture; motion estimation; segmentation; stereo vision; 3-D modeling; and statistical recognition.

Outline

• What is Vision?
• Course outline
• Applications
The Vision Problem

How to infer salient properties of 3-D world from time-varying 2-D image projection

Computer Vision Outline

• Image formation
• Low-level
  – Single image processing
  – Multiple views
• Mid-level
  – Estimation, segmentation
• High-level
  – Recognition
Image Formation

- 3-D geometry
- Physics of light
- Camera properties
  - Focal length
  - Distortion
- Sampling issues
  - Spatial
  - Temporal

Low-level: Single Image Processing

- Filtering
  - Edge
  - Color
  - Local pattern similarity
- Texture
  - Appearance characterization from the statistics of applying multiple filters
- 3-D structure estimation from...
  - Shading
  - Texture
Low-level: Multiple Views

• Stereo
  – Structure from two views
• Structure from motion
  – What can we learn in general from many views, whether they were taken simultaneously or sequentially?

Mid-Level: Estimation, Segmentation

• Estimation: Fitting parameters to data
  – Static (e.g., shape)
  – Dynamic (e.g., tracking)
• Segmentation/clustering
  – Breaking an image or image sequence into a few meaningful pieces with internal similarity
High-level: Recognition

• Recognition:
  – Finding and parametrizing a known object
  – Assignment to known categories using statistics/probability to make best choice

Applications

• Inspection
  – Factory monitoring: Analyze components for deviations
  – Character recognition for mail delivery, scanning
• Biometrics (face recognition, etc.), surveillance
• Image databases: Image search on Google, etc.
• Medicine
  – Segmentation for radiology
  – Motion capture for gait analysis
• Entertainment
  – 1st down line in football, virtual advertising
  – Matchmove, rotoscoping in movies
  – Motion capture for movies, video games
• Architecture, archaeology: Image-based modeling, etc.
• Robot vision
  – Obstacle avoidance, object recognition
  – Motion compensation/image stabilization
Applications: Factory Inspection

Cognex’s “CapInspect” system

courtesy of H. Shum and R. Szeliski

Applications: Mosaicing

courtesy of H. Shum and R. Szeliski
Applications: Face Detection
courtesy of H. Rowley

Applications: Face Recognition
ORL face dataset
Applications: Text Detection & Recognition

Applications: MRI Interpretation

Coronal slice of brain

Segmented white matter
Detection and Recognition: How?

- Build models of the appearance characteristics (color, texture, etc.) of all objects of interest
- Detection: Look for areas of image with similar appearance to a targeting category of objects
- Recognition: Decide the identity of an object
- Segmentation: Categorize every pixel

Applications:
Football First-Down Line

courtesy of Sportvision
Applications: Virtual Advertising

First-Down Line, Virtual Advertising: How?

- Sensors that measure pan, tilt, zoom and focus are attached to calibrated cameras at surveyed positions.
- Knowledge of the 3-D position of the line, advertising rectangle, etc. can be directly translated into where in the image it should appear for a given camera.
- The part of the image where the graphic is to be inserted is examined for occluding objects like the ball, players, and so on. These are recognized by being a sufficiently different color from the background at that point. This allows pixel-by-pixel compositing.
Applications: Inserting Computer Graphics with a Moving Camera

Opening titles from the movie “Panic Room”

Applications: Inserting Computer Graphics with a Moving Camera

courtesy of 2d3
CG Insertion with a Moving Camera: How?

- This technique is often called *matchmove*.
- Once again, we need camera calibration, but also information on how the camera is moving—its *egomotion*. This allows the CG object to correctly move with the real scene, even if we don’t know the 3-D parameters of that scene.
- Estimating camera motion:
  - Much simpler if we know camera is moving sideways (e.g., some of the “Panic Room” shots), because then the problem is only 2-D.
  - For general motions: By identifying and following scene features over the entire length of the shot, we can solve retrospectively for what 3-D camera motion would be consistent with their 2-D image tracks. Must also make sure to ignore independently moving objects like cars and people.

Applications: Rotoscoping

2d3’s Pixeldust
Applications: Motion Capture

Vicon software:
12 cameras, 41 markers for body capture;
6 zoom cameras, 30 markers for face

Applications: Motion Capture without Markers

courtesy of C. Bregler
Motion Capture: How?

- Similar to matchmove in that we follow features and estimate underlying motion that explains their tracks
- Difference is that the motion is not of the camera but rather of the subject (though camera could be moving, too)
  - Face/arm/person has more degrees of freedom than camera flying through space, but still constrained
- Special markers make feature identification and tracking considerably easier
- Multiple cameras gather more information

Applications: Image-Based Modeling

Façade project: UC Berkeley Campanile
Image-Based Modeling: How?

• 3-D model constructed from manually-selected line correspondences in images from multiple calibrated cameras
• Novel views generated by texture-mapping selected images onto model

Applications: Robotics

Autonomous driving: Lane & vehicle tracking (with radar)