BİL-613
Computer Vision

Instructor: Emre SÜMER, PhD

What is Computer Vision?

Every image tells a story

- Goal of computer vision: perceive the "story" behind the picture
- Compute properties of the world
  - 3D shape
  - Names of people or objects
  - What happened?

What is Computer Vision?

Can the computer match human perception?

- Yes and no (mainly no)
  - computers can be better at "easy" things
  - humans are much better at "hard" things
- But huge progress has been made
  - Especially in the last 10 years
  - What is considered "hard" keeps changing

What is Computer Vision?

Human perception has its shortcomings

[Sinha and Poggio, *Nature*, 1996]

What is Computer Vision?

But humans can tell a lot about a scene from a little information…

Source: "80 million tiny images" by Torralba, et al.

What is Computer Vision?

Make computers understand images and video.

What kind of scene?

Where are the cars?

How far is the building?

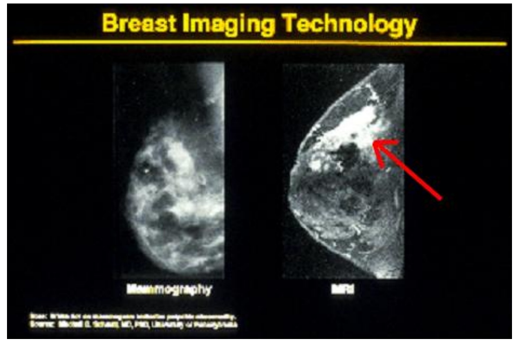
…

Vision is really hard

- Vision is an amazing feature of natural intelligence
  - Visual cortex occupies about 50% of Macaque brain
  - More human brain devoted to vision than anything else

Why computer vision matters

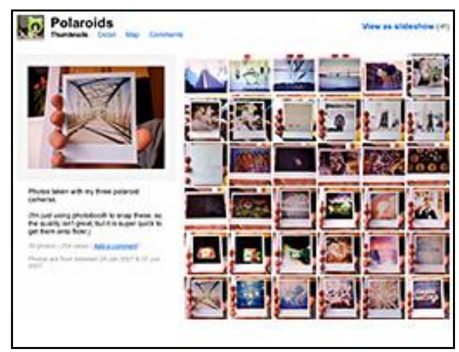
Safety



Health

Security

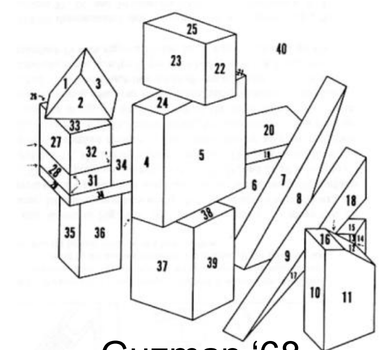
Fun



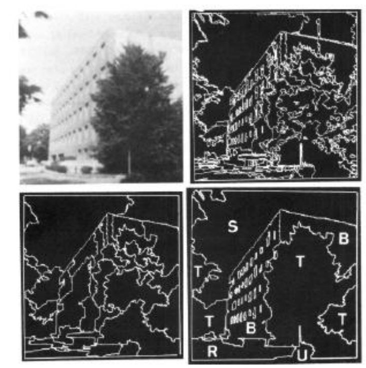
Access

Brief history of computer vision

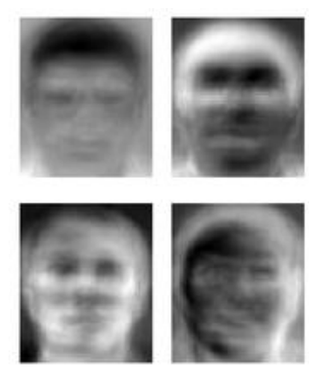
- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis in vogue
- 2000's: broader recognition; large annotated datasets available; video processing starts
- 2030's: robot uprising?



Guzman '68



Ohta Kanade '78



Turk and Pentland '91

How vision is used now

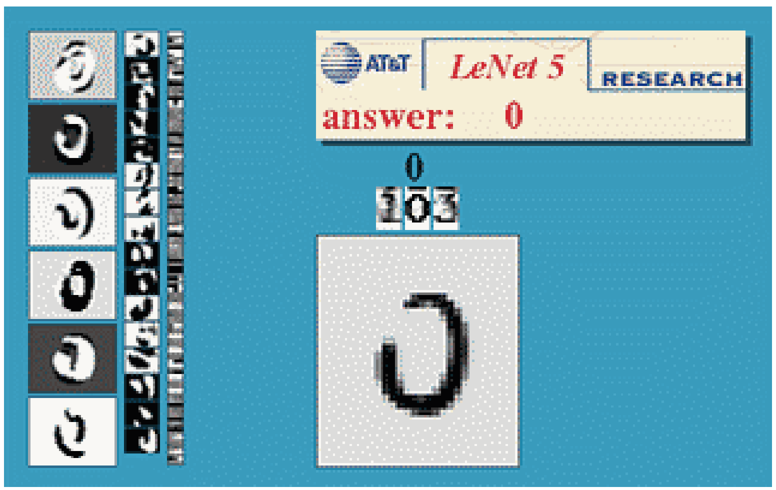
- Examples of state-of-the-art

Some of the following slides by Steve Seitz

Optical character recognition (OCR)

Technology to convert scanned docs to text

- If you have a scanner, it probably came with OCR software



Digit recognition, AT&T labs
http://www.research.att.com/~yann/

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

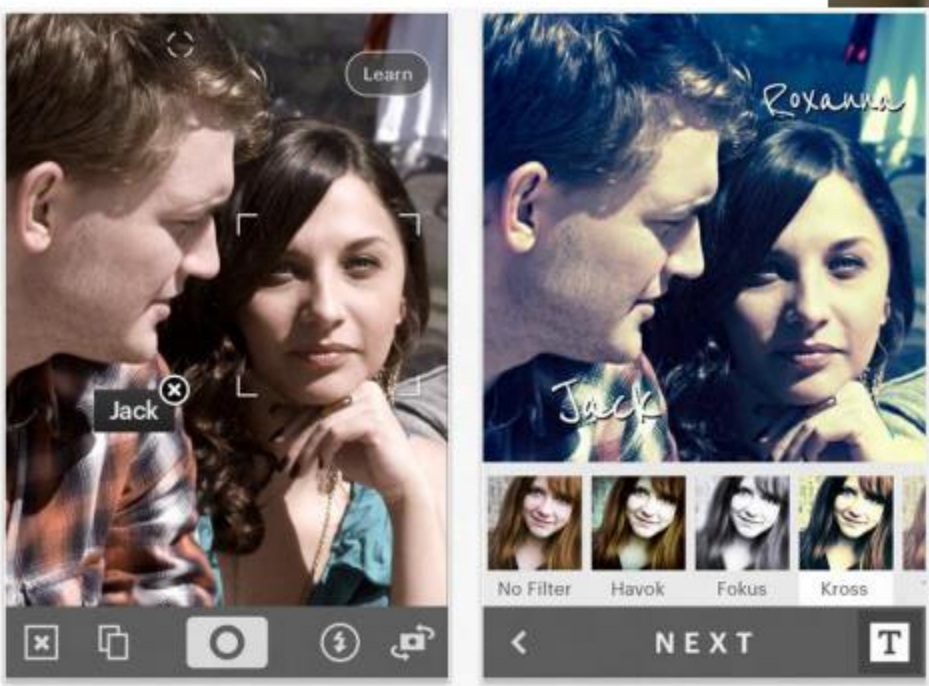
Face detection

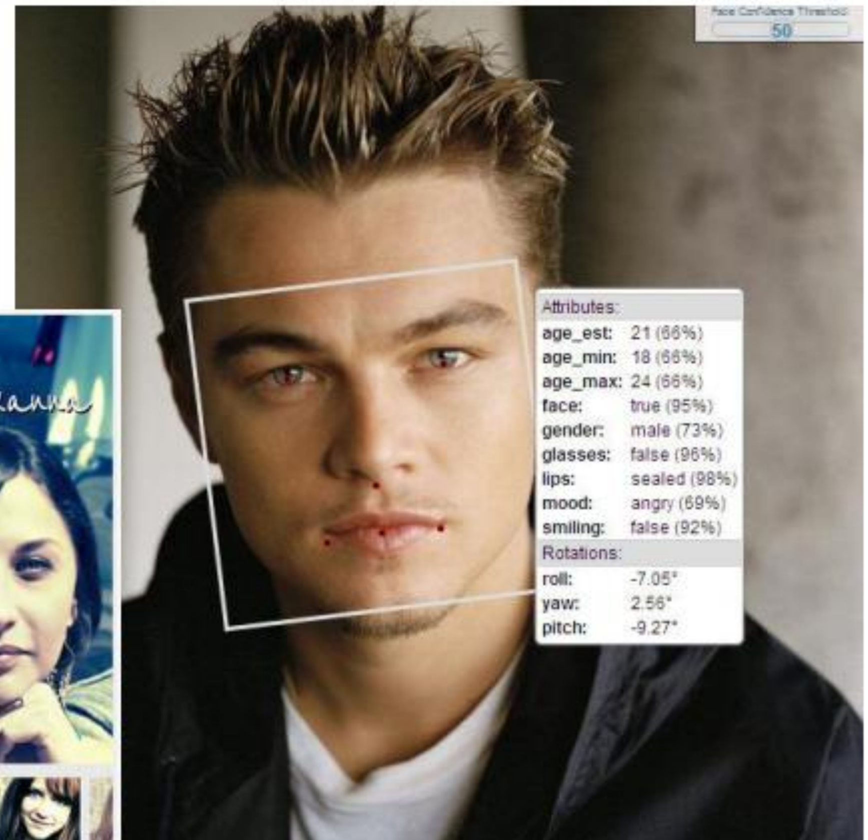
- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, …

Face recognition

face.com

http://developers.face.com/tools/

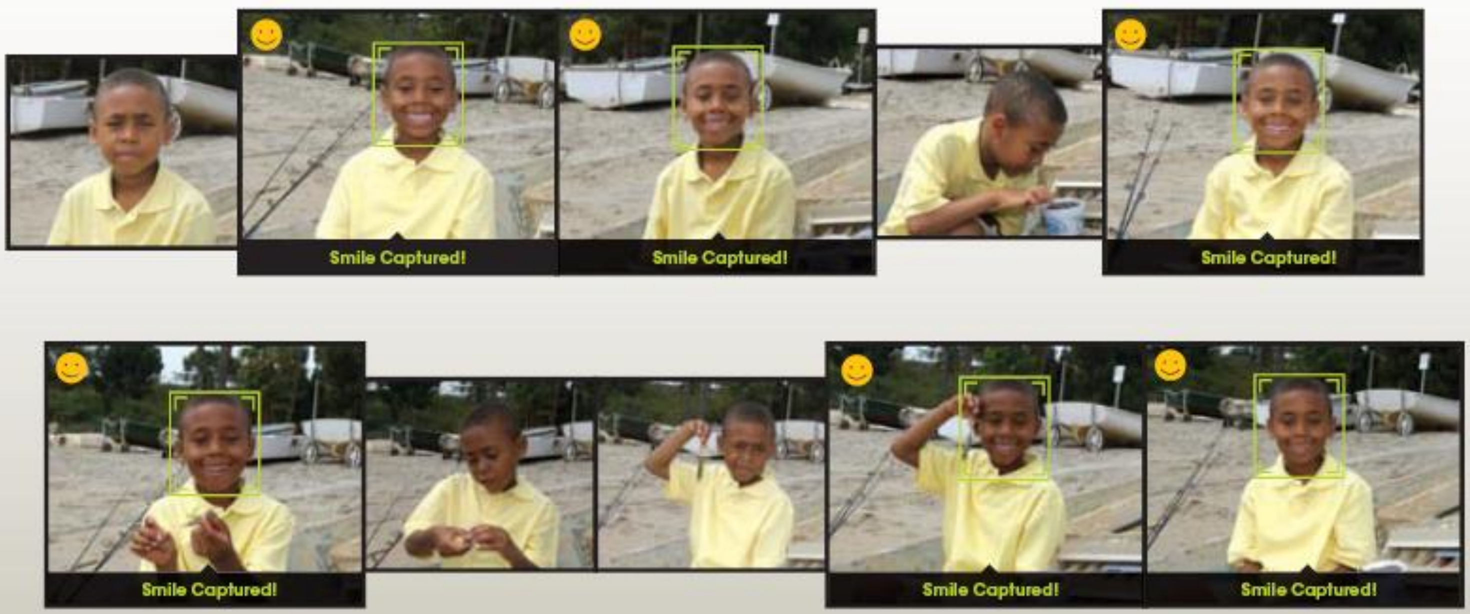




Smile detection

The Smile Shutter flow

Imagine a camera smart enough to catch every smile! In Smile Shutter Mode, your Cyber-shot® camera can automatically trip the shutter at just the right instant to catch the perfect expression.



[Sony Cyber-shot® T70 Digital Still Camera](#)

3D from thousands of images

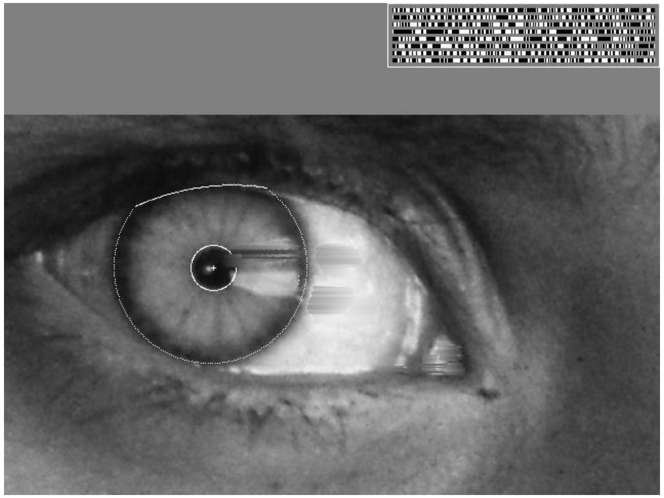
Object recognition (in supermarkets)

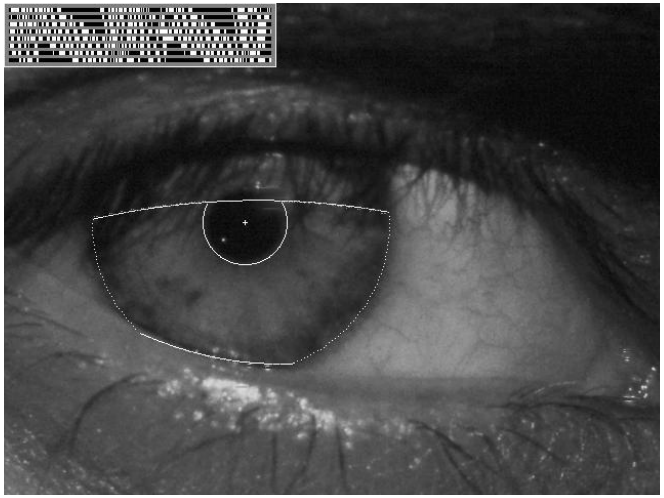
[LaneHawk by EvolutionRobotics]

“A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you are assured to get paid for it… “

Vision-based biometrics

"*How the Afghan Girl was Identified by Her Iris Patterns*" Read the story
wikipedia





Login without a password…

Fingerprint scanners on
many new laptops,
other devices

Face recognition systems now
beginning to appear more widely
http://www.sensiblevision.com/

Object recognition (in mobile phones)

[Point & Find](), [Nokia]()

[Google Goggles]()

Object recognition (in mobile phones)

iPhone Apps: kooaba (www.kooaba.com)

Object recognition (in mobile phones)

Special effects: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Special effects: motion capture

Pirates of the Carribean, Industrial Light and Magic

Sports

Sportvision first down line
Nice [explanation](#) on www.howstuffworks.com

http://www.sportvision.com/video.html

Smart cars

Slide content courtesy of Amnon Shashua

- Mobileye
  - Vision systems currently in high-end BMW, GM, Volvo models
  - By 2010: 70% of car manufacturers.

Google cars

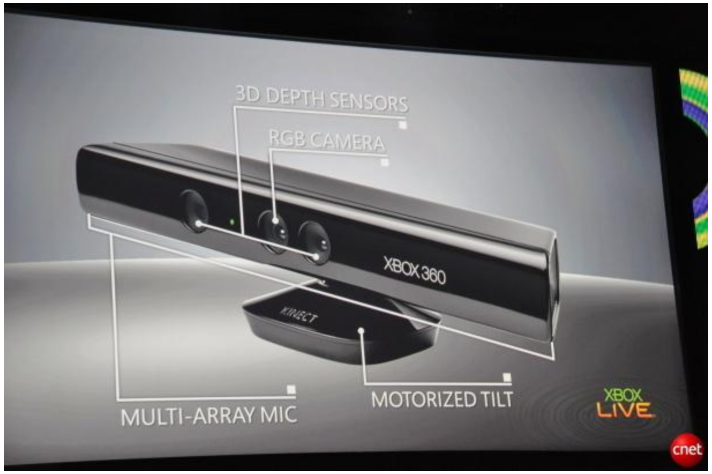
Oct 9, 2010. "Google Cars Drive Themselves, in Traffic". *The New York Times*. John Markoff

June 24, 2011. "Nevada state law paves the way for driverless cars". *Financial Post*. Christine Dobby

Aug 9, 2011, "Human error blamed after Google's driverless car sparks five-vehicle crash". *The Star* (Toronto)

Interactive Games: Kinect

- Object Recognition: http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o
- Mario: http://www.youtube.com/watch?v=8CTJL5lUjHg
- 3D: http://www.youtube.com/watch?v=7QrnwoO1-8A
- Robot: http://www.youtube.com/watch?v=w8BmgtMKFbY



Vision in space

NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

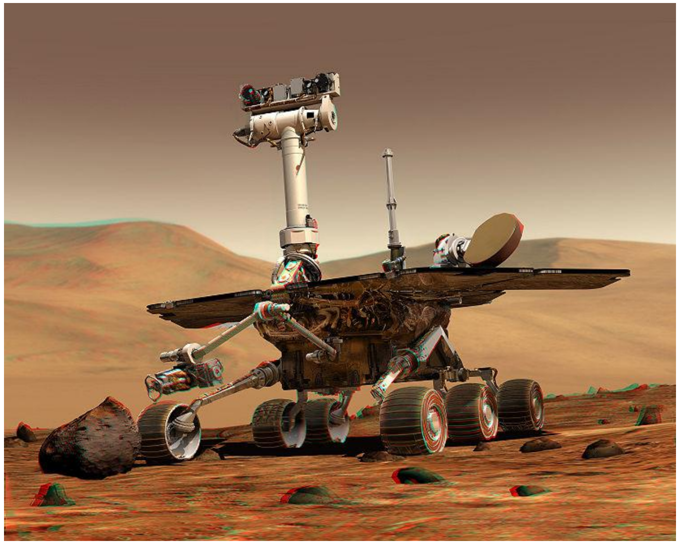
Vision systems (JPL) used for several tasks

- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

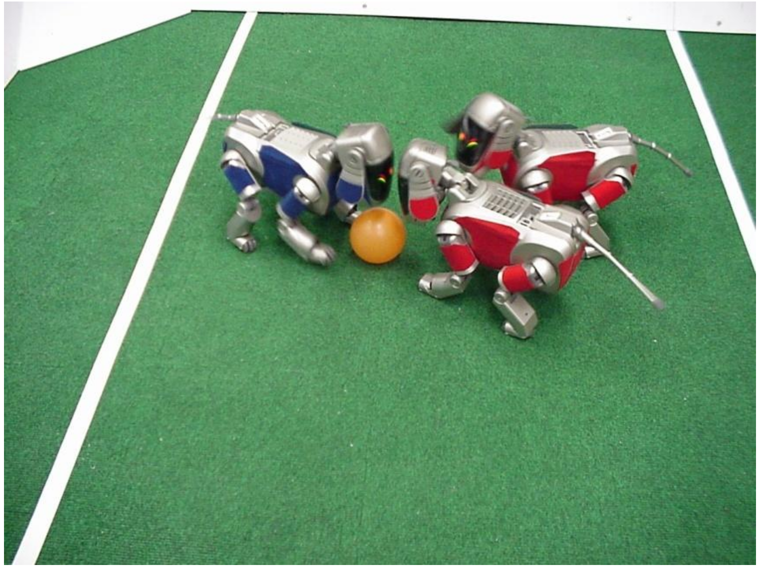
Industrial robots

Vision-guided robots position nut runners on wheels

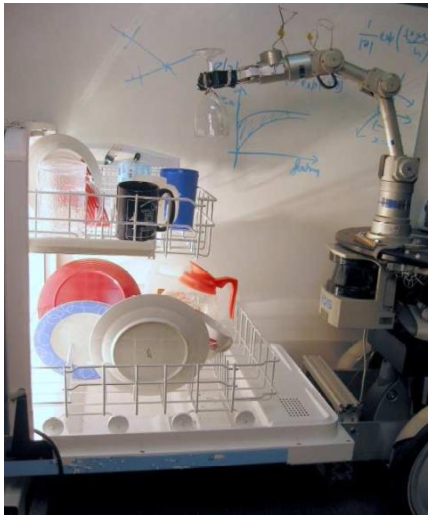
Mobile robots



NASA’s Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover

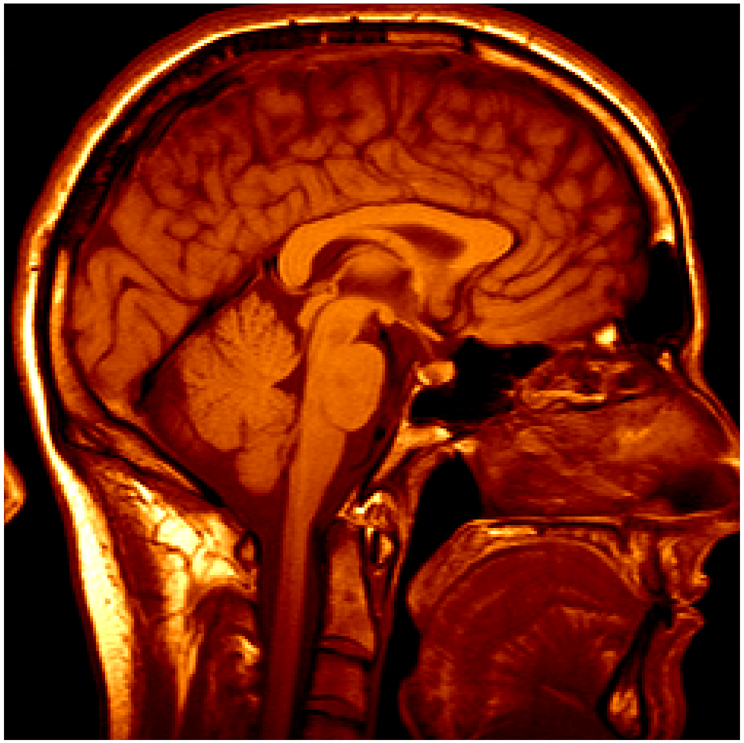


http://www.robocup.org/



Saxena et al. 2008
STAIR at Stanford

Medical imaging



3D imaging
MRI, CT

Image guided surgery
[Grimson et al., MIT]

Why is Computer Vision Difficult?

Viewpoint variation

Illumination

Scale

Why is Computer Vision Difficult?

Intra-class variation

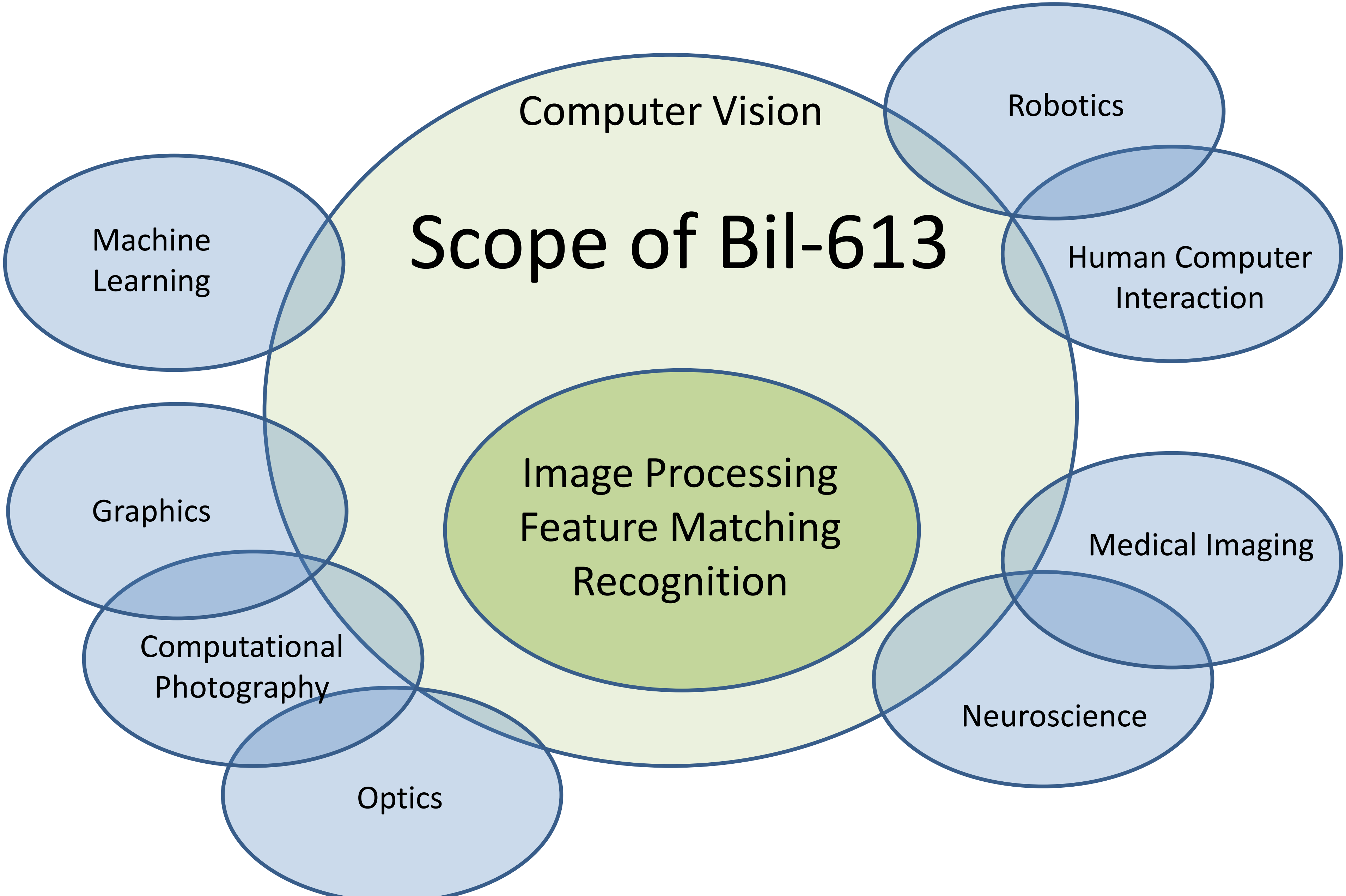
Motion (Source: S. Lazebnik)

Background clutter

Occlusion

Computer Vision and Nearby Fields

- Computer Graphics: Models to Images
- Comp. Photography: Images to Images
- Computer Vision: Images to Models


Computer Vision
Scope of Bil-613
Image Processing
Feature Matching
Recognition
Robotics
Human Computer Interaction
Machine Learning
Graphics
Computational Photography
Optics
Medical Imaging
Neuroscience

Prerequisites

- **Linear algebra**, basic calculus, and probability
- Experience with image processing or Matlab will help but is not necessary

Course Syllabus

Course Description: The course includes the topics: Pinhole cameras, pixel shading models, human color perception, color physics, color spaces, image color model, linear filters, building local features, texture representation, texture synthesis, image denoising, segmentation ideas & applications, segmentation by clustering pixels, Hough transform, fitting lines, robustness, RANSAC, simple tracking strategies, tracking by matching, Kalman filters, data association and image registration.

Course Objective: The objective of this course is to give image formation basics, early (just one image) and mid-level vision concepts, and an introduction to high level vision.

Instructor	Hours	Grading
Name: Emre SÜMER, PhD.	**Class Day:** Thursday	Review Paper...............%15
Office: D-414	**Class Hours:** 18:00 – 21:00	Programming Projects.....%15
Phone: 246 6666 / 1305	**Location:** B-308	Midterm Exam.............%30
E-mail: esumer@baskent.edu.tr	**Office Hours:** Any proper time	Final Exam...................%40
Web: www.baskent.edu.tr/~esumer		

Course Topics

<u>Tentative Weekly Course Schedule</u>

1- (20 Feb.) Introduction to Computer Vision

2- (27 Feb.) Geometric Camera Models

3- (06 Mar.) Light and Shading

4- (13 Mar.) Color

5- (20 Mar.) Linear Filters

6- (27 Mar.) Local Image Features

7- (03 Apr.) Texture

8- (10 Apr.) ***Midterm Exam***

9- (17 Apr.) Segmentation by Clustering

10- (24 Apr.) Grouping and Model Fitting

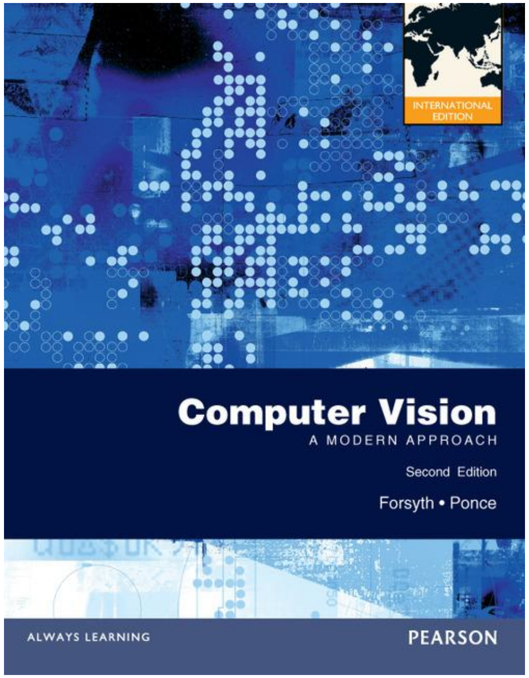
11- (01 May) Holiday – No Class

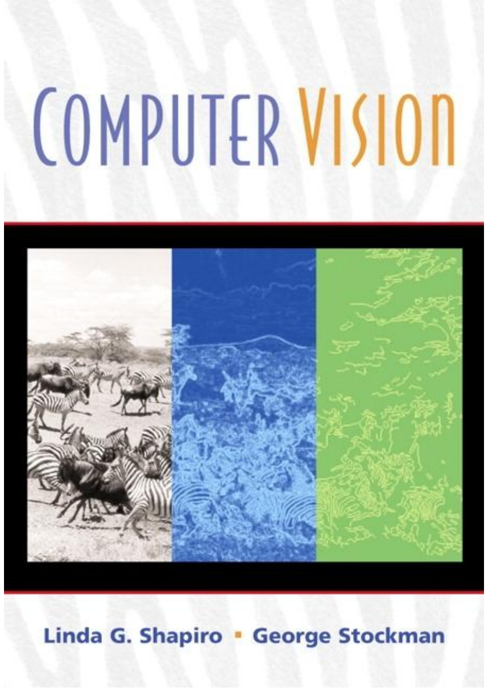
12- (08 May) Tracking

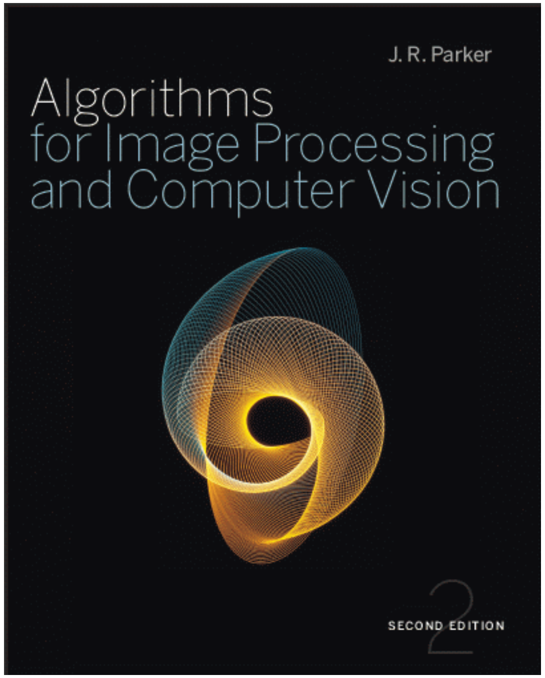
13- (15 May) Registration

14- (22 May) Review Paper Presentations

Textbooks







1. Forsyth, D.A. & Ponce, J., "Computer Vision: A Modern Approach", 2nd ed., Prentice Hall, 2011.
2. Shapiro, L.G. & Stockman, G.C., "Computer Vision", Prentice Hall, 2001.
3. Parker, J.R., "Algorithms for Image Processing and Computer Vision", Wiley, 2010.

The Physics of Imaging

- How images are formed
  - Cameras
  - Light
  - Color

Early Vision in One Image

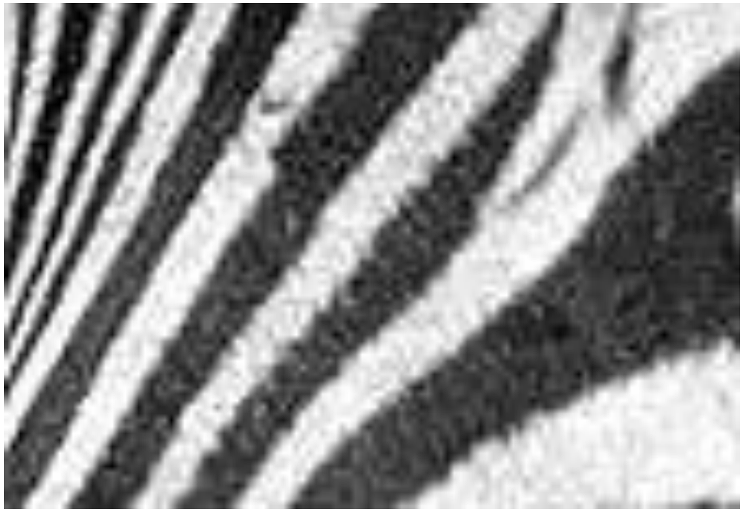
- Representing small patches of image
  - For three reasons
    - We wish to establish correspondence between (say) points in different images, so we need to describe the neighborhood of the points
    - Sharp changes are important in practice --- known as “edges”
    - Representing texture by giving 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

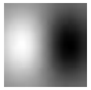
- Filter outputs
  - essentially form a dot-product between a pattern and an image, while shifting the pattern across the image
  - strong response -> image locally looks like the pattern
  - e.g. derivatives measured by filtering with a kernel that looks like a big derivative (bright bar next to dark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

Filtering

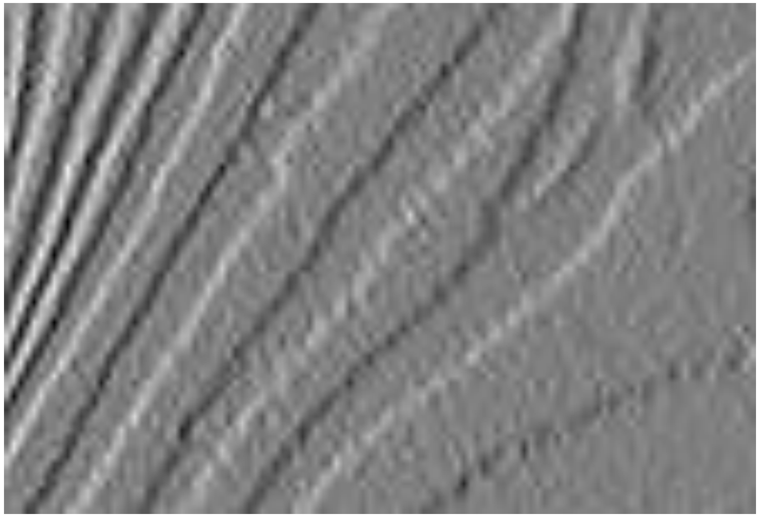
Convolve this image



With this kernel



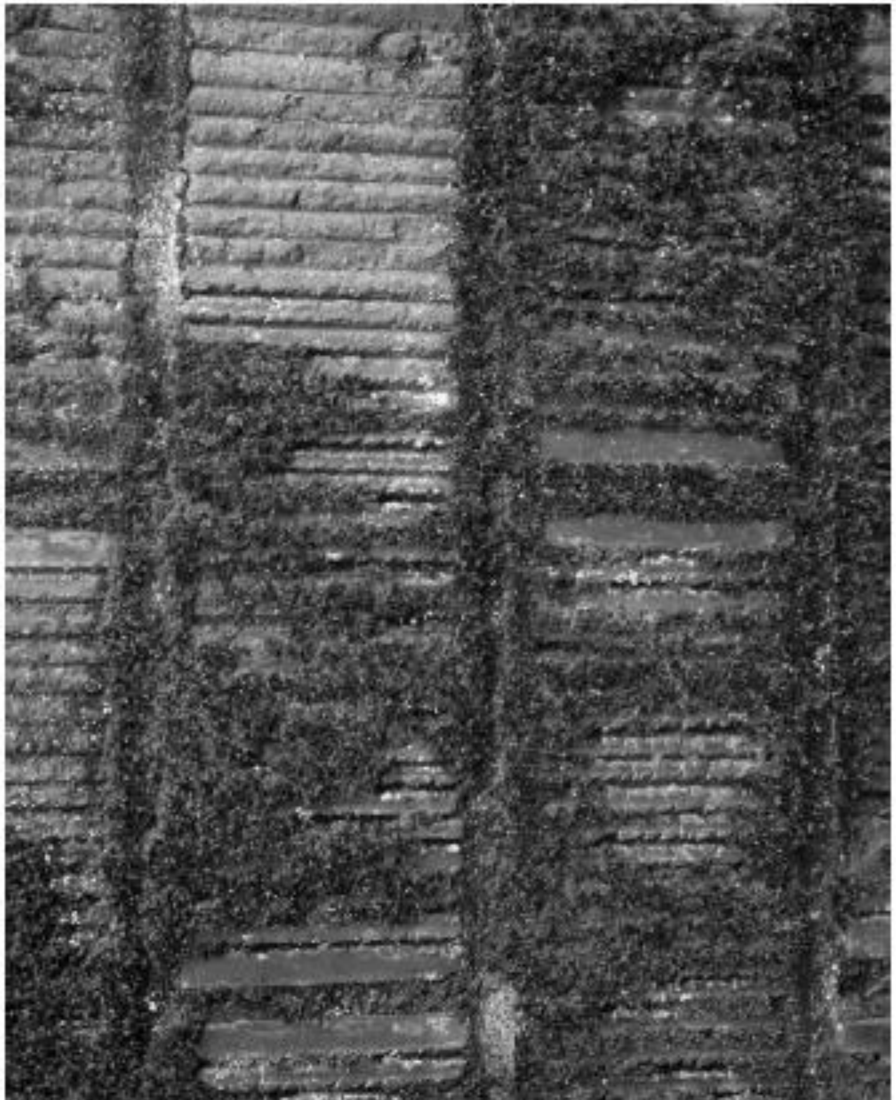
To get this



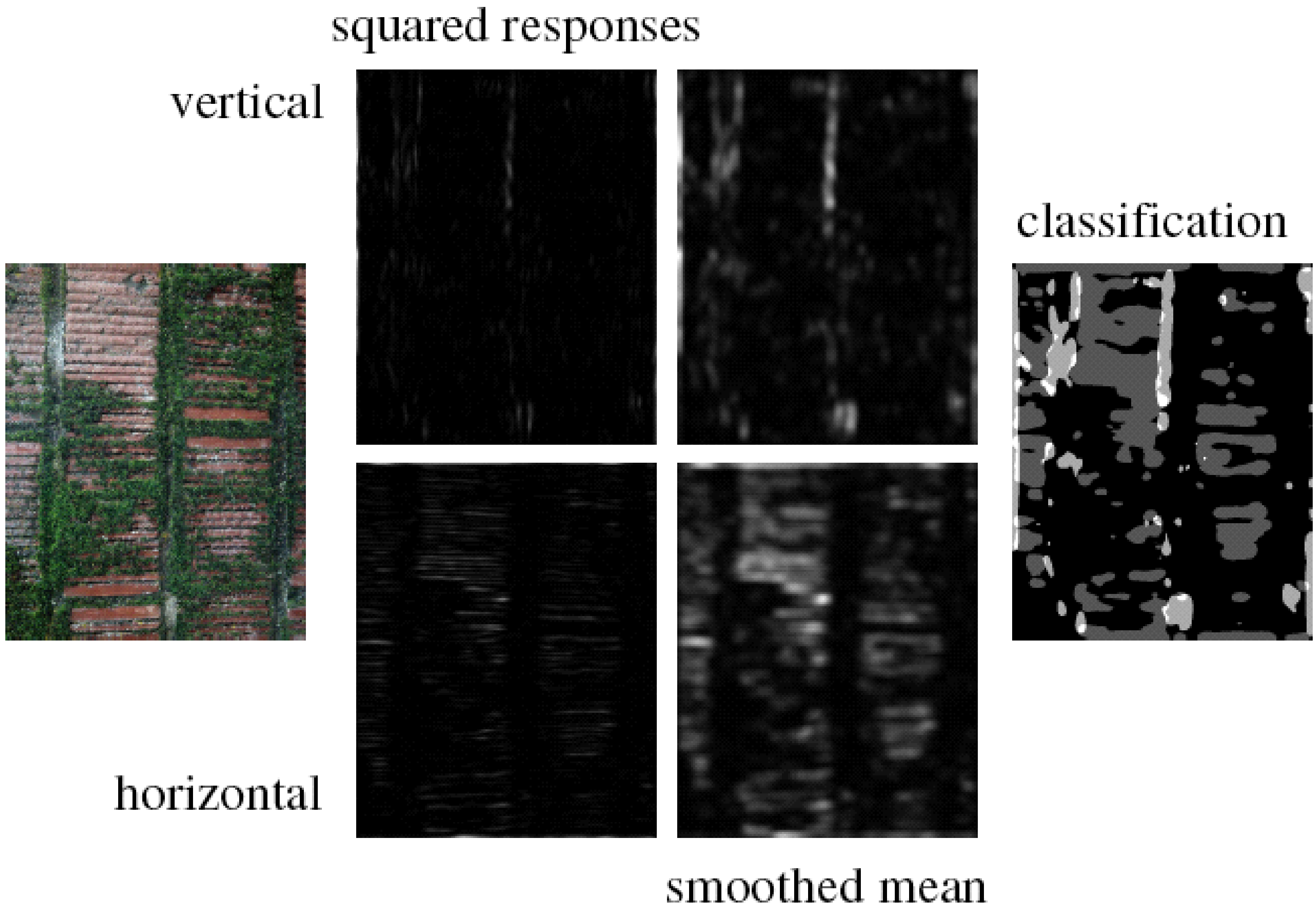
Texture

- 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

Texture



Texture


squared responses
vertical
classification
horizontal
smoothed mean

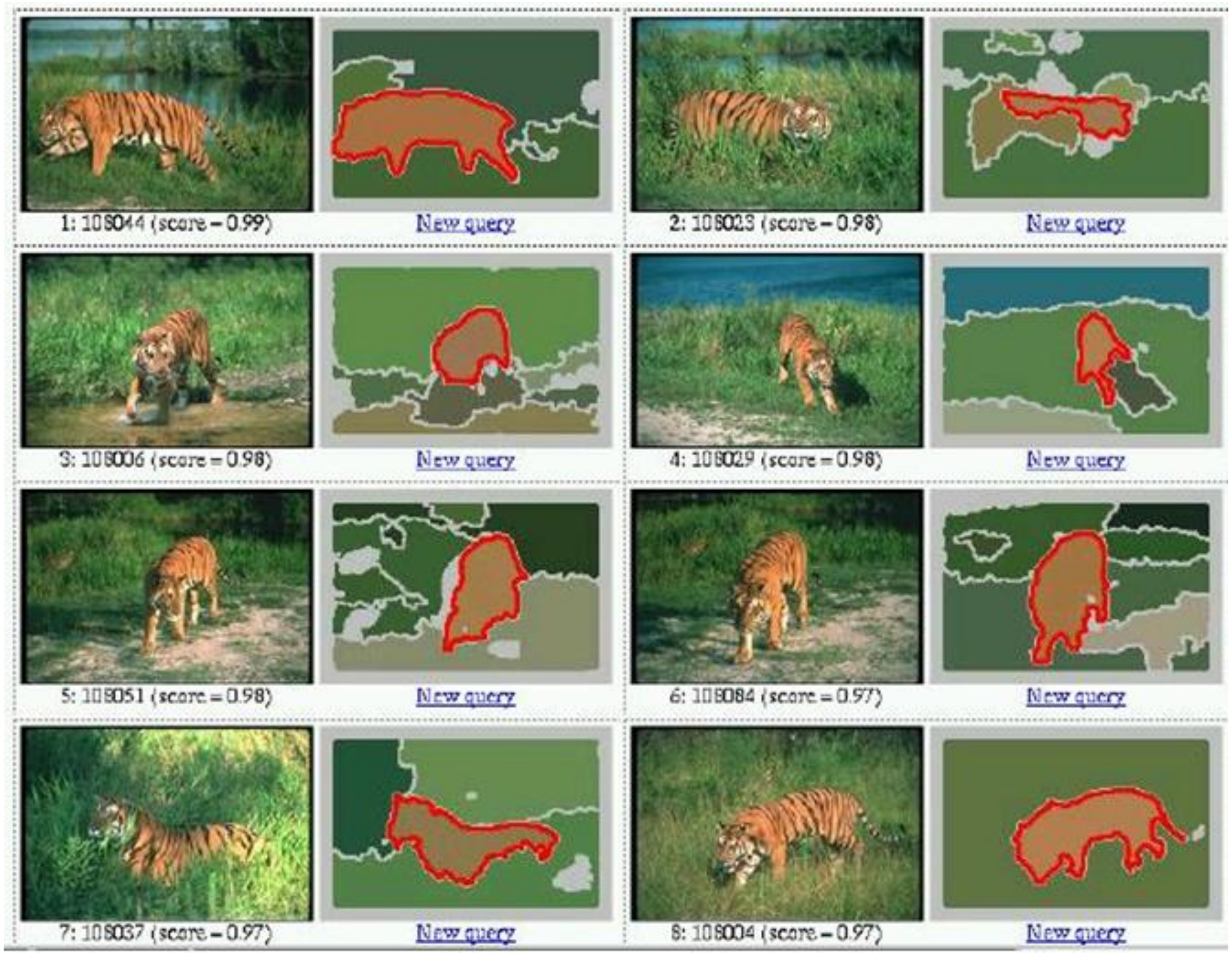
Mid-Level Vision

- 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

Segmentation

- Which image components “belong together”?
- Belong together=lie on the same object
- Cues
  - similar colour
  - similar texture
  - not separated by contour
  - form a suggestive shape when assembled

Segmentation



Tracking

- Use a model to predict next position and refine using next image
- Model:
  - simple dynamic models
  - kinematic models
  - etc.
- Face tracking and eye tracking now work rather well

Tracking

(a) (b) (c)

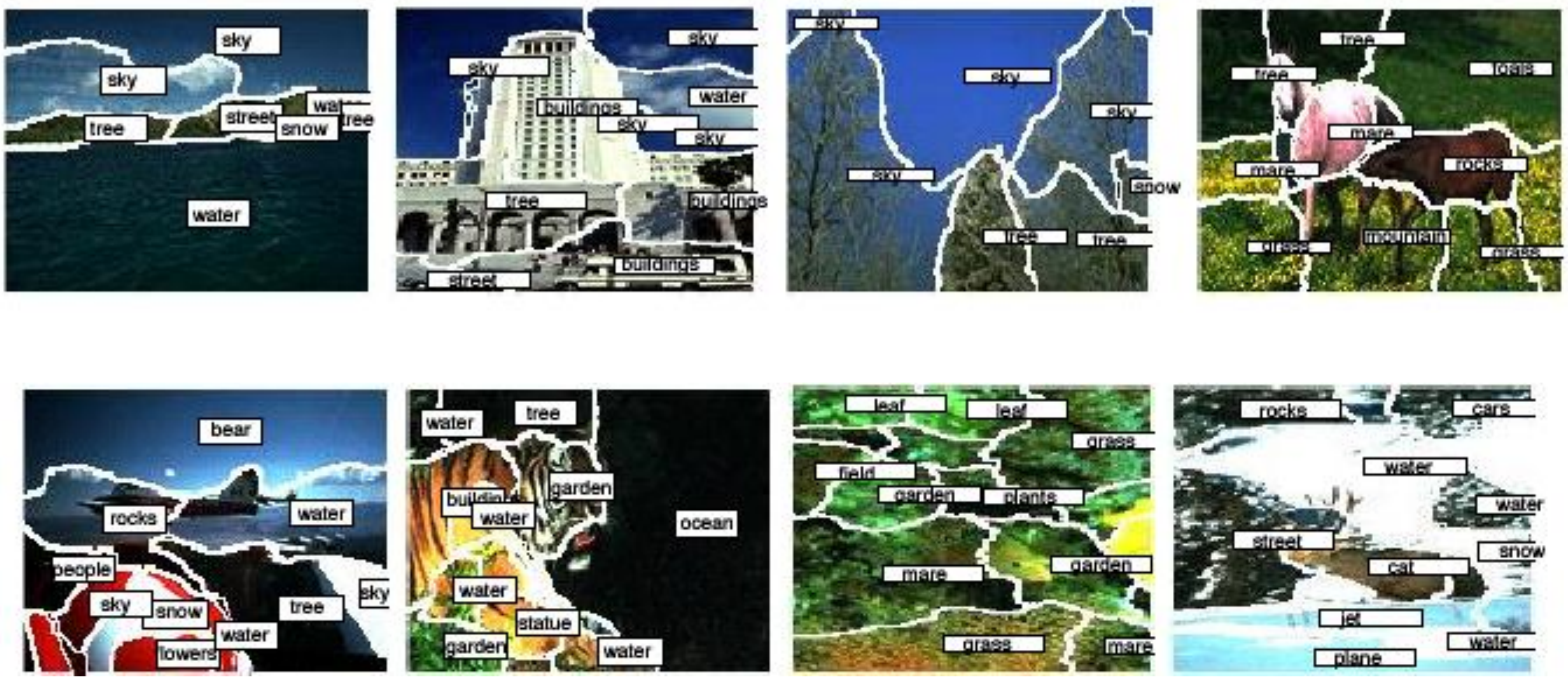
High Level Vision

- Using classifiers and probability to 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

Matching templates

http://www.ri.cmu.edu/projects/project_271.html

Classifiers



My Research Interests

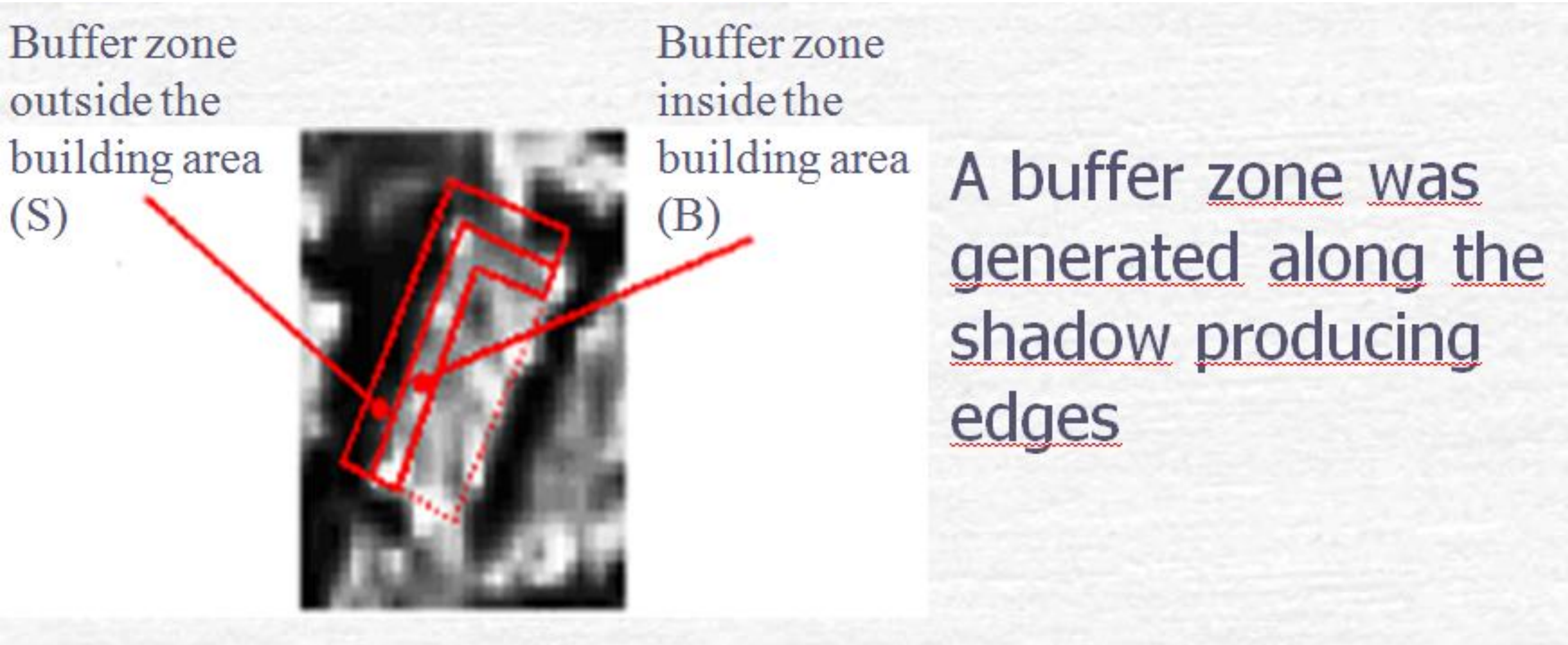
Earthquake Damage Detectıon using Image Processıng Technıques and GIS (MS Thesis)

* A post-earthquake panchromatic aerial imagery

* Vector building boundaries

My Research Interests

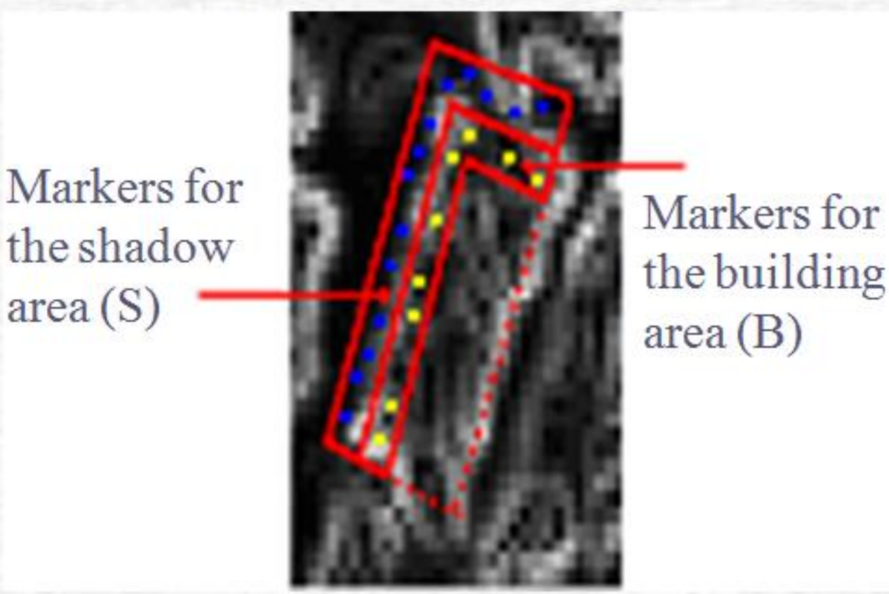
Earthquake Damage Detectıon using Image Processıng Technıques and GIS (MS Thesis)



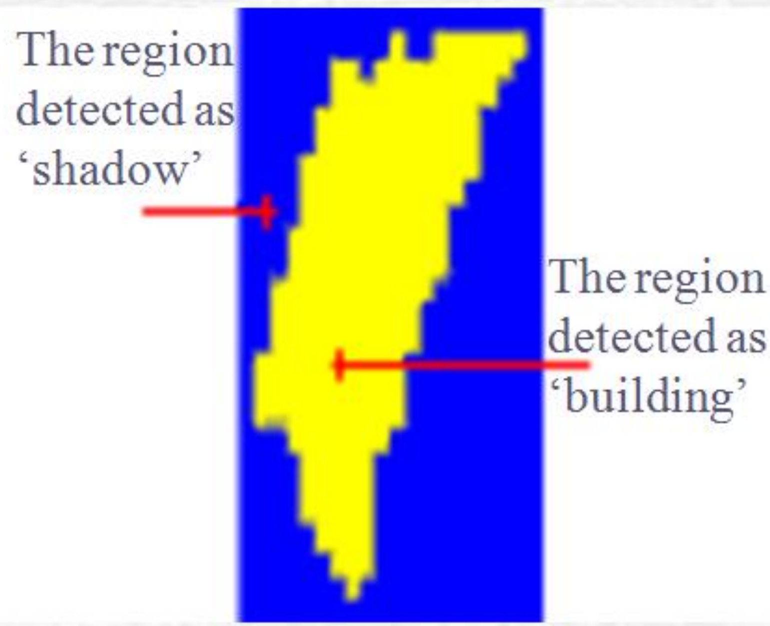
My Research Interests

Earthquake Damage Detectıon using Image Processıng Technıques and GIS (MS Thesis)

- Locations of the markers were seeded randomly in the gradient image

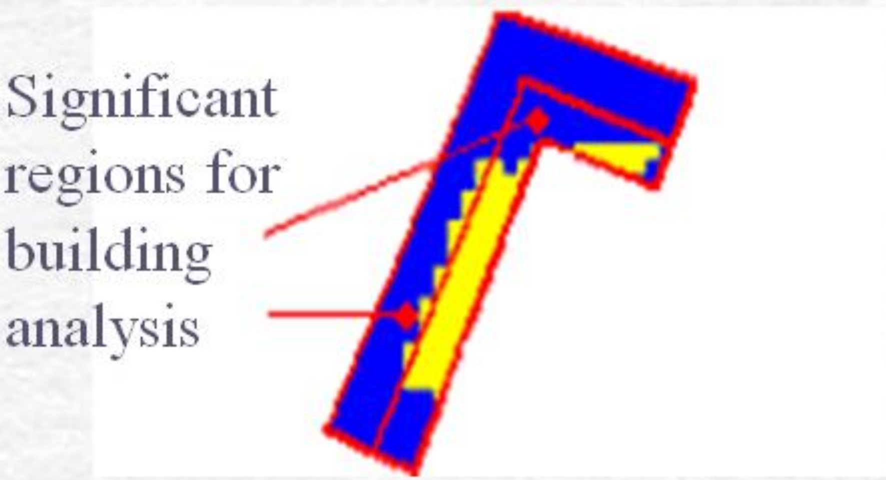


- A two-colored output image was generated



My Research Interests

Earthquake Damage Detectıon using Image Processıng Technıques and GIS (MS Thesis)

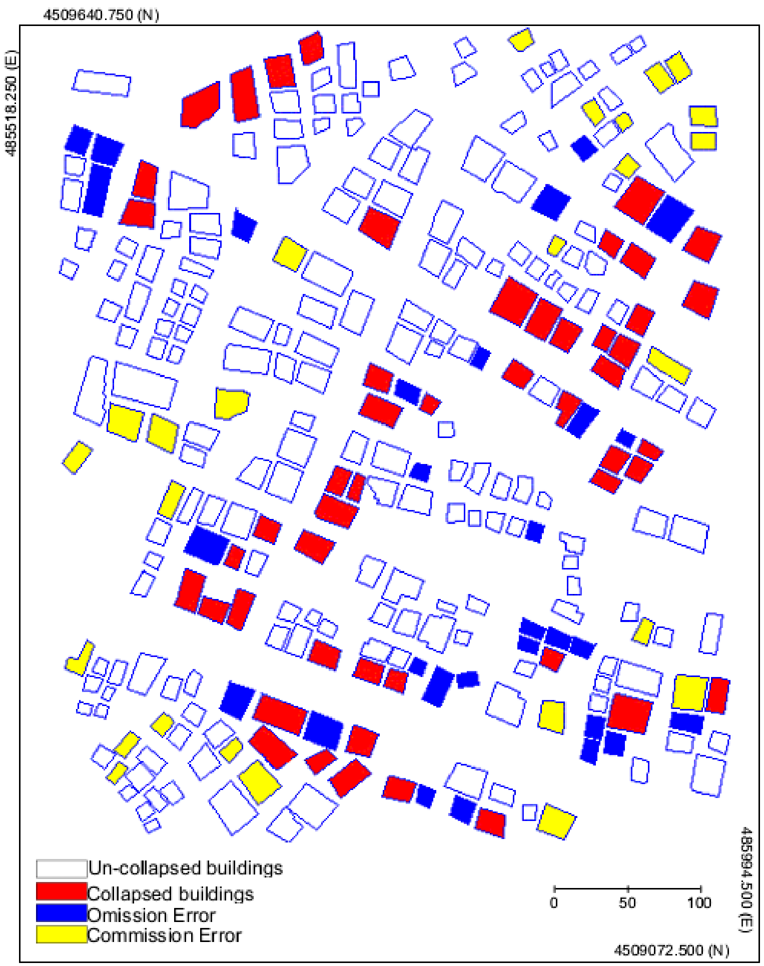


- Shadow and building pixels were counted
- Ratios were computed for building and shadow pixels

An example : Total Assessed Pixels: **99**
Detected Shadow Pixels : **91**
Detected Building Pixels : **66**
Shadow Ratio: **0.9192**
Building Ratio: **0.6667**

My Research Interests

Earthquake Damage Detectıon using Image Processıng Technıques and GIS (MS Thesis)



My Research Interests

Reconstruction of Photorealistic Building Models (PhD Thesis)

* 2-D building extraction
  ** from high resolution satellite imagery
  ** using adaptive fuzzy-genetic approach

* Automated building facade texture extraction
  ** from ground-level building images
  ** using an iterative image segmentation approach

* Automated facade texture rectification
  ** from extracted building facade images
  ** using a fully automated edge-based approach

* Texture occlusion removal
  ** from rectified building facade images
  ** using an image matching approach

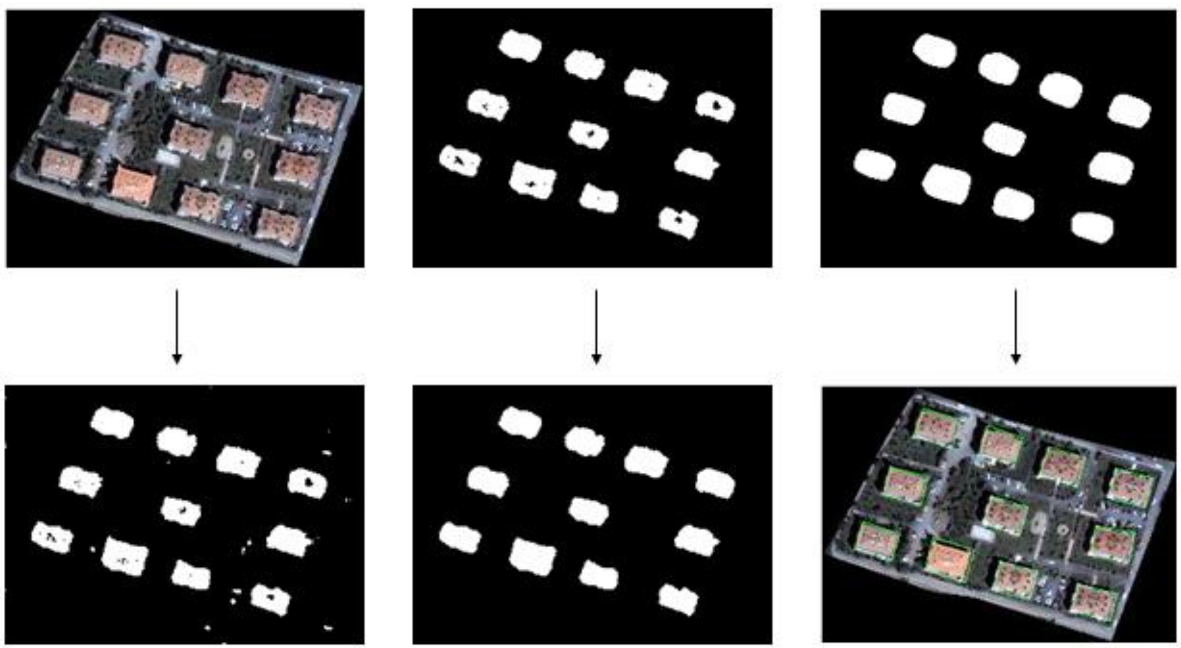
* 3-D extrusion model generation and visualization
  ** from 2-D building coordinates with building height info
  ** using Virtual Reality Modeling (VRM)

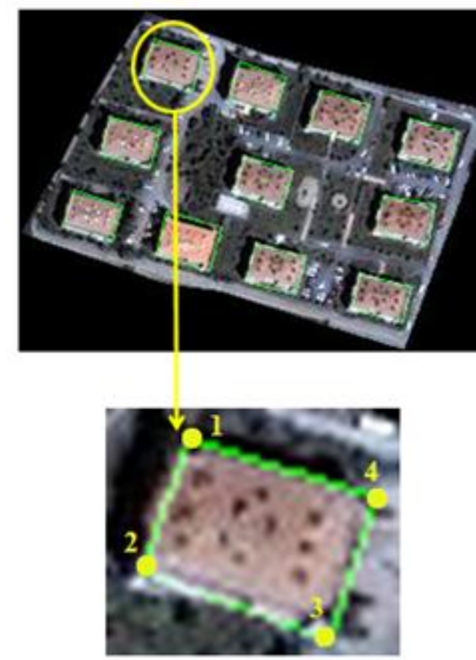
* Texture mapping onto 3-D model
  ** from occlusion-free building facade texture images
  ** using a GPS assisted approach

My Research Interests

Reconstruction of Photorealistic Building Models (PhD Thesis)

2-D building extraction

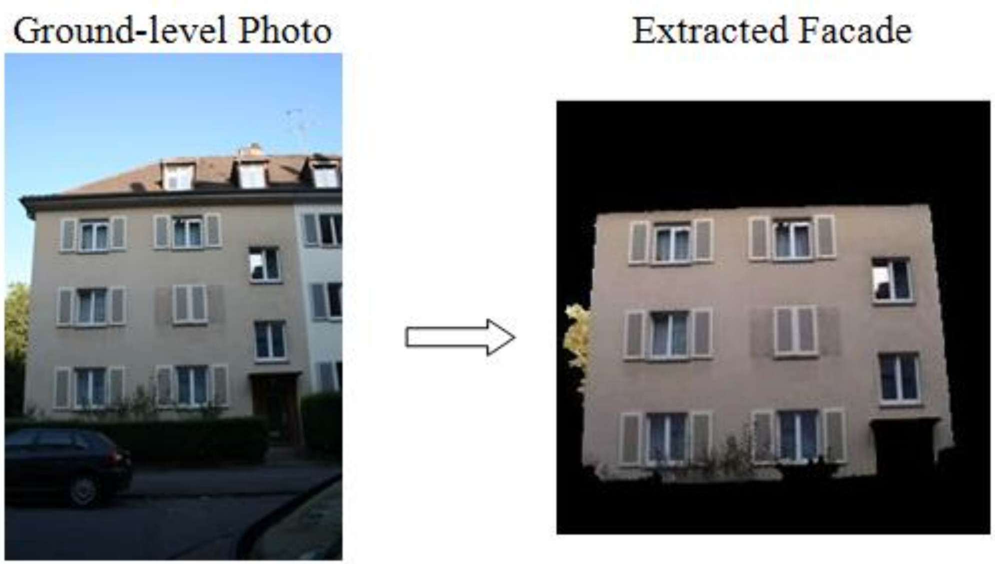




My Research Interests

Reconstruction of Photorealistic Building Models (PhD Thesis)

Automated building facade texture extraction



My Research Interests

Reconstruction of Photorealistic Building Models (PhD Thesis)

Automated facade texture rectification

My Research Interests

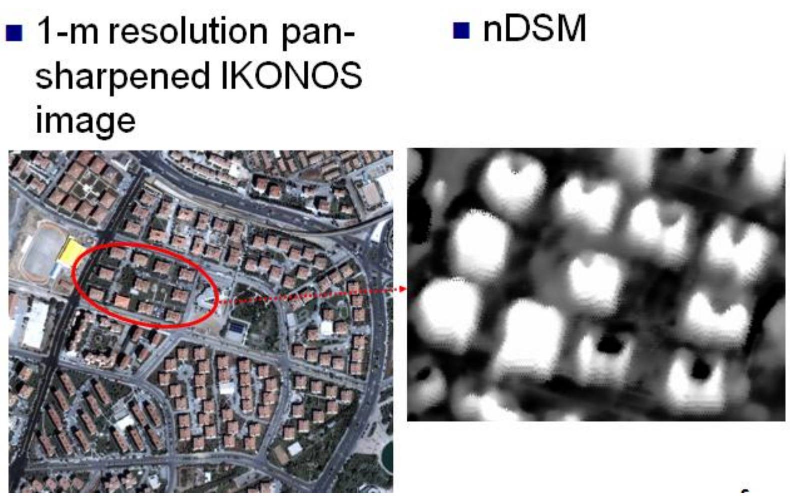
Reconstruction of Photorealistic Building Models (PhD Thesis)

Texture occlusion removal

My Research Interests

Reconstruction of Photorealistic Building Models (PhD Thesis)

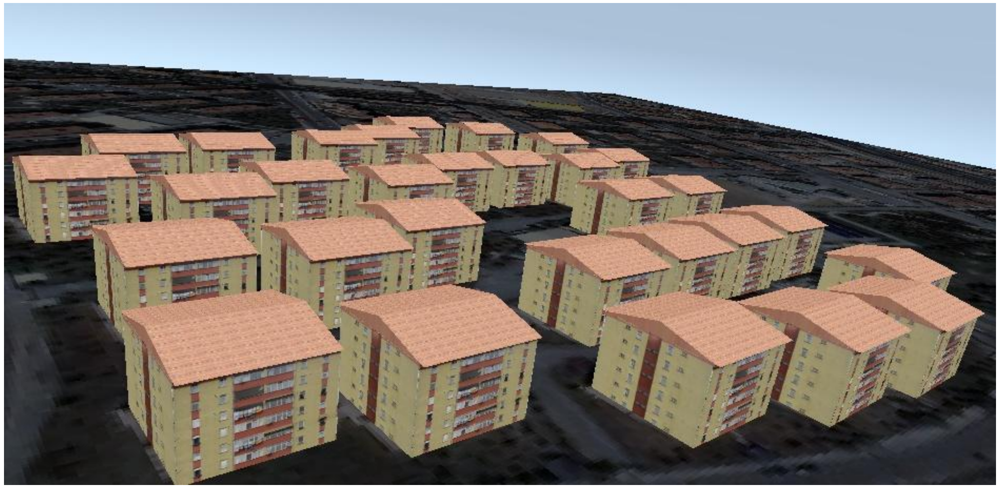
3-D extrusion model generation and visualization



My Research Interests

Reconstruction of Photorealistic Building Models (PhD Thesis)

Texture mapping onto 3-D model



My Publications

2008 (SCI)

- “Building-based Damage Detection due to Earthquake using the Watershed Segmentation of Post-event Aerial Images”, *International Journal of Remote Sensing*, Vol.29, No.11, pp.3073-3089.

2013 (SSCI)

-"An adaptive fuzzy-genetic algorithm approach for building detection using high-resolution satellite images", *Computers, Environment and Urban Systems*, Vol.39, pp.48-62.

2014 (SCI)

- “Automated extraction of photorealistic facade textures from single ground-level building images”, International Journal of Pattern Recognition and Artifical Intelligence, (In Revision)