Computer Vision

Dr. Zoran Duric

Department of Computer Science
George Mason University

Office: Nguyen Engineering Building 4443
Email: zduric@cs.gmu.edu
URL: http://www.cs.gmu.edu/∼zduric/
URL: http://www.cs.gmu.edu/∼vislab/

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

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”

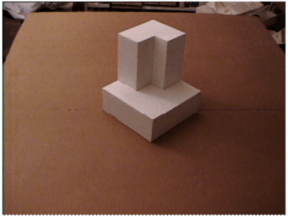
but they do not see alike

Vision is deceivingly easy

The MIT 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 including support relationships

Formally ended in 1985



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

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 0.1% 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
- Ultrasound vision
- X-ray vision!
- RADAR vision
- Lidar 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 haven't our eyes evolved to see into 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
- Microsoft Kinect

Vision is critical to many applications in

- Manufacturing
- Communications
  - Video teleconferencing
- Medicine
- Transportation
  - Self-driving cars
  - Warning drowsy drivers, obstacle detection/avoidance
  - Traffic monitoring, license plate recognition
- Entertainment
  - Microsoft Kinect
  - Virtual and Augmented Reality applications
- Agriculture
- Defense & Security
  - Many R&D programs at DoD
  - Video Surveillance

- Idea and motivation:
  - Process video sequences and make them more useful, intuitive, interesting
  - Allow user to quickly browse large video data
  - Allow user to get summaries of objects of interest
- Challenge:
  - Video is highly redundant
  - Decisions are based on the content and relative importance of parts of the video
  - Large intervals have no activity, or have activity that occur in a small image region
- Representative work: Video Synopsis [49]

Video Synopsis and Indexing

Video Summarizations

To generate and visualize summarizations, need:

- Information about scene
- List of objects
- Information about objects
- Images and trajectories of objects

A surveillance video

Information about the scene: Ground plane calibration

Ground plane calibration is achieved by computing a *homography* between the scene and corresponding location in Google Earth.

Image of the scene

Google Earth view of the scene

$$h_{11}x + h_{12}y + h_{13} - h_{31}xX - h_{32}yX - h_{33}X = 0$$

$$h_{21}x + h_{22}y + h_{23} - h_{31}xY - h_{32}yY - h_{33}Y = 0$$

$$A_i = \begin{pmatrix} -x & -y & -1 & 0 & 0 & 0 & Xx & Xy & X \\ 0 & 0 & 0 & -x & -y & -1 & Yx & Yy & Y \end{pmatrix}$$

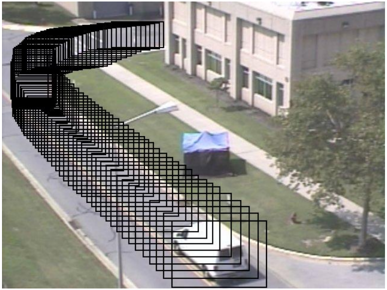
$$h = \begin{pmatrix} h_{11} & h_{12} & h_{13} & h_{21} & h_{22} & h_{23} & h_{31} & h_{32} & h_{33} \end{pmatrix}^T$$

$$X = \frac{h_{11}x+h_{12}y+h_{13}}{h_{31}x+h_{32}y+h_{33}} \quad Y = \frac{h_{21}x+h_{22}y+h_{23}}{h_{31}x+h_{32}y+h_{33}}$$

x	y	Y (Latitude)	X (Longitude)
261	100	39.028879	-76.965515
207	362	39.028600	-76.965645

List of objects: Detection and tracking

- The results of the tracker are used as an input data for the method



Bounded boxes for tracked object

Object ID
Location of the object
Height of the object's bounding box
Width of the object's bounding box
Number of foreground pixels inside the box

Output of the tracker

- The goal of the foreground objects detection is to identify regions of a frame that contain moving objects
- The goal of the foreground objects tracking is to track detected regions through the sequence

List of objects: Example of detection and tracking

Frame 1970

Frame 2000

Frame 2050

Object ID: 116
Location: X=260, Y=134
Height: 116
Width: 162
Foreground pixels: 11556

Tracker result for frame 1970

Object ID: 116
Location: X=036, Y=330
Height: 78
Width: 106
Foreground pixels: 5108

Tracker result for frame 2000

Object ID: 116
Location: X=160, Y=446
Height: 50
Width: 102
Foreground pixels: 3424

Tracker result for frame 2050

Video summarization – Frames 4000-5000

Frame 4000

Video summarization – Frames 4000-5000

Frame 4000

Frame 4090

Video summarization – Frames 4000-5000

Frame 4000

Frame 4090

Frame 4248

Video summarization – Frames 4000-5000

Frame 4000

Frame 4090

Frame 4248

Frame 4323

Video summarization – Frames 4000-5000

Frame 4000

Frame 4090

Frame 4248

Frame 4323

Frame 4442

Video summarization – Frames 4000-5000


Frame 4000


Frame 4090


Frame 4248


Frame 4323


Frame 4442


Frame 4538

Video summarization – Frames 4000-5000

Frame 4000

Frame 4090

Frame 4248

Frame 4323

Frame 4442

Frame 4538

Frame 4797

Video summarization – Frames 4000-5000

Frame 4000

Frame 4090

Frame 4248

Frame 4323

Frame 4442

Frame 4538

Frame 4797

Frame 4890

Video summarization – Frames 4000-5000

Frame 4000

Frame 4090

Frame 4248

Frame 4323

Frame 4442

Frame 4538

Frame 4797

Frame 4890

Frame 5000

Video synopsis for frames 4000-5000

Video synopsis for frames 4000-5000

Video synopsis for frames 4000-5000

Video synopsis for frames 4000-5000

Video synopsis for frames 4000-5000

Video synopsis for frames 4000-5000

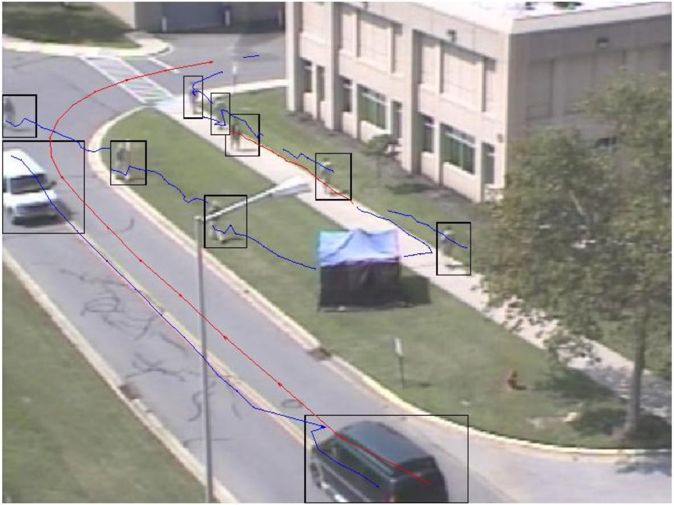
Video synopsis for frames 4000-5000

Video synopsis for frames 4000-5000

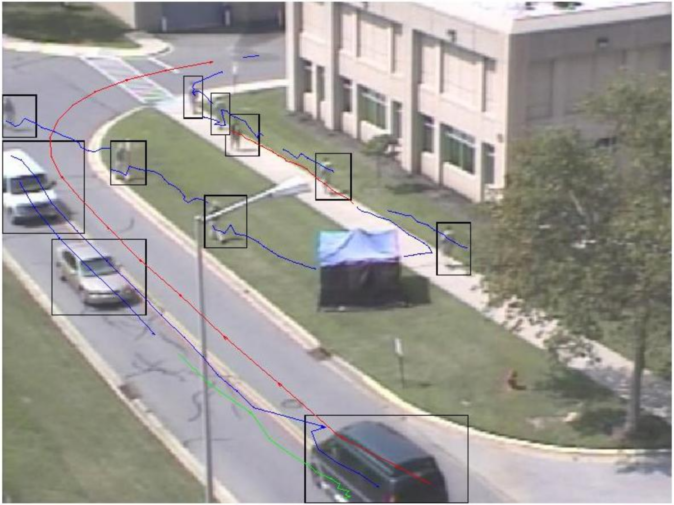
Video synopsis for frames 4000-5000

Video synopsis for frames 4000-5000

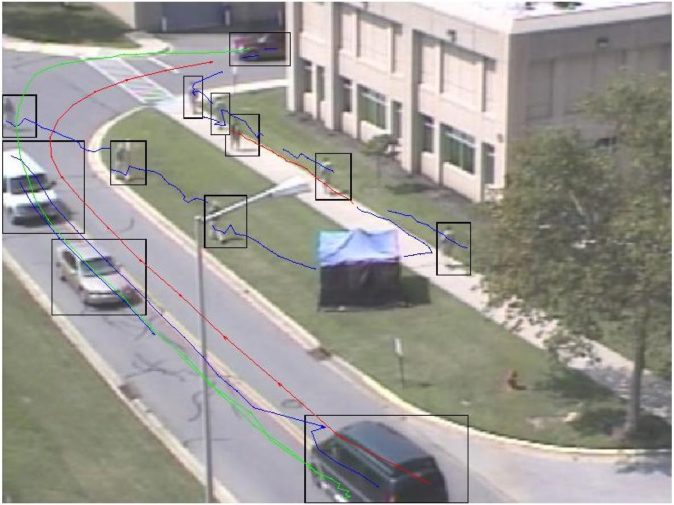
Video synopsis for frames 4000-5000



Video synopsis for frames 4000-5000



Video synopsis for frames 4000-5000



Creating texture mapped 3D models of buildings

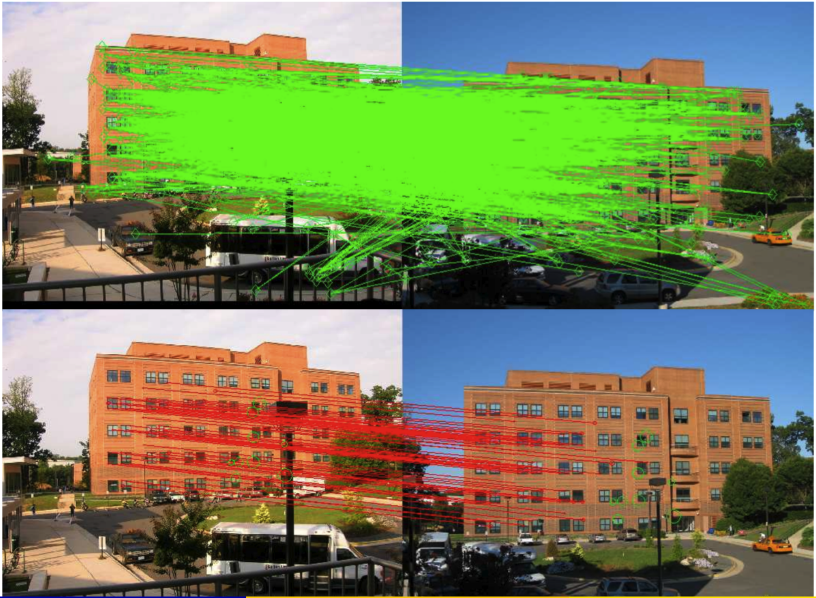
Applications in

- virtual reality
- homeland security
- entertainment
- location recognition
- organizing large image databases
- real estate site selection
- engineering site survey

Detect Distinct Features: Harris Corners

Detect Distinct Features: SIFT Features

Match Features in Different Images



Create 3D Models from Images: Bundle Adjusment

Bundle adjustment Using N views, fit best camera models to all views.