

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?

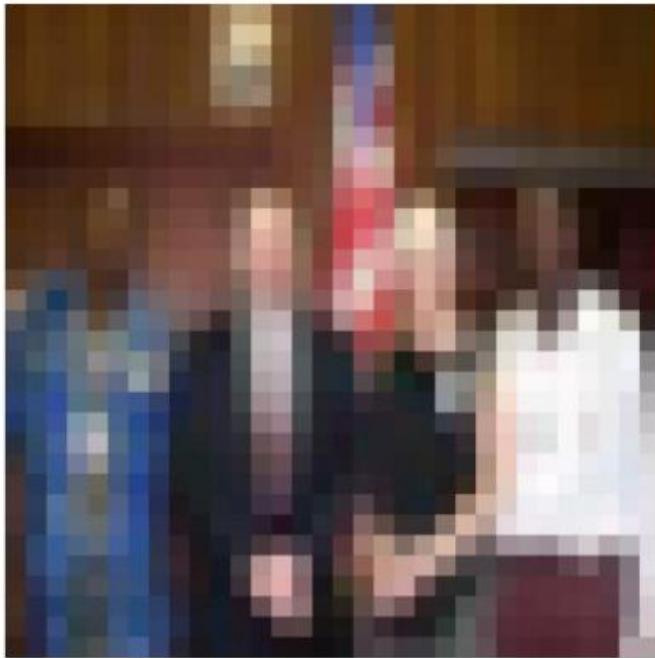
Human perception has its
shortcomings



[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

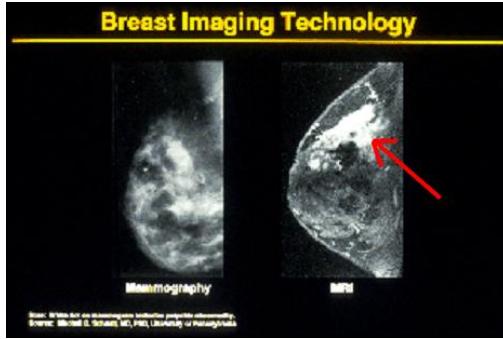
- Vision is an amazing feature of natural intelligence
 - Visual cortex occupies about 50% of Macaque brain
 - More human brain devoted to vision than anything else



Why computer vision matters



Safety



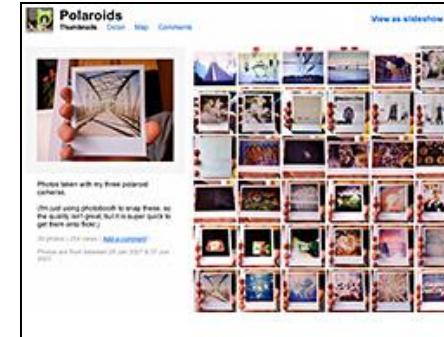
Health



Security



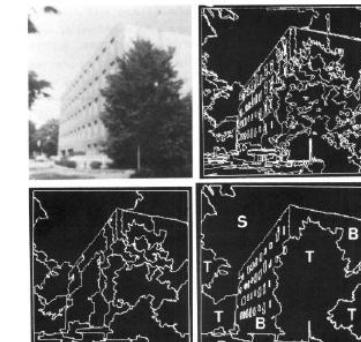
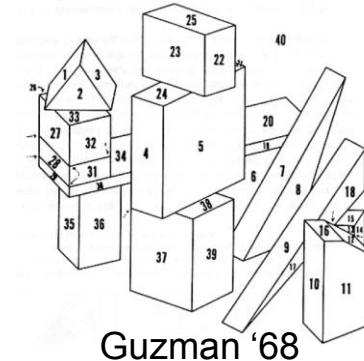
Fun



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?



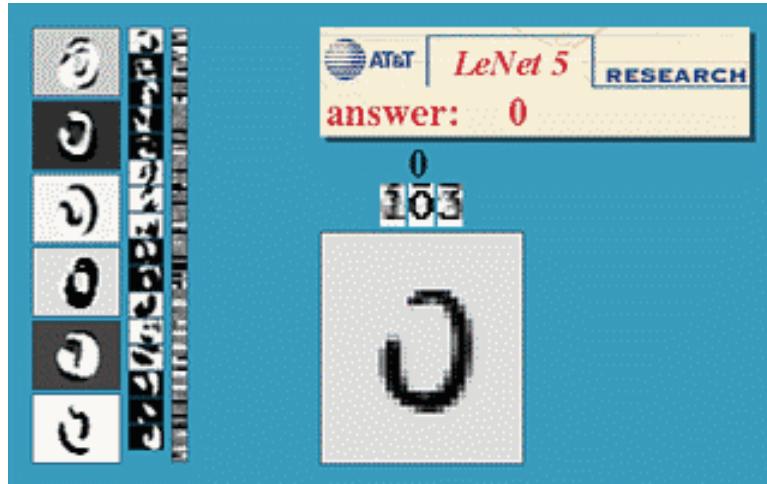
How vision is used now

- Examples of state-of-the-art

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

The image shows a screenshot of the face.com mobile application. At the top left is the face.com logo. Below it is a blue link: <http://developers.face.com/tools/>. The main interface features two photo thumbnails. The left thumbnail shows a man and a woman; the man's name, "Jack", is displayed with a "Learn" button above his head. The right thumbnail shows the same couple, with the woman's name, "Roxanne", written above her head. Below these thumbnails is a row of four small profile pictures labeled "No Filter", "Havok", "Fokus", and "Kross". At the bottom are standard mobile navigation icons: a camera icon, a flash icon, a back arrow, a "NEXT" button, and a "T" button.

Face Confidence Threshold: 50

Attributes:

- age_est: 21 (66%)
- age_min: 18 (66%)
- age_max: 24 (66%)
- face: true (95%)
- gender: male (73%)
- glasses: false (96%)
- lips: sealed (98%)
- mood: angry (69%)
- smiling: false (92%)

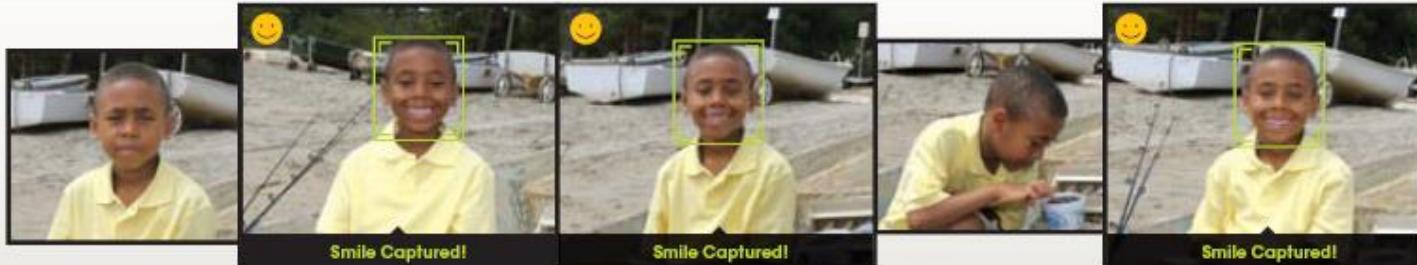
Rotations:

- roll: -7.05°
- yaw: 2.56°
- pitch: -9.27°

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)



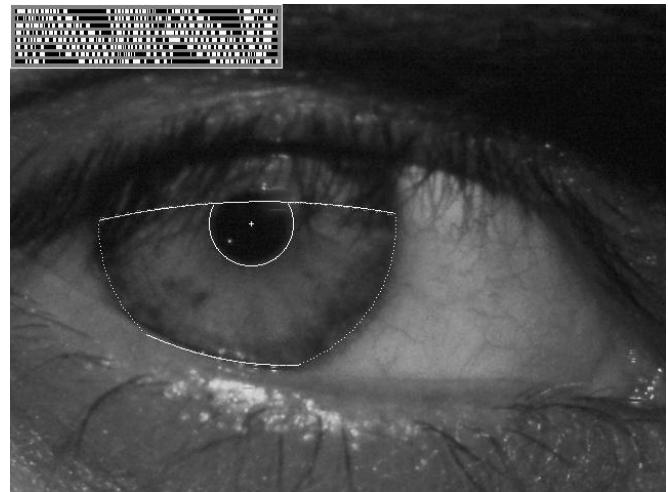
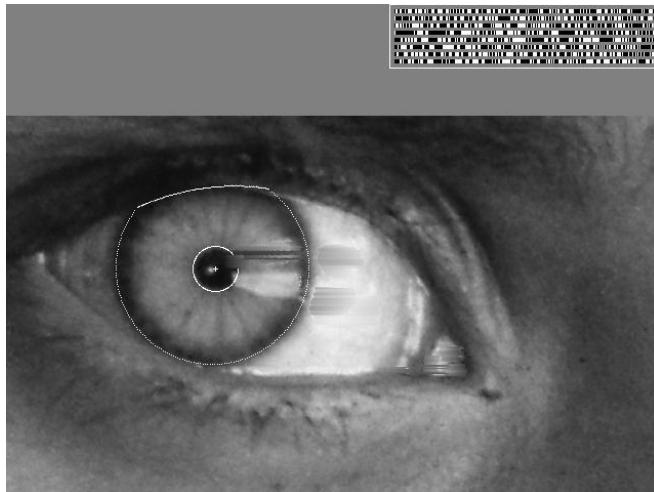
LaneHawk by EvolutionRobotics

“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)

A cartoon illustration of a person with orange hair, wearing a light-colored sweater and pants, standing and holding a small mobile phone. They are pointing the phone towards a movie poster for the film 'Panic Room'. The poster features Jodie Foster's face and the title 'PANIC ROOM' in large red letters. A text box at the top left of the poster says 'MOBILE IMAGE RECOGNITION? TRY IT OUT NOW!!!'. At the bottom left of the poster, there is a link 'Show another poster'.

1. POINT
YOUR MOBILE PHONE CAMERA TO THE MOVIE POSTER.

2. SNAP A PICTURE AND SEND IT:

IN SWITZERLAND:
MMS TO 5555 (OR
079 394 57 00
FOR ORANGE CUSTOMERS)

IN GERMANY:
MMS TO 84000

EVERYWHERE:
EMAIL TO
M@KOOABA.COM

3. FIND ALL RELEVANT INFORMATION ABOUT THE MOVIE ON YOUR MOBILE PHONE

Movie data provided by:

Object recognition (in mobile phones)

leafsnap

Home Species Collectors About

Leaf of the Bottlebrush Buckeye

Leafsnap: An Electronic Field Guide

Leafsnap is the first in a series of electronic field guides being developed by researchers from [Columbia University](#), the [University of Maryland](#), and the [Smithsonian Institution](#). This free mobile app uses visual recognition software to help identify tree species from photographs of their leaves.

Leafsnap contains beautiful high-resolution images of leaves, flowers, fruit, petiole, seeds, and bark. Leafsnap currently includes the trees of the Northeast and will soon grow to include the trees of the entire continental United States.

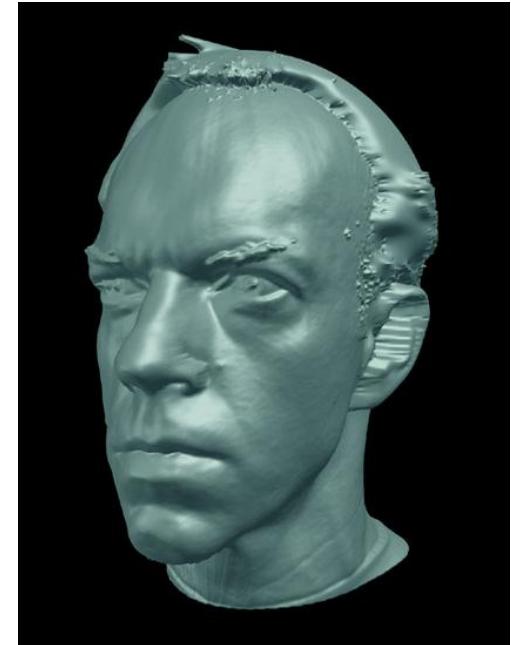
This website shows the tree species included in Leafsnap, the collections of its users, and the team of research volunteers working to produce it.

Free for iPhone:  and iPad: 

 [guardian.co.uk](#)



Special effects: shape capture



The Matrix movies, ESC Entertainment, XYZRGB, NRC

Special effects: motion capture



Pirates of the Caribbean, 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

The screenshot shows the Mobileye website's homepage. At the top, there are navigation tabs: 'manufacturer products' (with a right arrow), 'consumer products' (with left and right arrows), and 'News'. Below the tabs, the slogan 'Our Vision. Your Safety.' is displayed. A central image shows a car from above with three camera systems highlighted: 'rear looking camera' (top left), 'forward looking camera' (top right), and 'side looking camera' (bottom). Below this, there are three main sections: 'EyeQ Vision on a Chip' featuring a chip image, 'Vision Applications' showing a pedestrian crossing, and 'AWS Advance Warning System' showing a dashboard display. To the right, a sidebar titled 'News' lists articles like 'Mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System' and 'Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end'. Another sidebar titled 'Events' lists 'Mobileye at Equip Auto, Paris, France' and 'Mobileye at SEMA, Las Vegas, NV'.

- > **EyeQ** Vision on a Chip  [read more](#)
- > **Vision Applications**  Road, Vehicle, Pedestrian Protection and more [read more](#)
- > **AWS** Advance Warning System  [read more](#)

News

- > **Mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System**
- > **Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end**

[all news](#)

Events

- > **Mobileye at Equip Auto, Paris, France**
- > **Mobileye at SEMA, Las Vegas, NV**

[read more](#)

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

Google cars



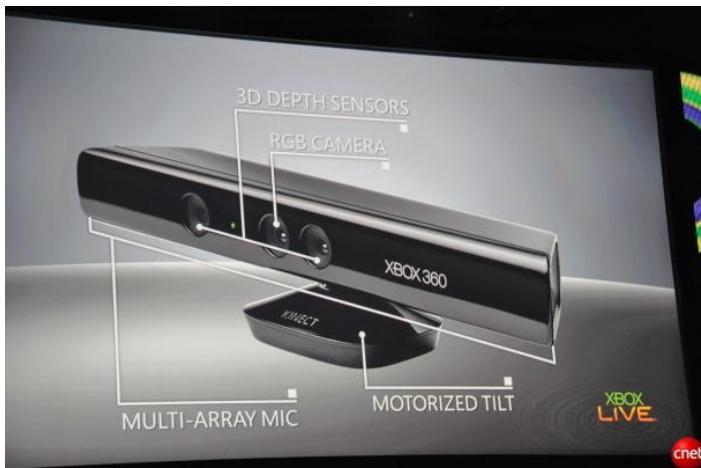
Oct 9, 2010. "[Google Cars Drive Themselves, in Traffic](#)". *The New York Times*. John Markoff

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=8CTJL5IUjHg>
- 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

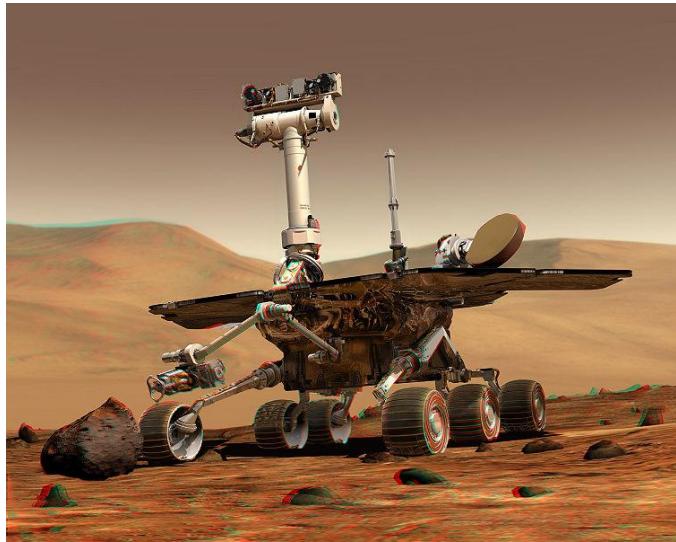
- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read “[Computer Vision on Mars](#)” by Matthies et al.

Industrial robots



Vision-guided robots position nut runners on wheels

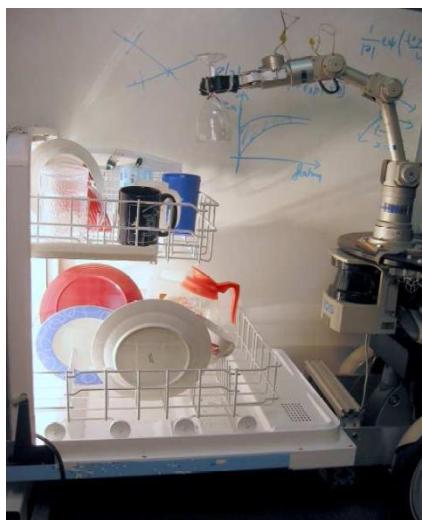
Mobile robots



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

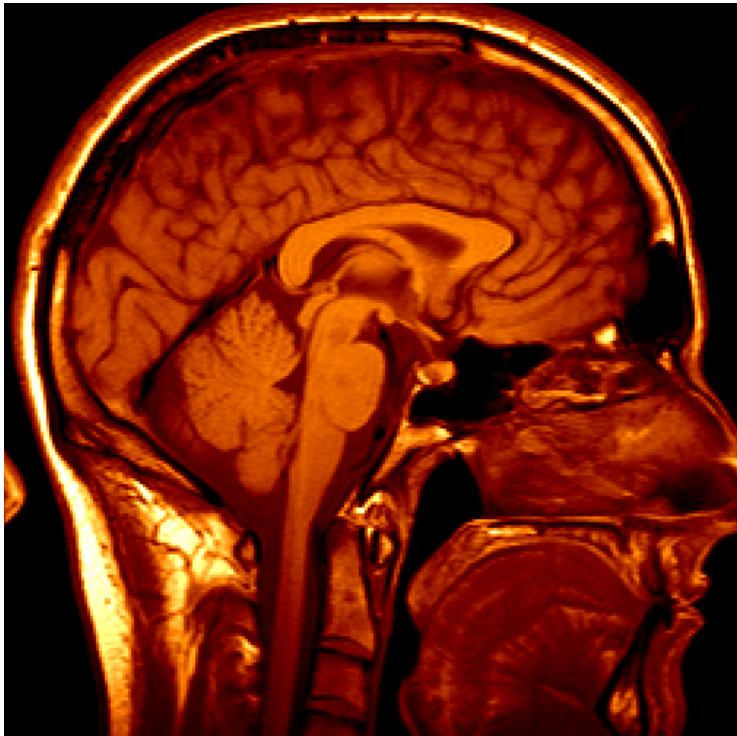


<http://www.robocup.org/>



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

Scope of Bil-613

Image Processing
Feature Matching
Recognition

Machine Learning

Graphics

Computational
Photography

Optics

Robotics

Human Computer
Interaction

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.

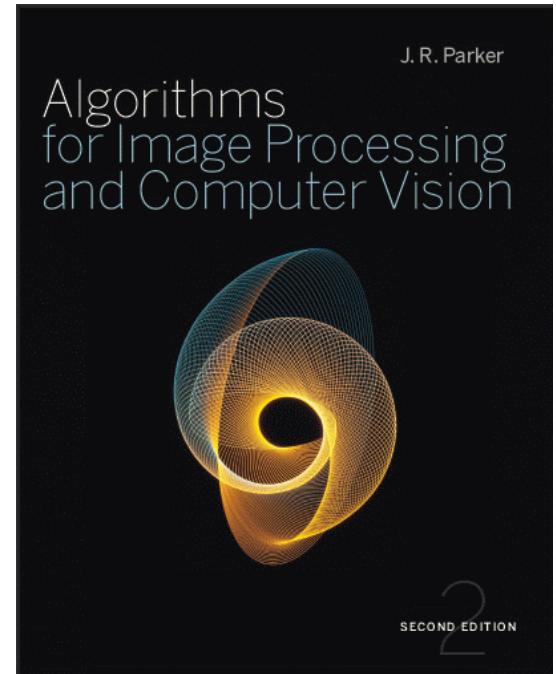
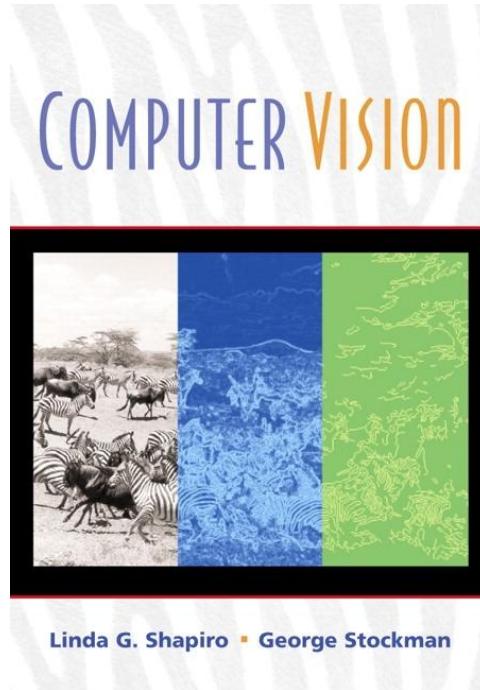
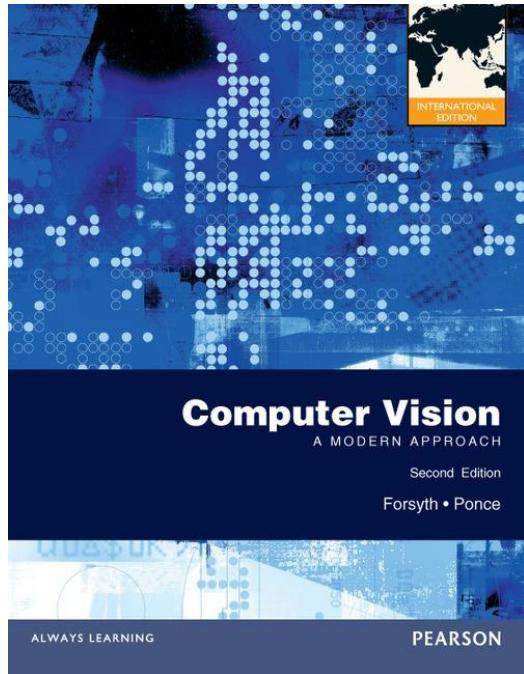
<u>Instructor</u>	<u>Hours</u>	<u>Grading</u>
Name: Emre SÜMER, PhD. Office: D-414 Phone: 246 6666 / 1305 E-mail: esumer@baskent.edu.tr Web: www.baskent.edu.tr/~esumer	Class Day: Thursday Class Hours: 18:00 – 21:00 Location: B-308 Office Hours: Any proper time	Review Paper.....%15 Programming Projects....%15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

Course Topics

Tentative Weekly Course Schedule

- 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

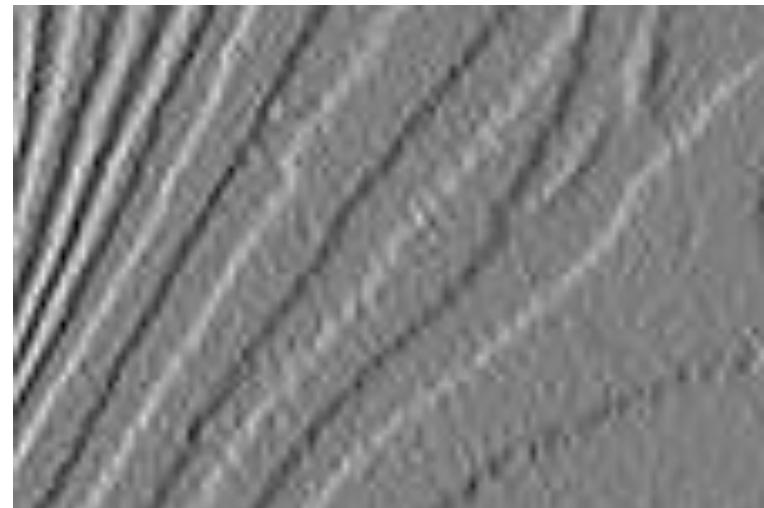
- Filter outputs
 - essentially form a dot-product between a pattern and an image, while shifting the pattern across the image
 - strong response -> image locally looks like the pattern
 - e.g. derivatives measured by filtering with a kernel that looks like a big derivative (bright bar next to dark bar)

Filtering

Convolve this image



To get this



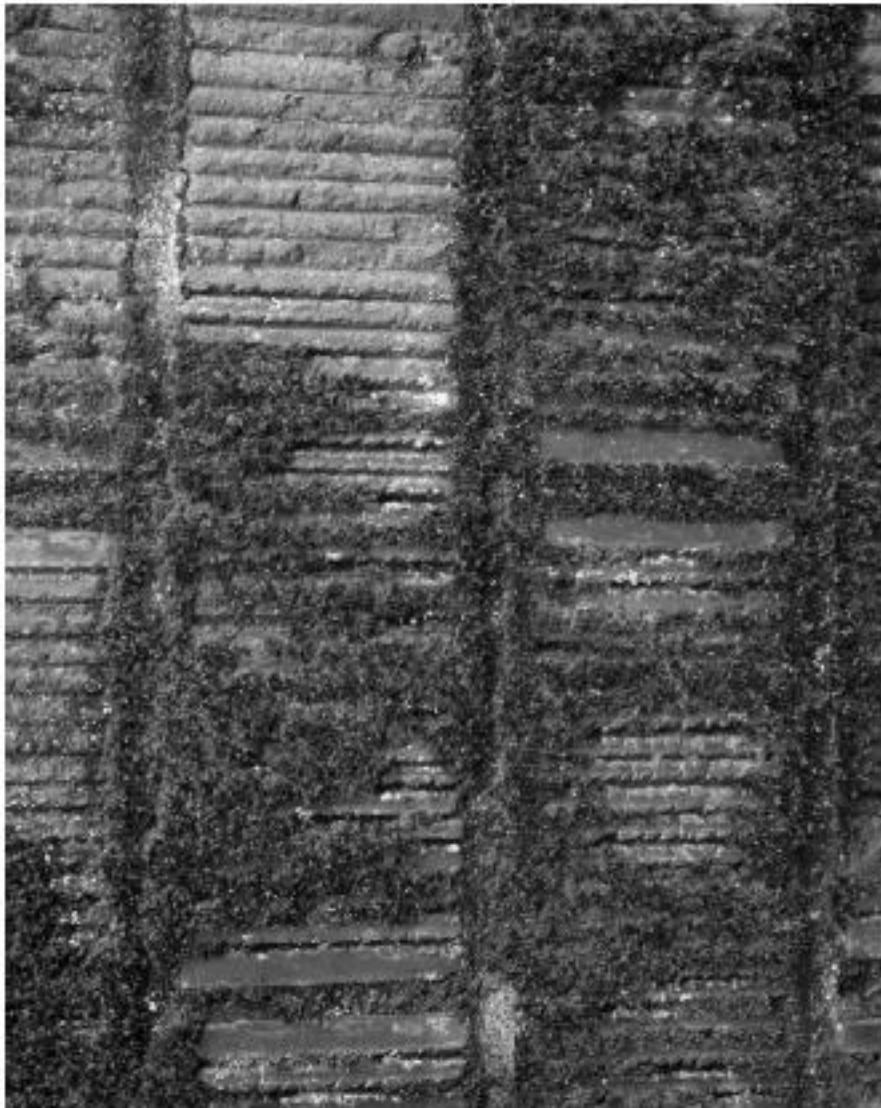
With this kernel



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

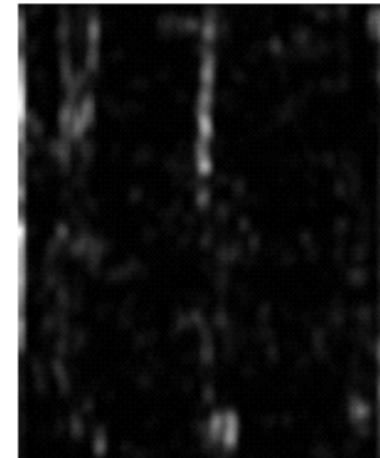
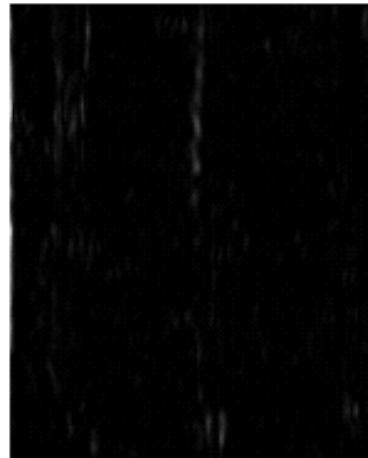
Texture



Texture

squared responses

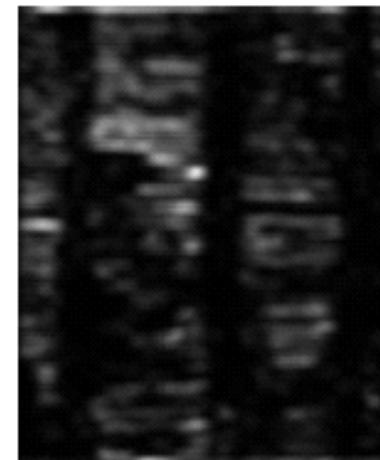
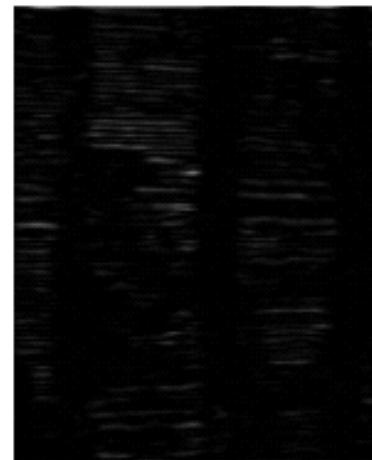
vertical



classification



horizontal



smoothed mean

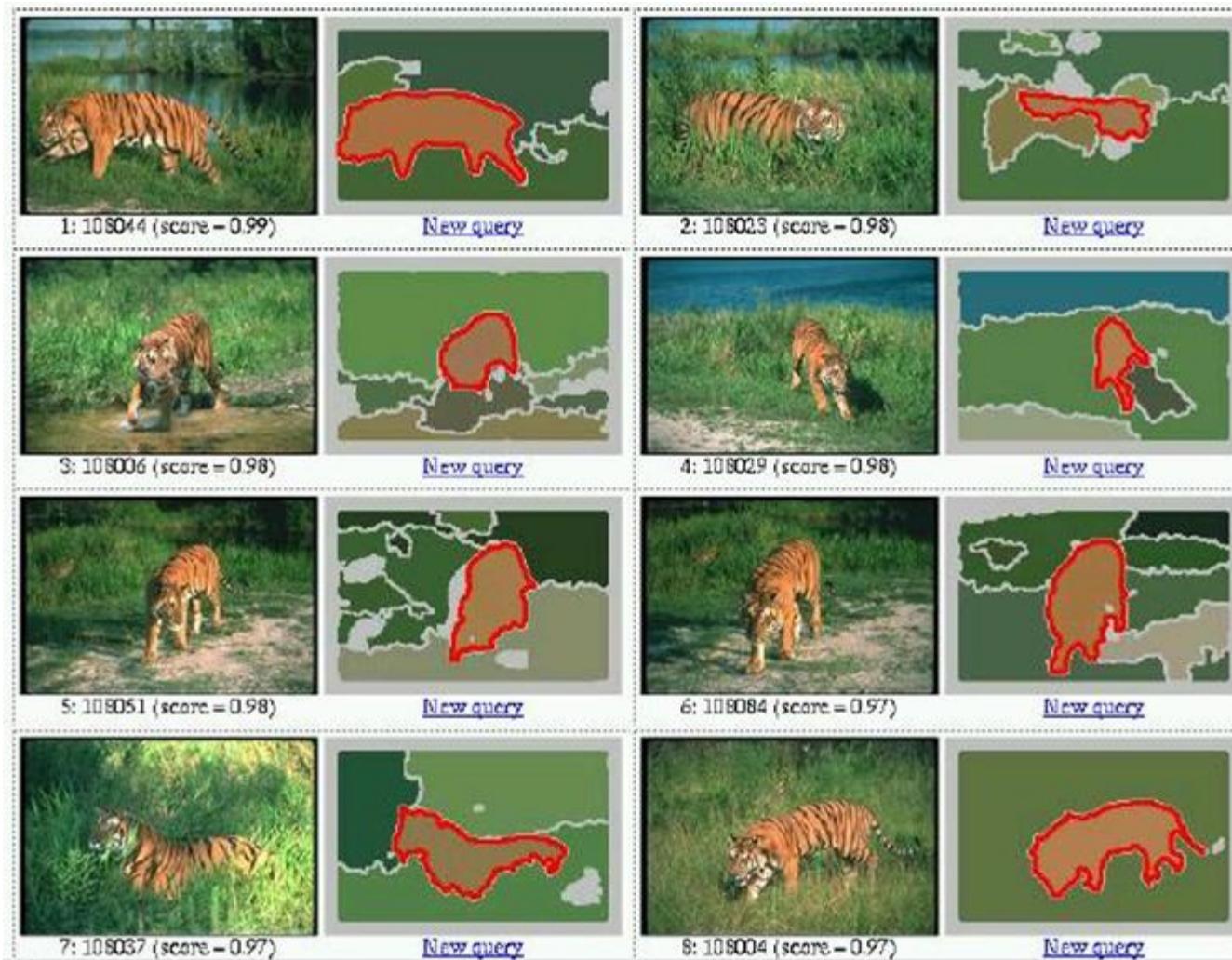
Mid-Level Vision

- 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



Classifiers



My Research Interests

Earthquake Damage Detection using Image Processing Techniques and GIS (MS Thesis)



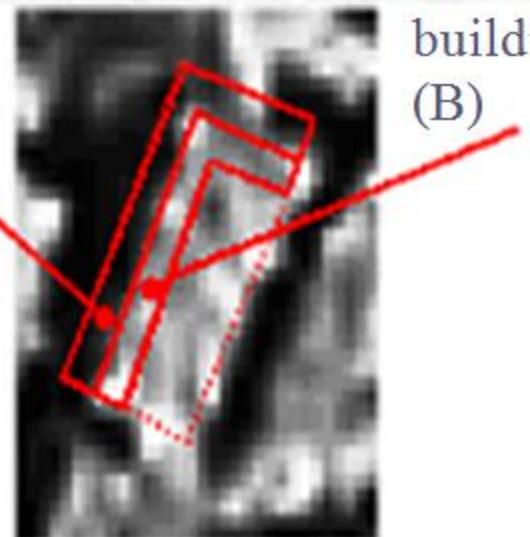
- * A post-earthquake panchromatic aerial imagery
- * Vector building boundaries

My Research Interests

Earthquake Damage Detection using Image Processing Techniques and GIS (MS Thesis)

Buffer zone
outside the
building area
(S)

Buffer zone
inside the
building area
(B)

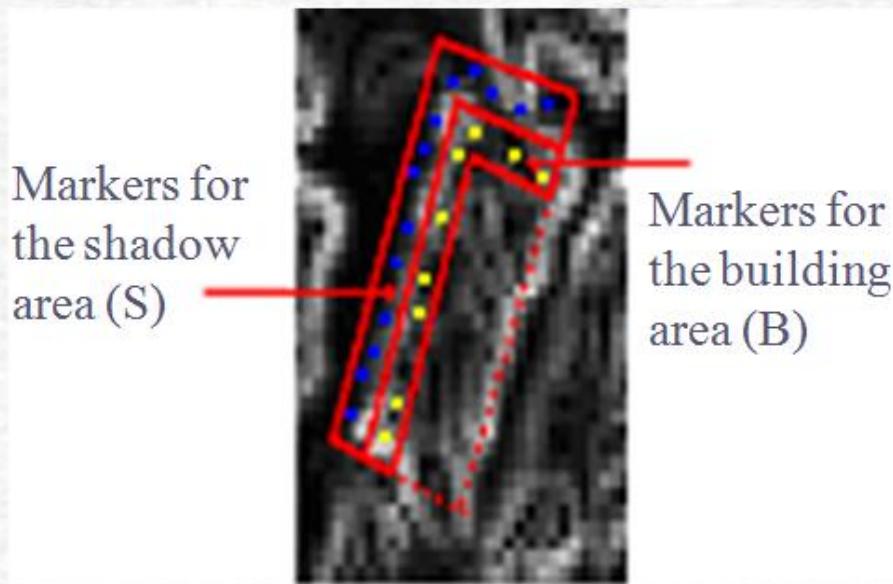


A buffer zone was generated along the shadow producing edges

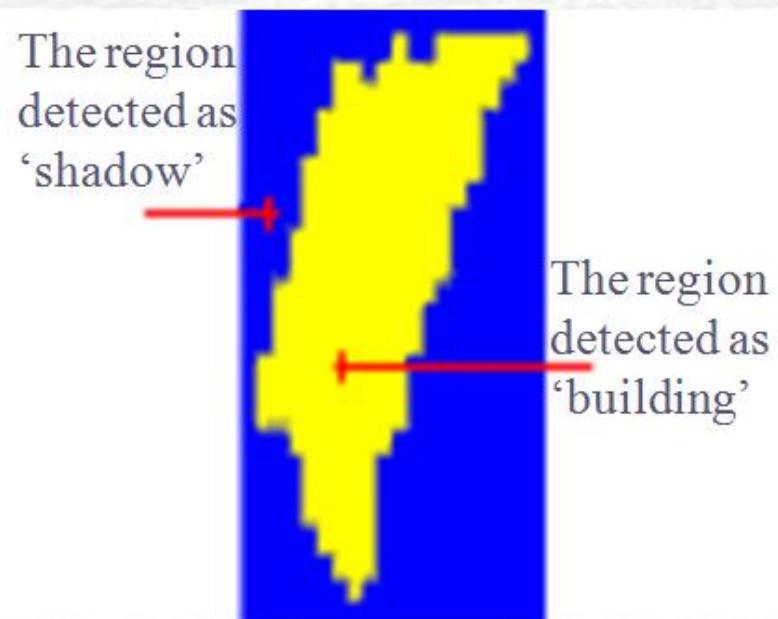
My Research Interests

Earthquake Damage Detection using Image Processing Techniques and GIS (MS Thesis)

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



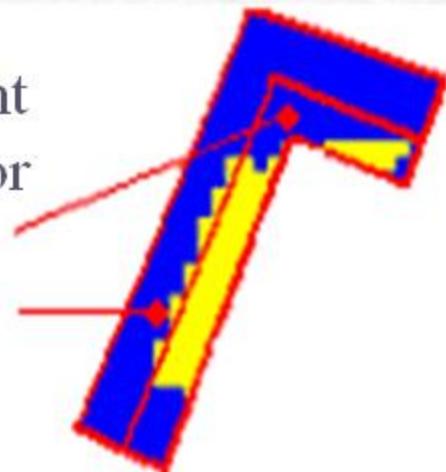
- A two-colored output image was generated



My Research Interests

Earthquake Damage Detection using Image Processing Techniques and GIS (MS Thesis)

Significant regions for building analysis

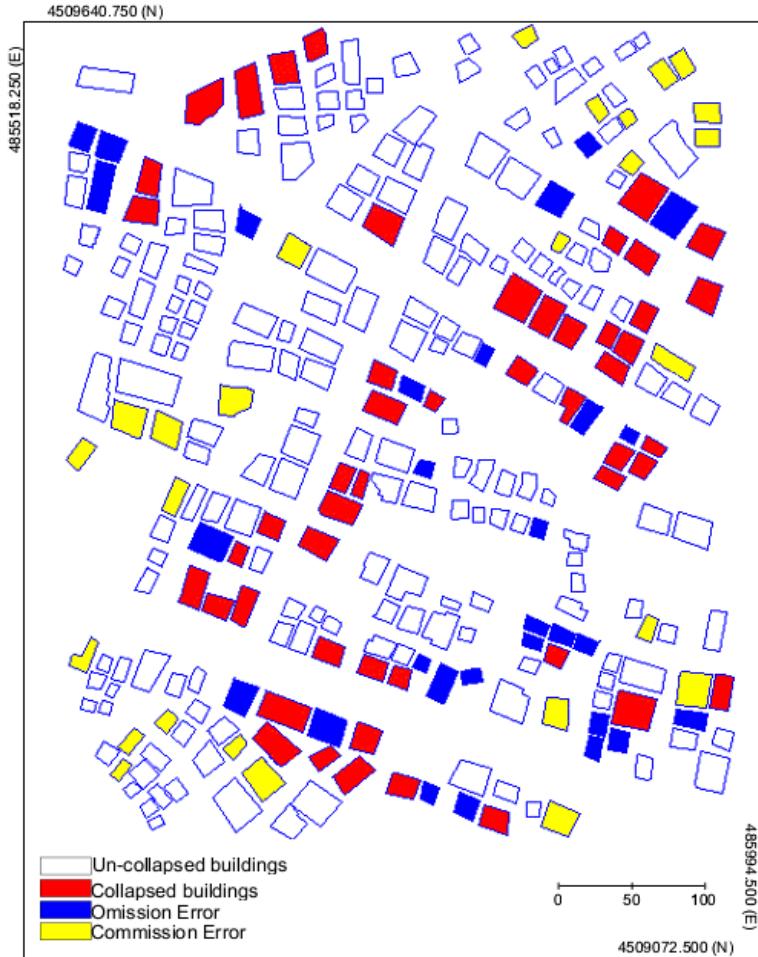


- Shadow and building pixels were counted
- Ratios were computed for building and shadow pixels

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 Detection using Image Processing Techniques and GIS (MS Thesis)



My Research Interests

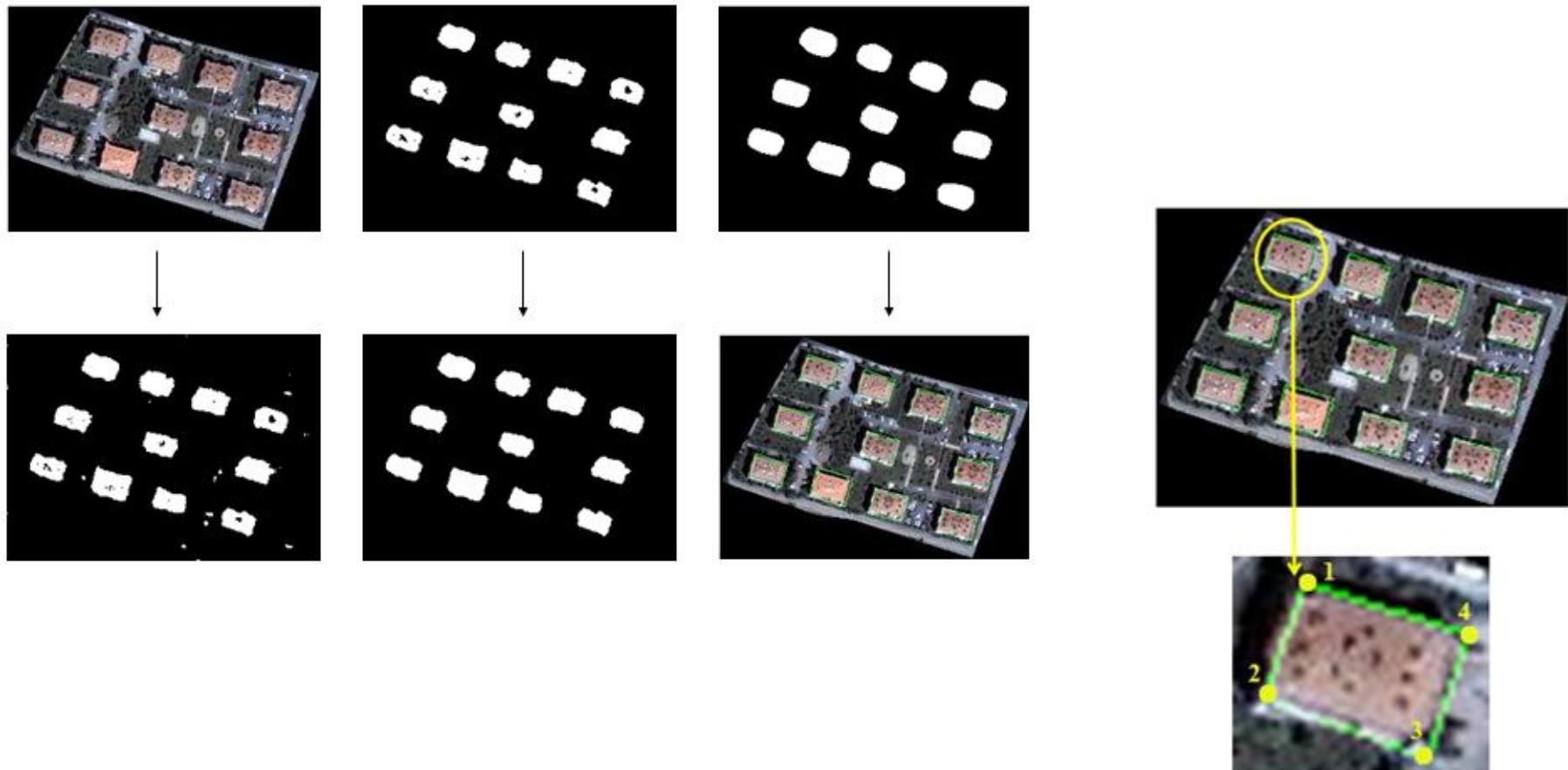
Reconstruction of Photorealistic Building Models (PhD Thesis)

- * 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

Ground-level Photo



Extracted Facade



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

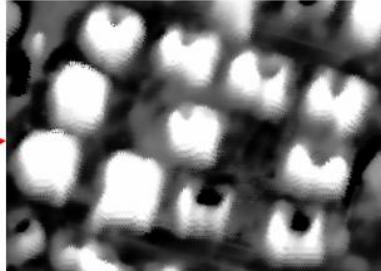
Reconstruction of Photorealistic Building Models (PhD Thesis)

3-D extrusion model generation and visualization

- 1-m resolution pan-sharpened IKONOS image



- nDSM



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 Artificial Intelligence*, (In Revision)