CS 534: Computer Vision

Ahmed Elgammal
Dept of Computer Science
Rutgers University

Robot Readable World - https://vimeo.com/36239715

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 ?

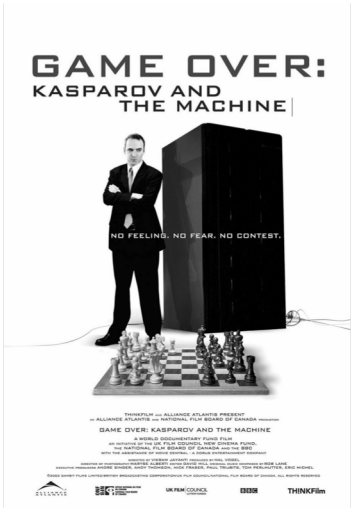
"The plain man's answer (and Aristotle's too) would be, to know what is where by looking. In other words, vision is the process of discovering from images what is present in the world, and where it is *" David Marr (1945-1980), Vision 1982*

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

(Human) Vision is

- Deceivingly easy
- Deceptive
- Computationally demanding
- Critical to many applications



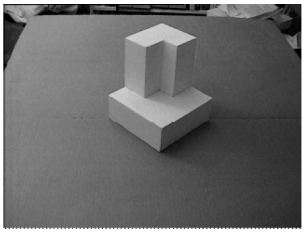
Vision is deceivingly easy

- 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! (vision, vision-touch, vision-motor, vision-navigation, etc.)
- 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 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



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

Vision is deceptive

- 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

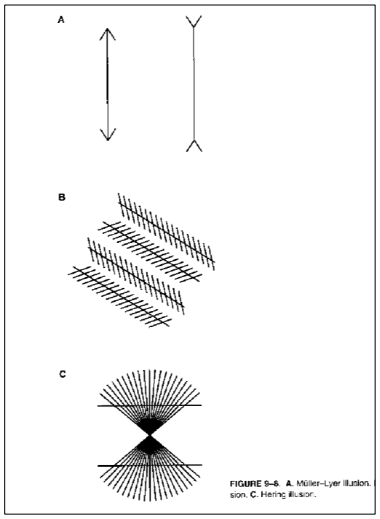
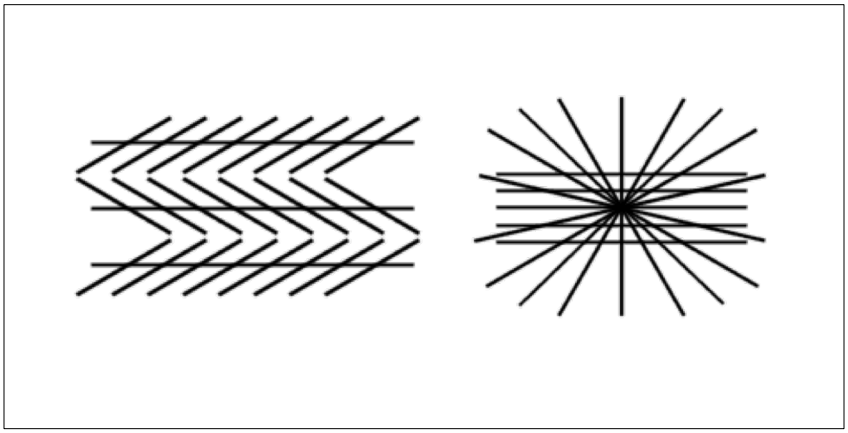


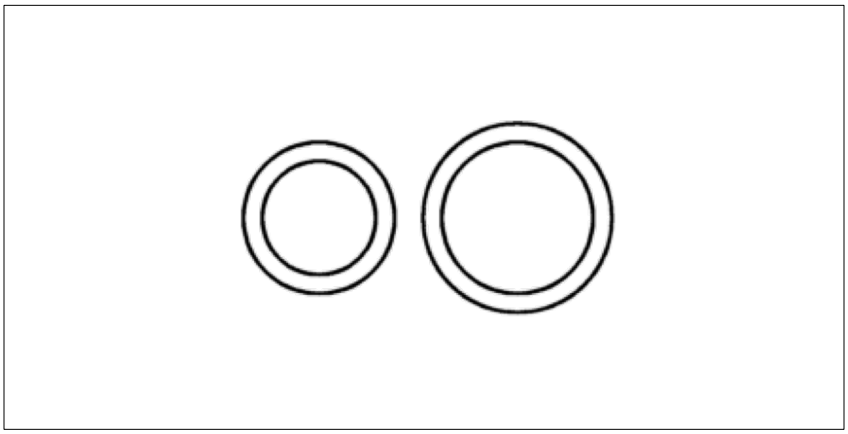
FIGURE 9–8. A. Müller–Lyer illusion. B. Zollner illusion. C. Hering illusion.

Illusions



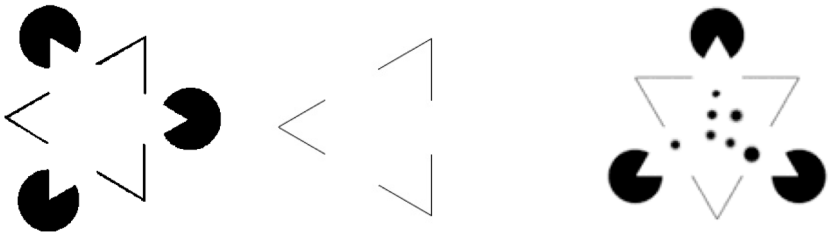
Zollner's illusion - 1860

The background of an image can distort the appearance of straight lines

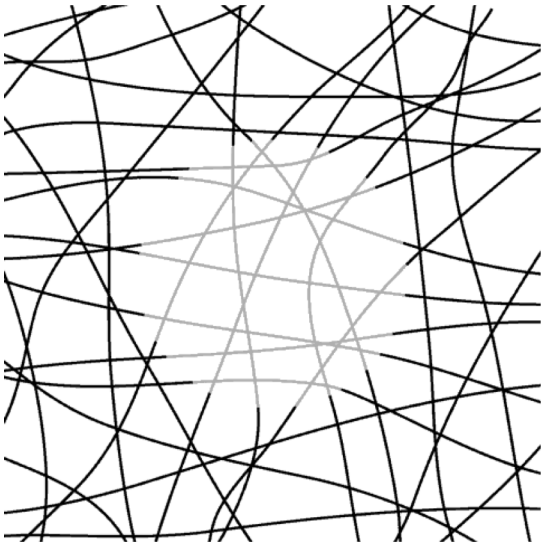


Delboenf's illusion - Context Sensitivity

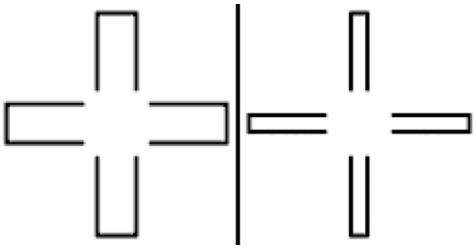
More illusions



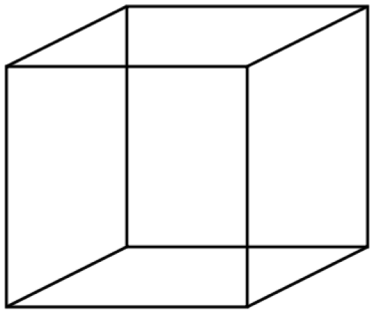
- Subjective contours



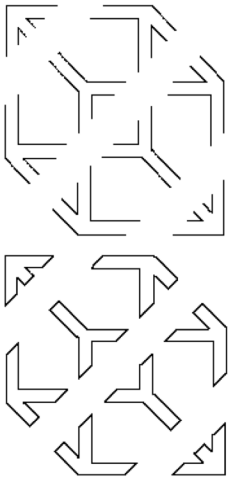
More illusions



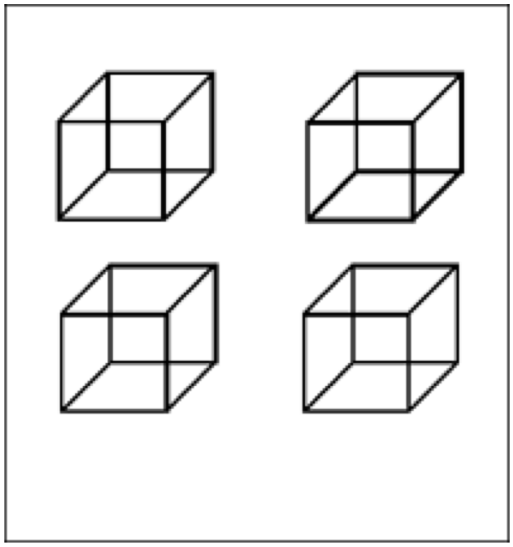
- Subjective contours
- Figure completion



Necker cube: The human visual system picks an interpretation of each part that makes the whole consistent.

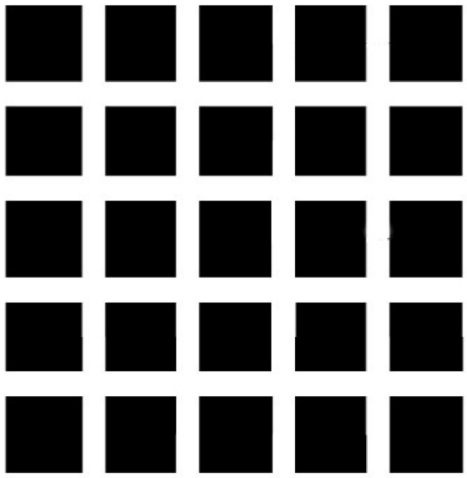


- Subjective contours
- Depth reversibility, Figure completion



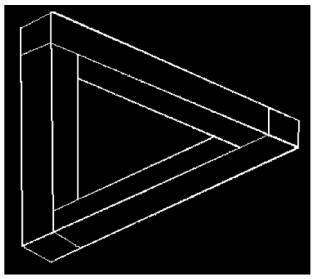
- Depth, reversibility, Do the cubes shift independently or as a unit

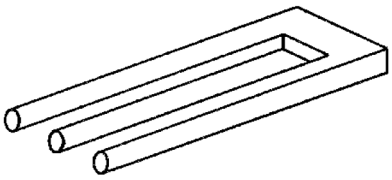
More illusions



- 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 inhibitian

More illusions

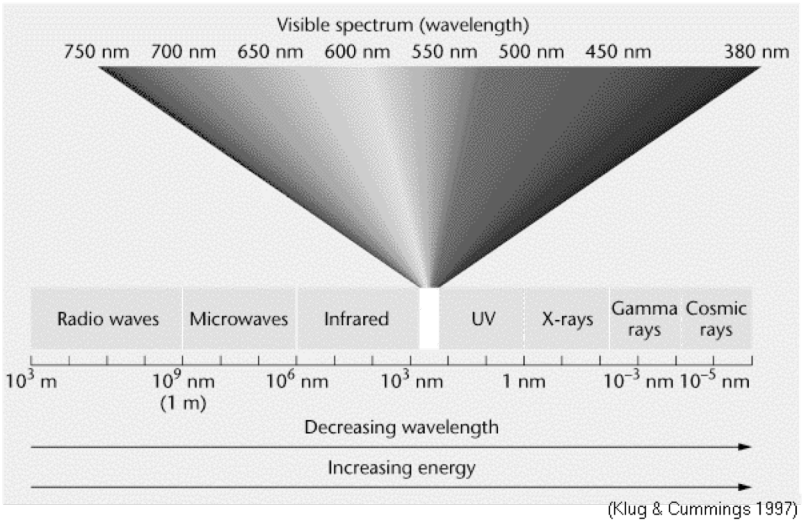




- We can see impossible figures

Spectral limitations of human vision

- We "see" only a small part of the energy spectrum of sunlight
  - we don't see infrared or lower frequencies of light
  - we don't see ultraviolet 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

Example: 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"

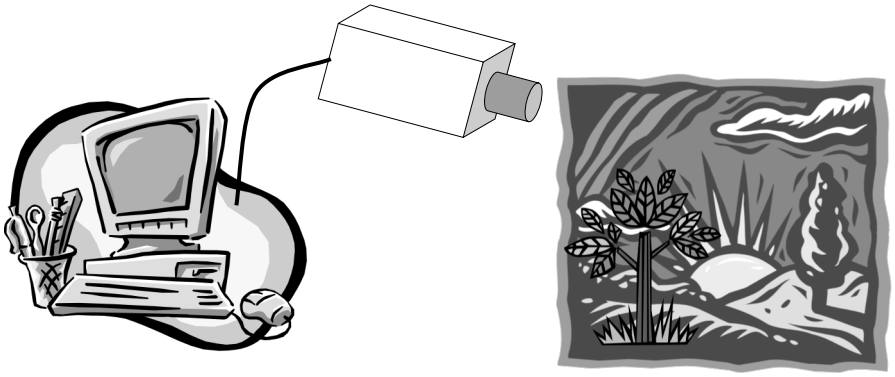
Representation + Processing

"if vision is an information-processing task, then I should be able to make my computer do it, provided that it has sufficient power, memory, and some way of being connected to a home television camera."

We wants to know how to program vision.

Computer Vision

- Understanding the content of images and videos



- Azriel Rosenfeld: Picture Processing By Computer 1969

Picture Processing by Computer

AZRIEL ROSENFELD

University of Maryland, College Park, Maryland

Techniques for processing pictorial information by computer are surveyed. The topics covered include efficient encoding and approximation; position-invariant operations and applications; picture properties useful for pattern recognition; picture segmentation and geometrical properties of picture subsets; picture descriptors and "picture languages."

Key words and phrases: picture processing, image processing, pattern recognition, bandwidth compression, efficient encoding, approximation, position-invariant operation, parallel computers, optical computers, matched filtering, template matching, spatial filtering, image restoration, image enhancement, picture properties, segmentation, geometrical properties, picture description, picture languages

CR categories: 3.63

INTRODUCTION

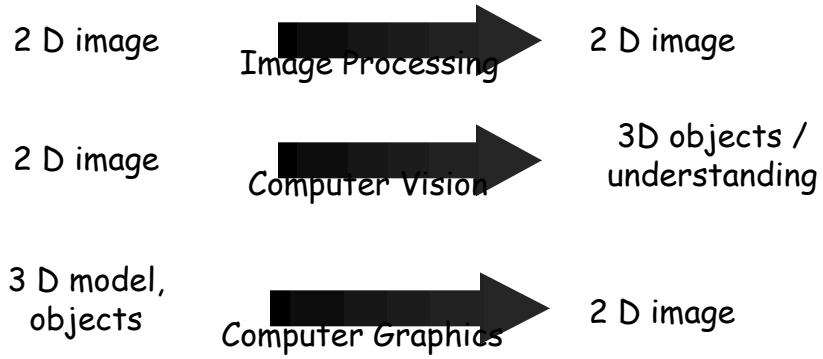
Over the past 15 years, much effort has been devoted to developing methods of processing pictorial information by computer. This work has had a number of different goals, among them television bandwidth compression, image "enhancement" and "restoration," and pictorial pattern recognition. Most of the studies in this field have been directed toward the solution of specific problems involving particular classes of pictures; but a body of "general purpose" picture processing techniques is gradually being built up. In this paper we review the picture-processing field from a technique-oriented standpoint.

We will deal here only with the processing of *given* pictures, not with pictures which have been synthesized by computer; this restriction rules out the areas of computer graphics, computer-generated movies, computer typography, and the like.

As a further restriction, we will be concerned only with pictures as *two-dimensional* objects; this excludes pictorial representation of three spatial dimensions (stereopairs, contour maps, the hidden line problem, etc.), as well as time-varying pictorial information (e.g. on-line character recognition in real time). We will not discuss techniques for computer input or output of pictures; on this subject see, e.g., A. Van Dam, "Computer driven displays and their use in man/machine interaction," *Advances in Computers*, Vol. 7, Academic Press, New York, 1967, pp. 239–290.

We will sometimes regard a *picture* as being a real-valued, nonnegative function of two real variables; the value of this function at a point will be called the *gray level* of the picture at the point. When necessary, we will assume that the function is analytically well-behaved (e.g. that it has an invertible Fourier transform). It will also often be convenient, in connection with picture processing by digital computer, to regard a picture as "discrete," being defined by specifying (say) an $n \times n$ *matrix* of gray levels; in this case the picture will be called a

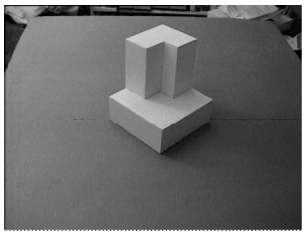
This paper is based on a report of the same title (University of Maryland Computer Science Center Technical Report 68-71) prepared with the support of the Office of Naval Research and the Office of Naval Research under Contract Nonr-4144 (00). An expanded version of the report has been published by Academic Press.



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 vidisector 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.

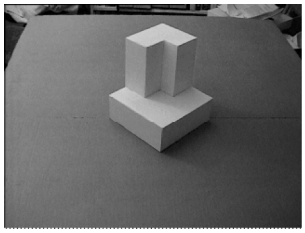
Extensions for August

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*)

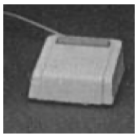
1978 International Association for Pattern Recognition (*IAPR*)

1979 IEEE Transactions on Pattern Analysis and Machine Intelligence (*PAMI*)

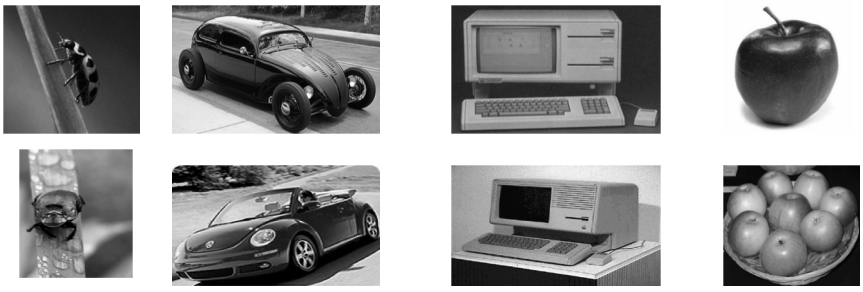
1987 1. International Conference on Computer Vision (*ICCV*)

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.




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 videos are everywhere:

Facebook alone:

- 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
- In 2013: 350M per day = 10.5 B per month

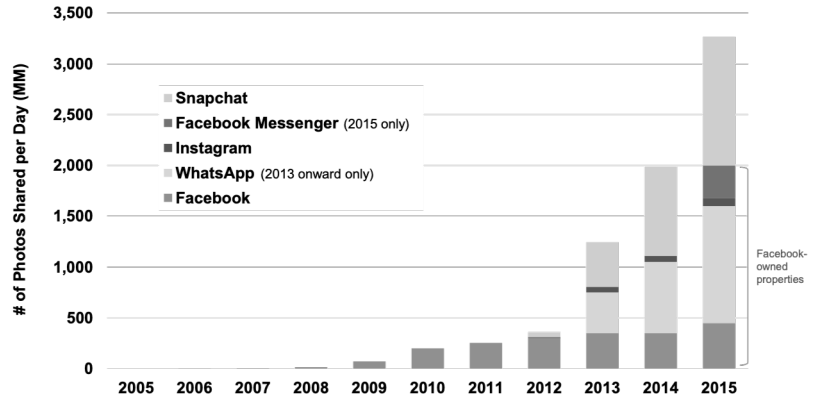
In 2014: In average people uploaded of 1.8 billion digital images every single day. That's 657 billion photos per year

Why to study Computer Vision?

- How many photos are uploaded to social media sites per second (2014 numbers – published in 2015):
  - Snapchat: 200 million users, 8,796 photos per second, 760 million per day
  - Whatsapp: 700 million users, 8,102 photos per second,700 million per day
  - Facebook: 1.39 Billion users, 4,501 photos per second, 350 million per day
  - Instagram: 300 million users, 810 photos per second, 70 million per day
  - Flicker: 12 photos per second, 1 million per day !
  - (2018) Everyday Instagrammers upload an average of 95 million photos.

Source: http://www.adweek.com/socialtimes/how-many-photos-are-uploaded-to-snapchat-every-second/621488

Daily Number of Photos Shared on Select Platforms, Global, 2005 – 2015



@KPCB Source: Snapchat, Company disclosed information, KPCB estimates
Note: Snapchat data includes images and video. Snapchat stories are a compilation of images and video. WhatsApp data estimated based on average of photos shared disclosed in Q1:15 and Q1:16. Instagram data per Instagram press release. Messenger data per Facebook (~9.5B photos per month). Facebook shares ~2B photos per day across Facebook, Instagram, Messenger, and WhatsApp (2015).

KPCB INTERNET TRENDS 2016 | PAGE 90

We are at the golden age for computer vision

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,…
- Videos:
  - Youtube:
  - 2008: 10 hours of videos per minute
  - 2010: 48 hours of video are uploaded every minute, resulting in nearly 8 years of content uploaded every day.
  - In 2011: 100 hours of video per minute
  - Recent 2016: 300 hours of video per minute

2020 statistics

- **3.2 billion images and 720,000 hours of video are shared online daily**

Source: https://theconversation.com/3-2-billion-images-and-720-000-hours-of-video-are-shared-online-daily-can-you-sort-real-from-fake-148630

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

- Manufacturing – machine vision: Visual inspection for quality control, Visual control of robots in assembly lines.
- 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

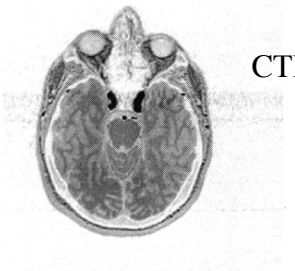
Communications

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

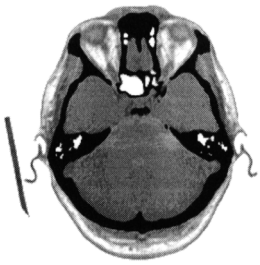
Medicine

- 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

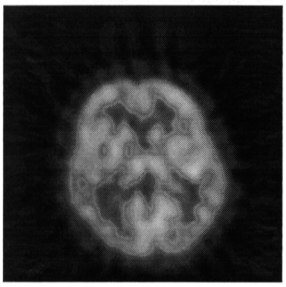

MRI

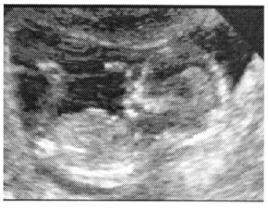


CTI



NMI

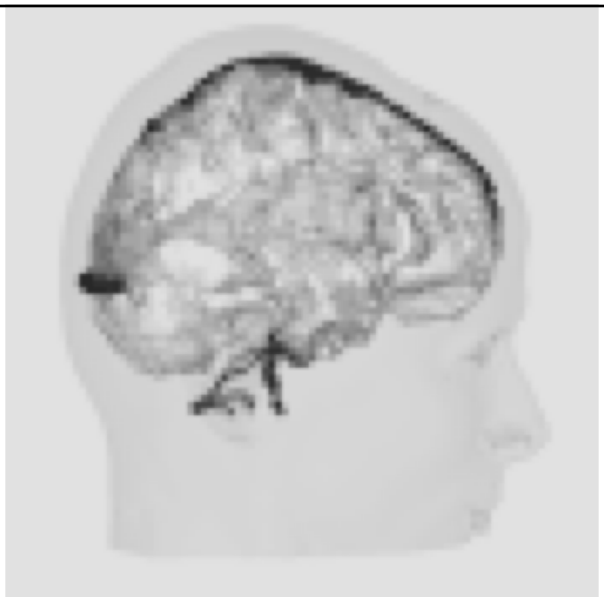




USI

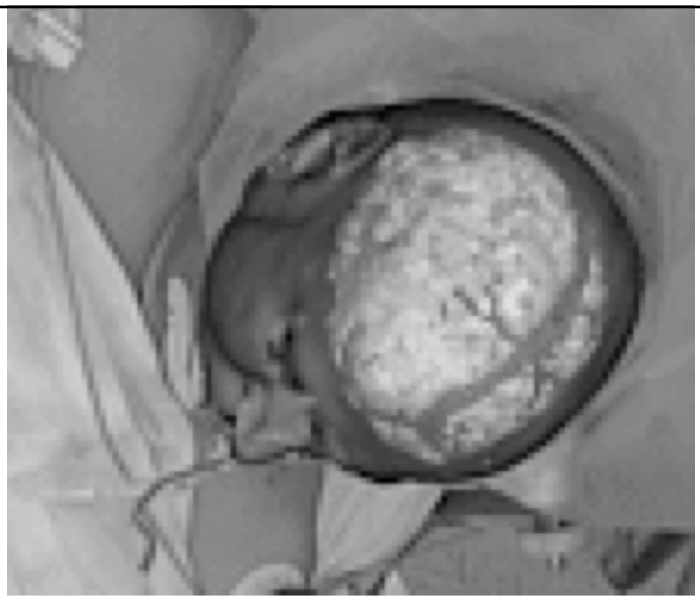
Reprinted from Image and Vision Computing, v. 13, N. Ayache, "Medical computer vision, virtual reality and robotics", Page 296, copyright, (1995), with permission from Elsevier Science

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.

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.

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

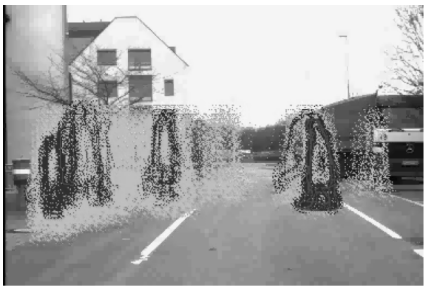
Traffic sign results

Daytime

Rain & Night

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

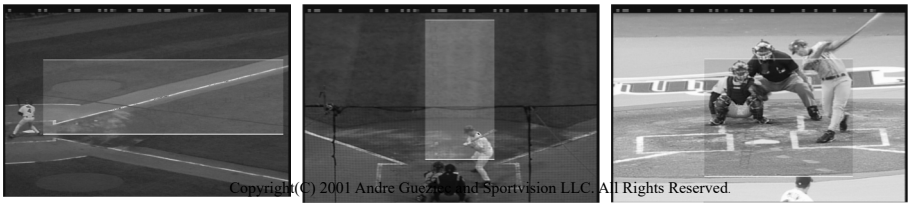
Entertainment

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

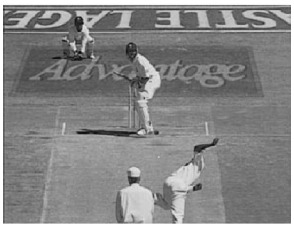
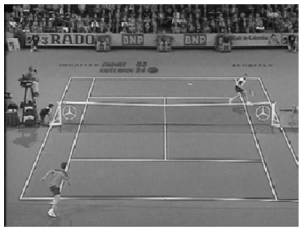
Applications – Sport Broadcasting

Tracking Baseball Pitches for Broadcast Television

- K Zone: system developed by Sportvision for ESPN in 2001.
- 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.



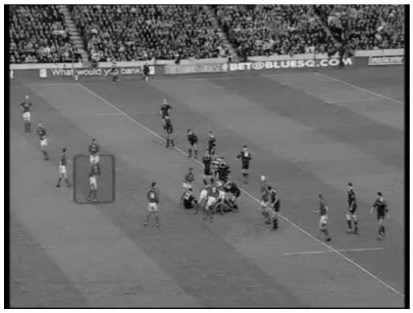
Copyright (C) 2001 Andre Guezec and Sportvision LLC. All Rights Reserved.

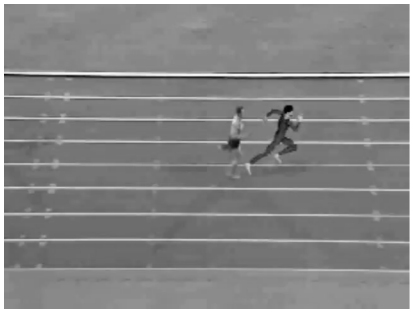




- Detect ground plan in video and introduce pictures on them

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
- https://www.youtube.com/watch?v=fVUvWOESZLY

Computational Photography

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

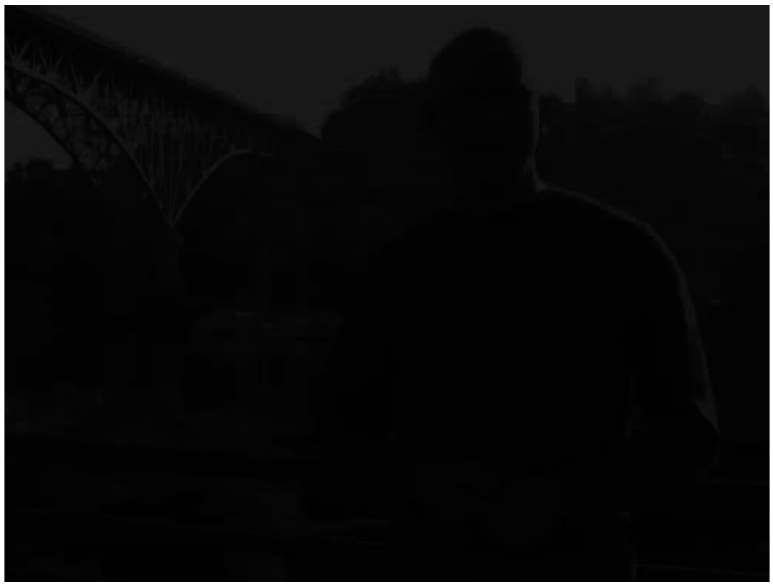
- Example Demos: check out vimeo.com
  - http://vimeo.com/5024379
  - http://vimeo.com/2345579

PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing

Connelly Barnes[1], Eli Shechtman[2,3], Adam Finkelstein[1], and Dan B Goldman[2]

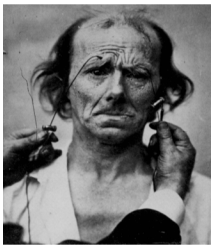
[1]Princeton University
[2]Adobe Systems
[3]University of Washington

SIGGRAPH 2009



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 mangement

Visual Surveillance


Consider a visual surveillance system

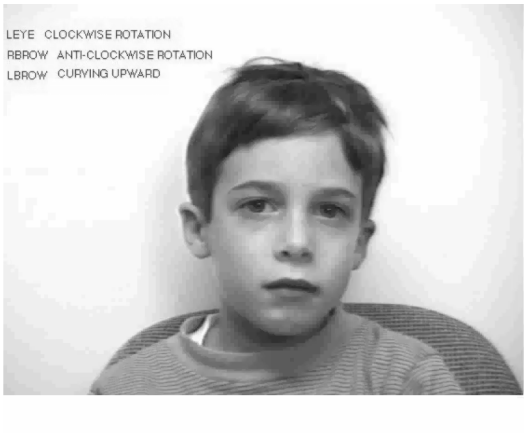
State of the art: archive huge volumes of video for eventual off-line human inspection

Goal : Automatic understanding of events happening in the site.

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

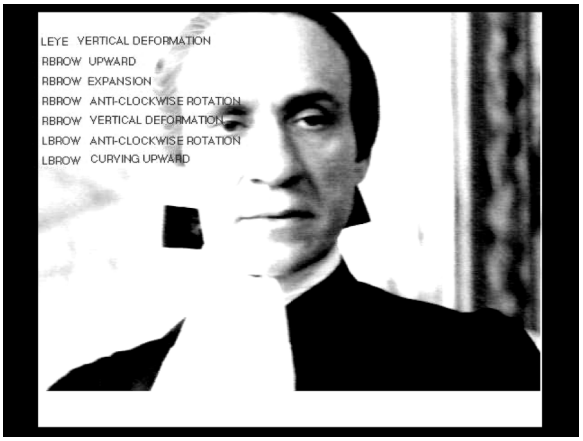
245 million video surveillance cameras installed globally in 2014 - according to IHS

Recognizing facial expressions



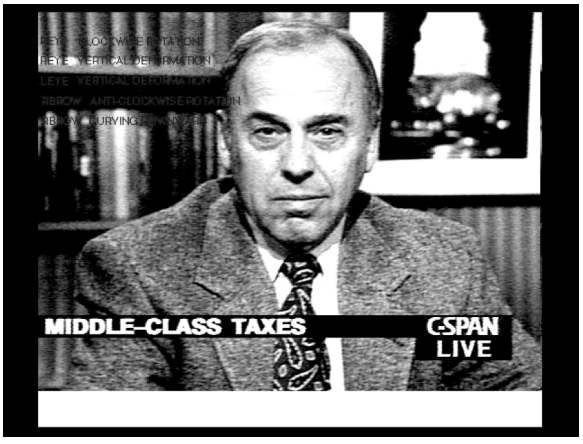
Videos by: Dr. Yaser Yacoob - Black, M.J., Yacoob, Y., Jepson, A.D., Fleet, D.J., Learning Parameterized Models of Image Motion, CVPR97

Recognizing facial expressions



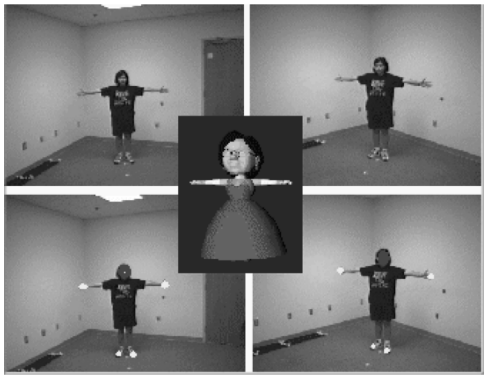
Videos by: Dr. Yaser Yacoob - Black, M.J., Yacoob, Y., Jepson, A.D., Fleet, D.J., Learning Parameterized Models of Image Motion, CVPR97

Recognizing facial expressions

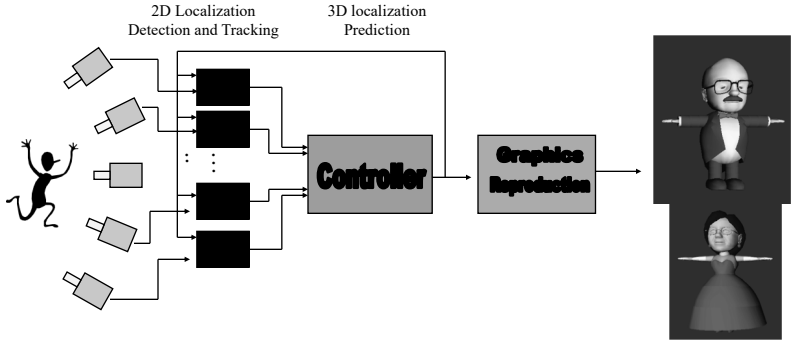


Videos by: Dr. Yaser Yacoob - Black, M.J., Yacoob, Y., Jepson, A.D., Fleet, D.J., Learning Parameterized Models of Image Motion, CVPR97

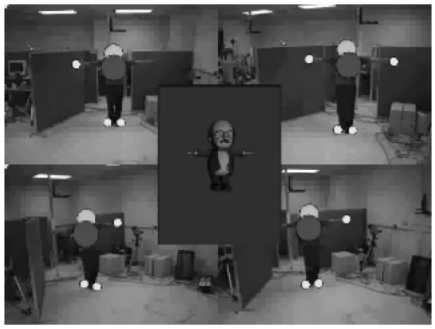
Motion Capture

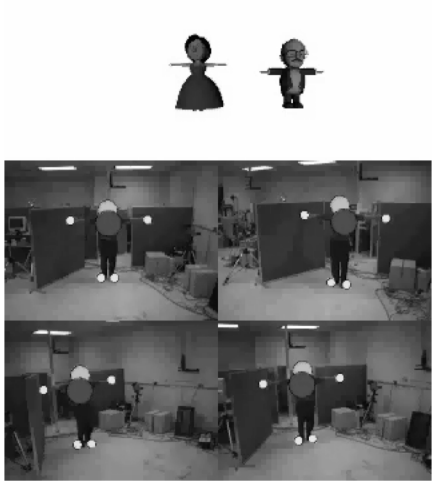


Motion Capture

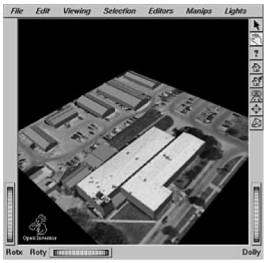


Videos by: Dr. Thanarat Horprasert





Defense

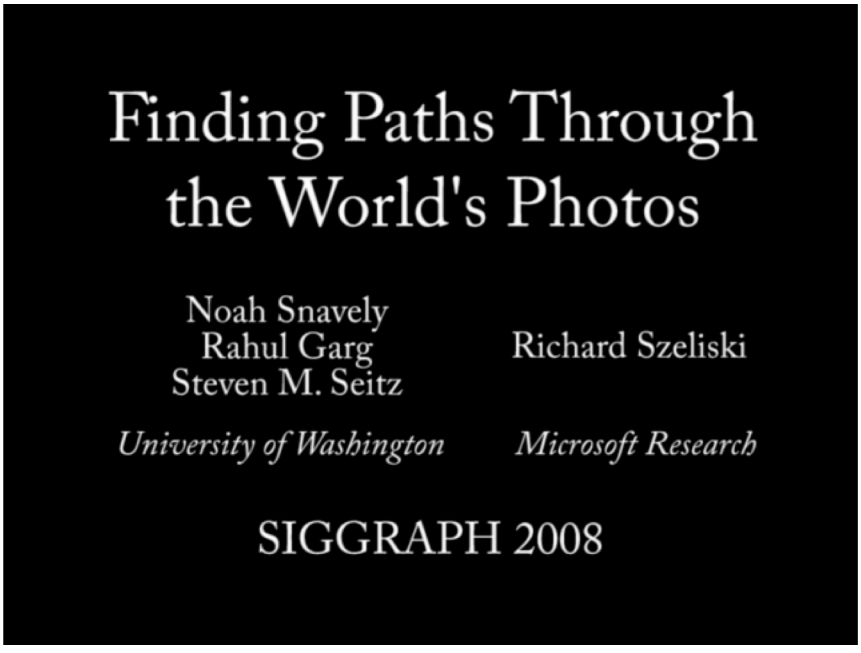


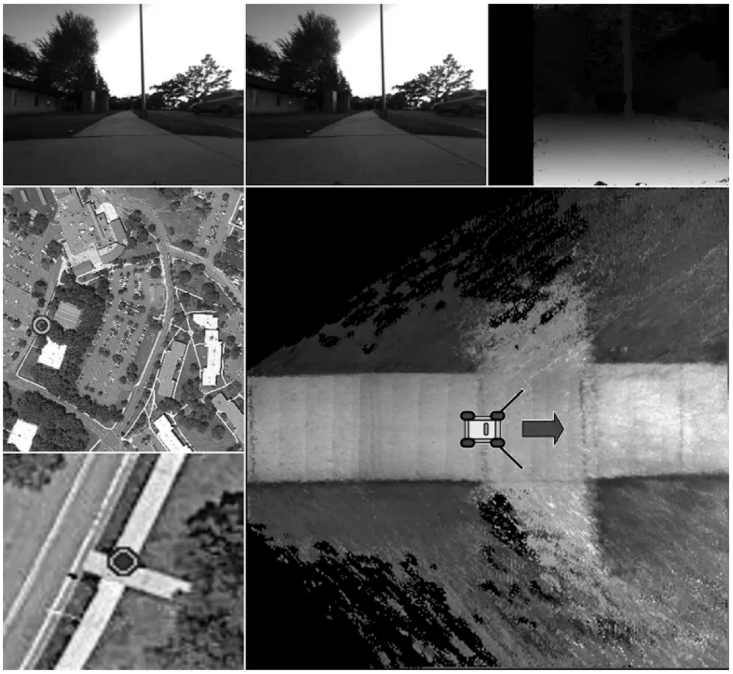
- Automatic target recognition systems
  - cruise missiles
  - air to surface "smart" missiles
- Reconnaissance
  - monitoring strategic sites
- Simulation
  - acquisition of terrain models from imagery
  - model acquisition of buildings, roads, etc.

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


Finding Paths Through the World's Photos
Noah Snavely
Rahul Garg
Steven M. Seitz
Richard Szeliski
University of Washington
Microsoft Research
SIGGRAPH 2008



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