

Computer Vision: Introduction

Raquel Urtasun

TTI Chicago

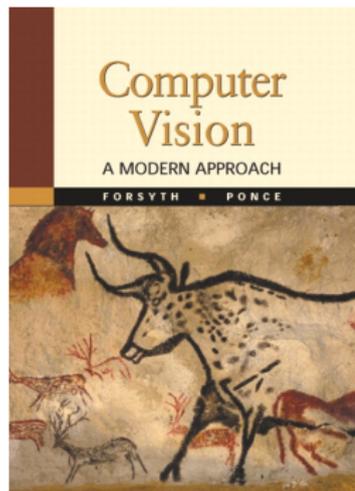
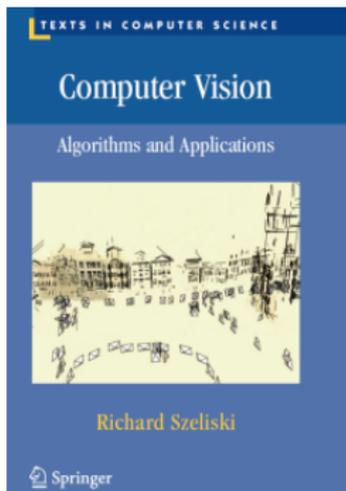
January 8th, 2013

- Instructor: Raquel Urtasun (rurtasun@ttic.edu)
- Lectures: Tuesday and Thursday 10:30-11:50
- Course Webpage:
<http://ttic.uchicago.edu/~rurtasun/courses/CV/cv.html>
- TA: TBD

Other useful info:

Materials:

- Rick Szeliski, Computer Vision: Algorithms and Applications, <http://szeliski.org/Book/>
- David Forsyth and Jean Ponce, Computer Vision: A Modern Approach



- Lot's of papers

Course requirements

Essential Prerequisites:

- Linear algebra
- Vector calculus
- Programming

Course does not knowledge about:

- Computer vision
- Image processing
- Graphics
- Robotics

- 1 Introduction to computer vision
- 2 Course overview

Readings for today

- Szeliski, CV: A&A, Ch 1.0 (Introduction)

Every image tells a story



- Goal of computer vision: perceive the story behind the picture
- Compute properties of the world
 - 3D shape and appearance
 - Names of people or objects
 - Track a person moving against a complex background
 - What happened?

[Source: N. Snavely]

The goal of Computer Vision



0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

[Source: N. Snavely]

Why is vision so difficult?

- It is an **inverse problem**: recover some unknowns given insufficient information to fully specify the solution
- In general to disambiguate between solutions we resort to
 - physics-based model, e.g., geometry, light
 - probabilistic models

Why is vision so difficult?

- It is an **inverse problem**: recover some unknowns given insufficient information to fully specify the solution
- In general to disambiguate between solutions we resort to
 - physics-based model, e.g., geometry, light
 - probabilistic models
- These are two different schools that are typically in conflict

Why is vision so difficult?

- It is an **inverse problem**: recover some unknowns given insufficient information to fully specify the solution
- In general to disambiguate between solutions we resort to
 - physics-based model, e.g., geometry, light
 - probabilistic models
- These are two different schools that are typically in conflict
- The future is in unifying both ... in my opinion

Why is vision so difficult?

- It is an **inverse problem**: recover some unknowns given insufficient information to fully specify the solution
- In general to disambiguate between solutions we resort to
 - physics-based model, e.g., geometry, light
 - probabilistic models
- These are two different schools that are typically in conflict
- The future is in unifying both ... in my opinion

... are usually developed in **physics** (radiometry, optics, and sensor design) and **graphics**, modeling

- how objects move and animate,
- how light reflects off their surfaces,
- is scattered by the atmosphere,
- refracted through camera lenses (or human eyes),
- and finally projected onto a flat (or curved) image plane

In computer vision, we

- describe the world that we see in one or more images
- and try to reconstruct its properties
 - shape
 - illumination
 - color distributions
 - ...

Can the computer match human perception?

Yes and no (but mainly no, so far)

- computers can be better at some easy things
- humans are much better at hard things

The notion of "hardness" is different for human and machine. Examples?

Is computer vision hard?

Is computer vision hard?

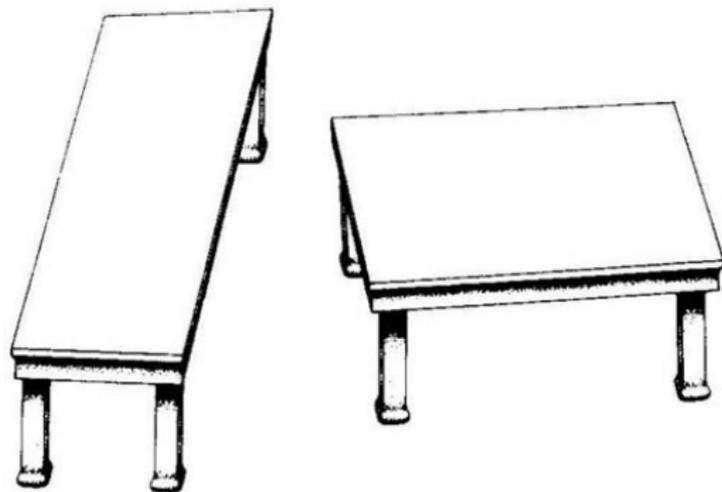
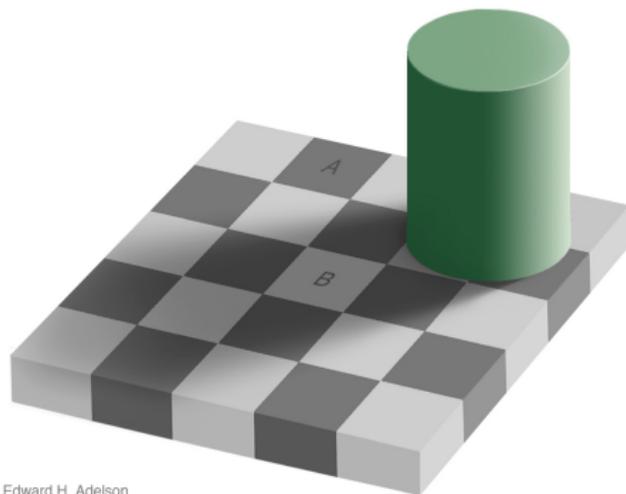


Figure: Turning the Tables by Roger Shepard

- Depth processing is automatic, and we can not shut it down

[Source: A. Torralba]

Is computer vision hard?

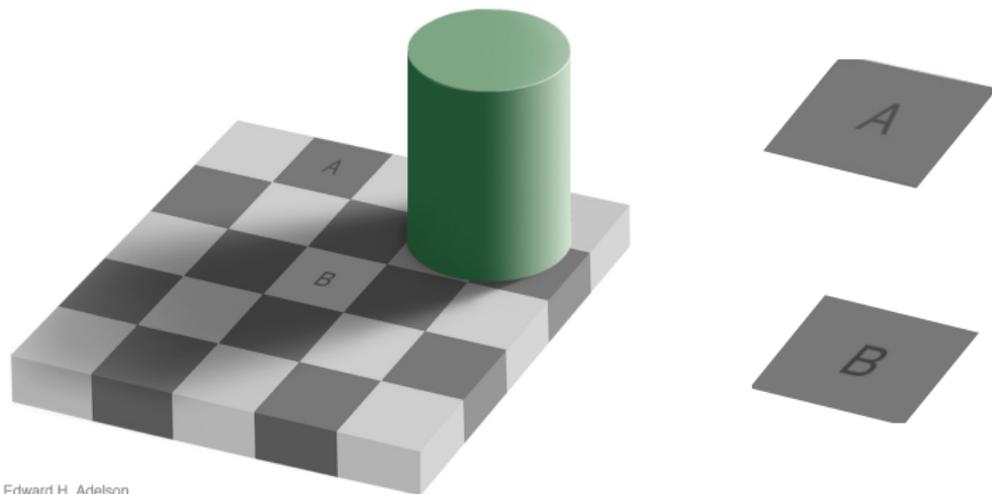


Edward H. Adelson

- Do A and B have the same gray level?

[Source: A. Torralba]

Is computer vision hard?



Edward H. Adelson

- Do A and B have the same gray level?

[Source: A. Torralba]

Is computer vision hard?

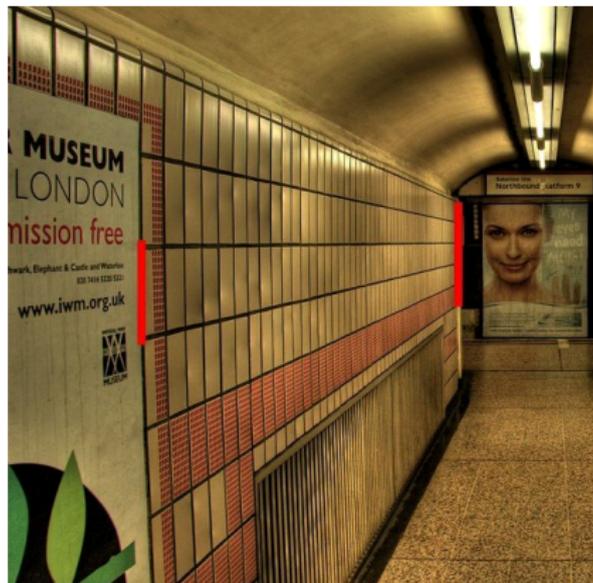


Figure: 2006 Walt Anthony

- Do they have the same length?

[Source: A. Torralba]

Is computer vision hard?

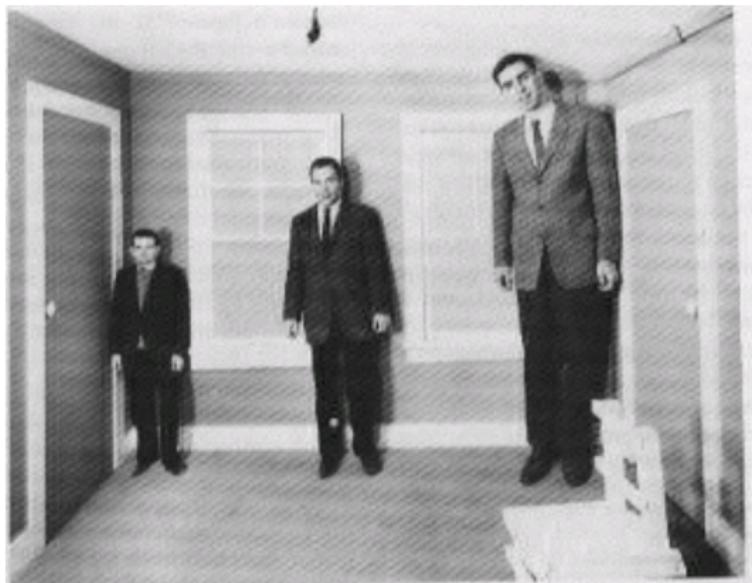


Figure: Ames room

- Assumptions can be wrong

[Source: A. Torralba]

Is computer vision hard?



Figure: Chabris & Simons

- Count number of times the white team pass the ball
- Concentrate, difficult task!

Is computer vision hard?



Figure: Simons et al.

- Is something happening in the picture?

A bit of history ...

The beginning of Computer Vision ...

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

[Source: A. Torralba]

Vision is hard ...

So let's make the problem more simple

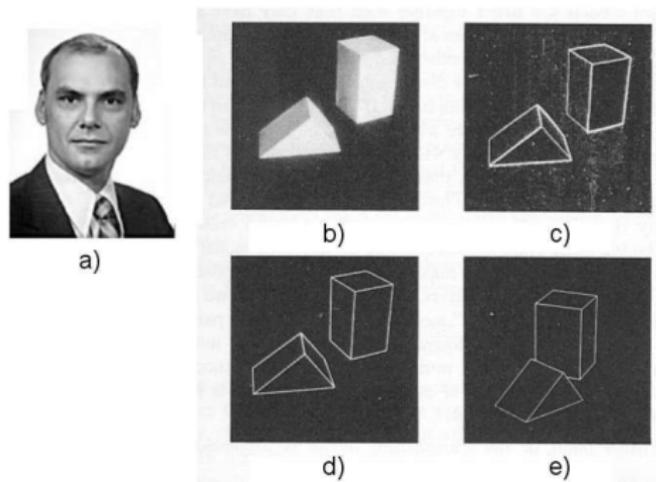


Fig. 1. A system for recognizing 3-d polyhedral scenes. a) L.G. Roberts. b) A blocks world scene. c) Detected edges using a 2×2 gradient operator. d) A 3-d polyhedral description of the scene, formed automatically from the single image. e) The 3-d scene displayed with a viewpoint different from the original image to demonstrate its accuracy and completeness. (b) - e) are taken from [64] with permission MIT Press.)

[Source: A. Torralba]

Vision is hard ...

- Initial focus on geometry.
- But, despite promising initial results, things did not work out so well for recognition (lack of data, processing power, lack of reliable methods for low-level and mid-level vision)
- Instead, a different way of thinking about object detection started making some progress: learning based approaches and classifiers, which ignored low and mid-level vision.
- Maybe the time is here to come back to some of the earlier models, more grounded in intuitions about visual perception

[Source: A. Torralba]

But humans are pretty good at it

Recognition even from tiny images

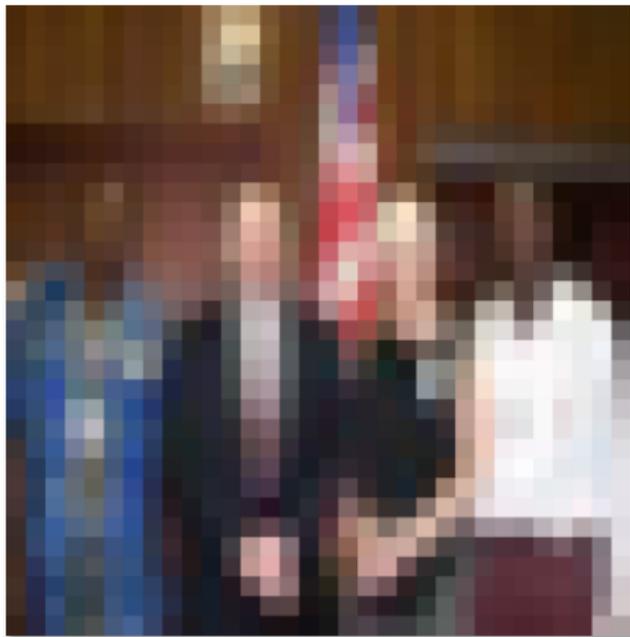


Figure: 80 million tiny images [Torralba et al.]

[Source: N. Snavely]

The goals of computer vision

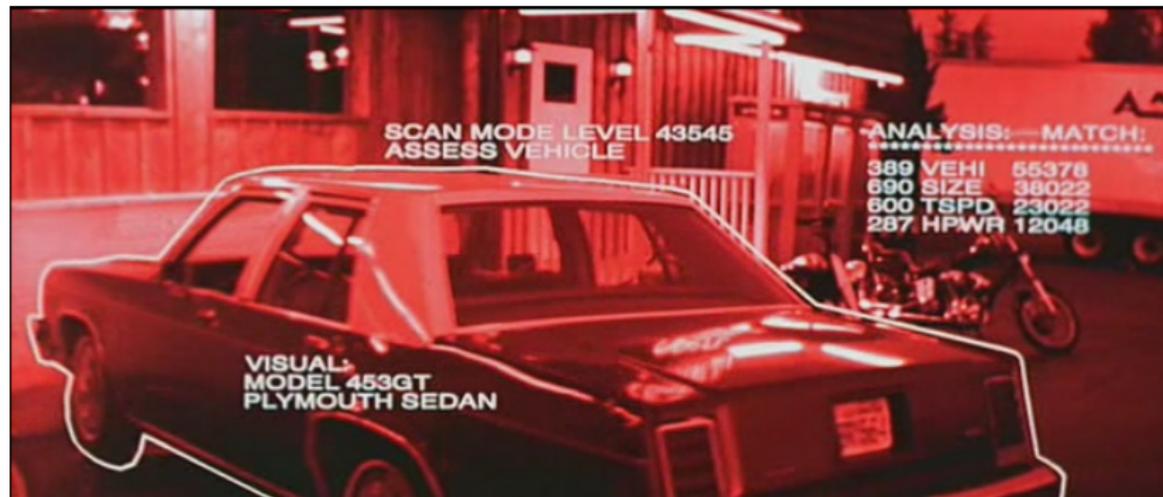
- Computing the 3D shape of the world



[Source: N. Snavely]

The goals of computer vision

- Recognizing objects and people



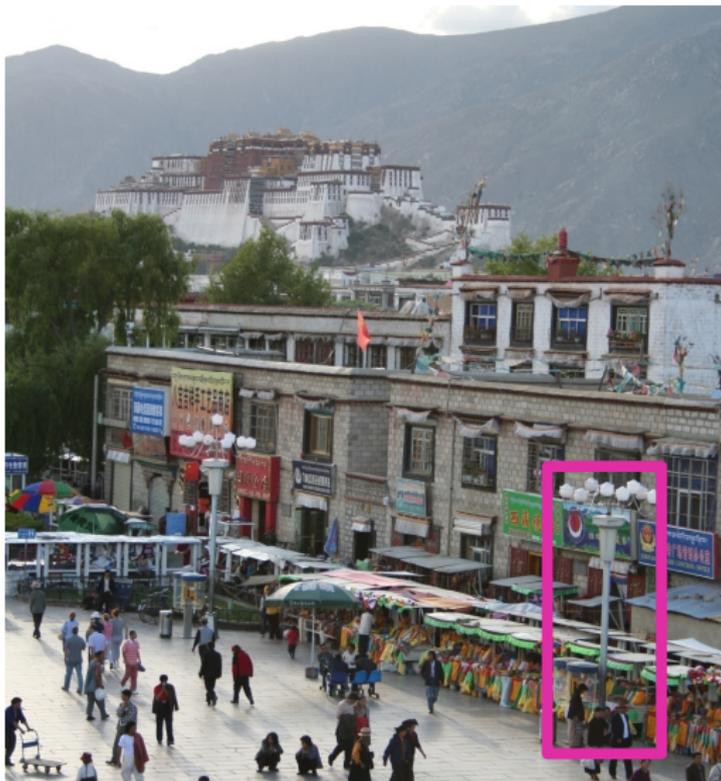
[Source: N. Snavely]

What does visual recognition involve?



[Source: R. Fergus]

Verification: Is that a lamp?



Detection: Where are the people?



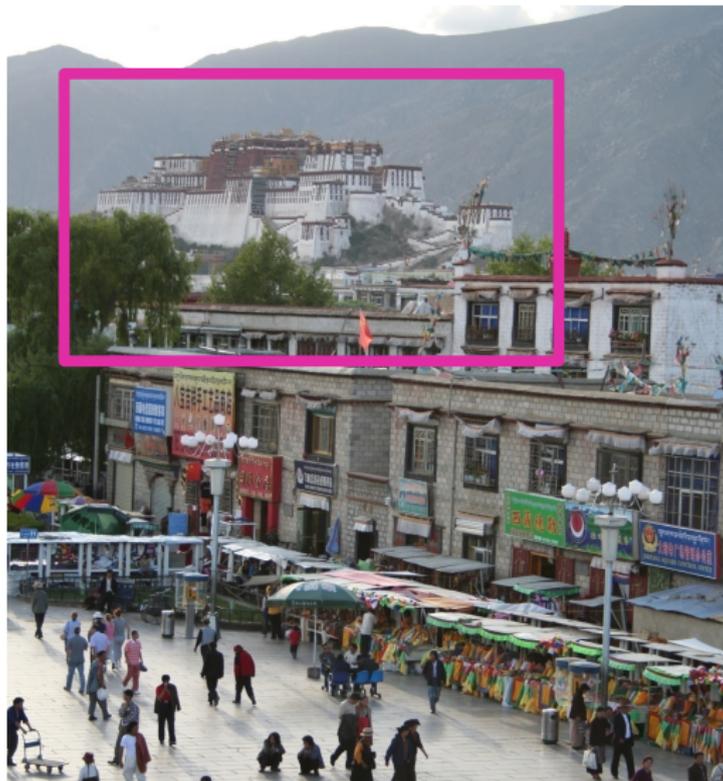
Activity Recognition: What are they doing?



Pose: Which pose do they have?

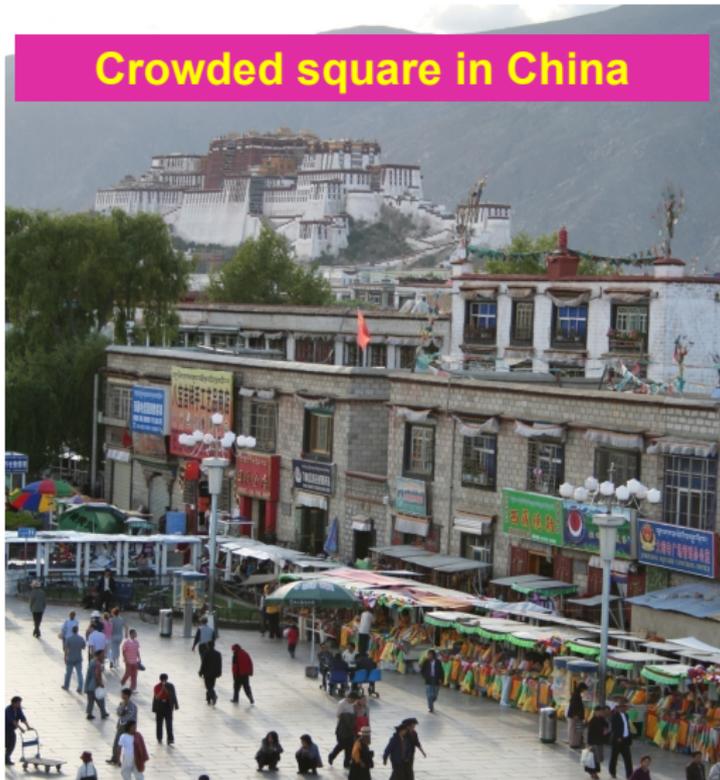


Identification: Is that the great wall?



Description: Attributes and relations

Crowded square in China



The goals of computer vision

- Enhancing images (c.f. Computational Photography)

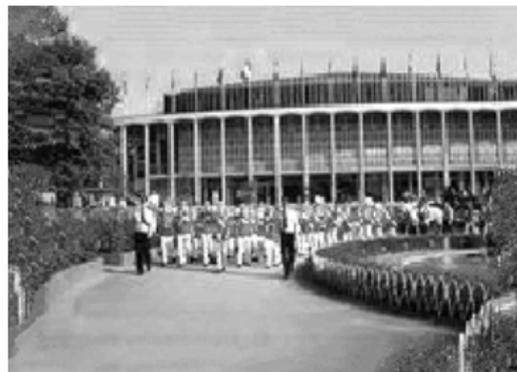


Super-resolution / denoising
(source: 2d3)

[Source: N. Snavely]

The goals of computer vision

- Enhancing images (c.f. Computational Photography)

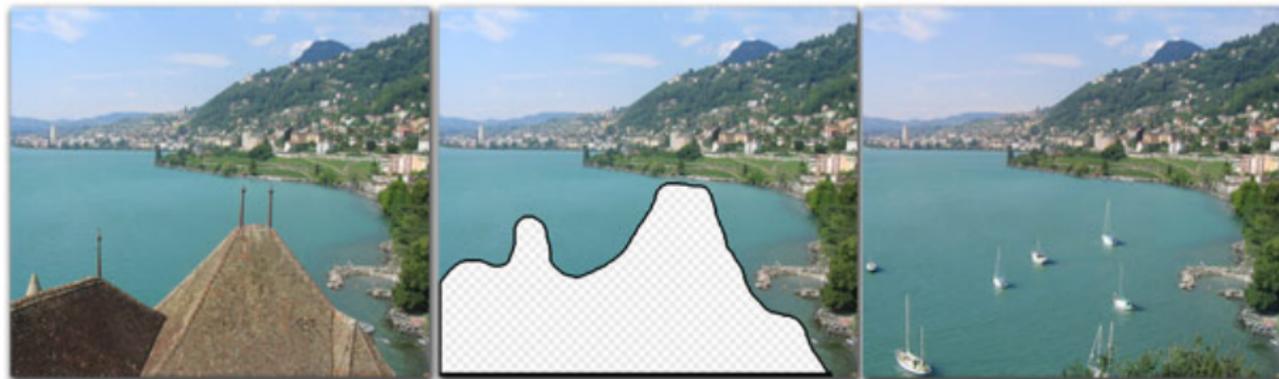


Texture synthesis / increased field of view (uncropping)
(image credit: Efros and Leung)

[Source: N. Snavely]

The goals of computer vision

- Enhancing images (c.f. Computational Photography)



Inpainting / image completion
(image credit: Hays and Efros)

[Source: N. Snavely]

The goals of computer vision

- Forensics



Figure: Source: Nayar and Nishino, Eyes for Relighting

[Source: N. Snavely]

The goals of computer vision



Source: Nayar and Nishino, "Eyes for Relighting"

[Source: N. Snavely]

The goals of computer vision



Figure: Source: Nayar and Nishino, Eyes for Relighting

[Source: N. Snavely]

Why study computer vision?

- Millions of images being captured all the time



Google™
Image Search

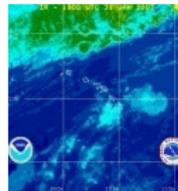
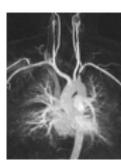
Picasa™

flickr™

webshots™

picsearch™

YouTube
Broadcast Yourself™



- Lots of useful applications
- The next slides show the current state of the art

[Source: S. Lazebnik]

Why study computer vision?

**UNITED STATES
SECURITIES AND EXCHANGE COMMISSION**
Washington, D.C. 20549

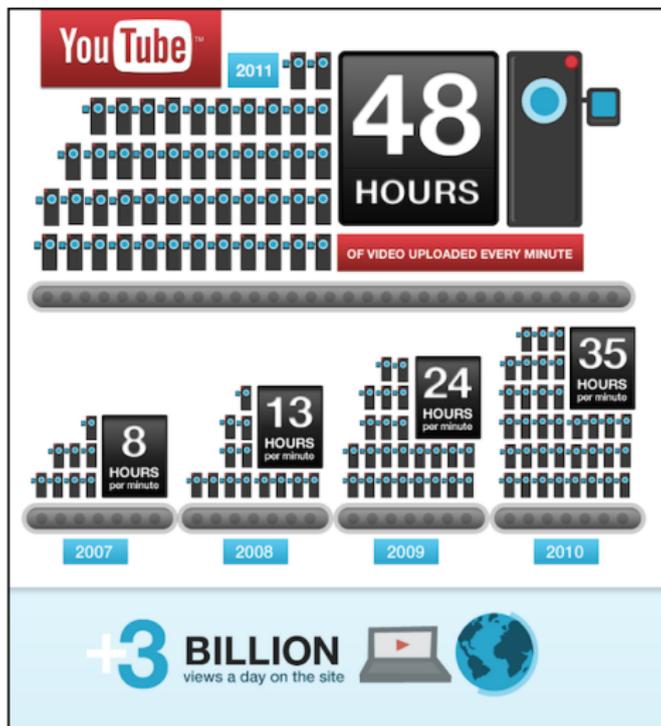
**Amendment No. 4 to
Form S-1
REGISTRATION STATEMENT**
*Under
The Securities Act of 1933*

Facebook, Inc.

**... On average more than 300 million photos per day were uploaded
to Facebook in the three months ended March 31, 2012 ...**

[Source: N. Snavey]

Why study computer vision?



<http://youtube-global.blogspot.com/2011/05/thanks-youtube-community-for-two-big.html>

[Source: N. Snavely]

Why study computer vision?

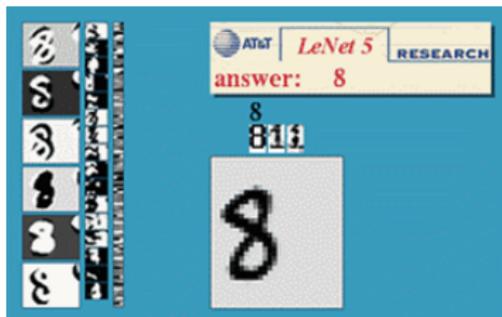


[Source: N. Snavely]

Let's look at some applications

Applications: Optical character recognition (OCR)

- 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



Automatic check processing



Sudoku grabber

<http://sudokugrab.blogspot.com/>

Source: S. Seitz

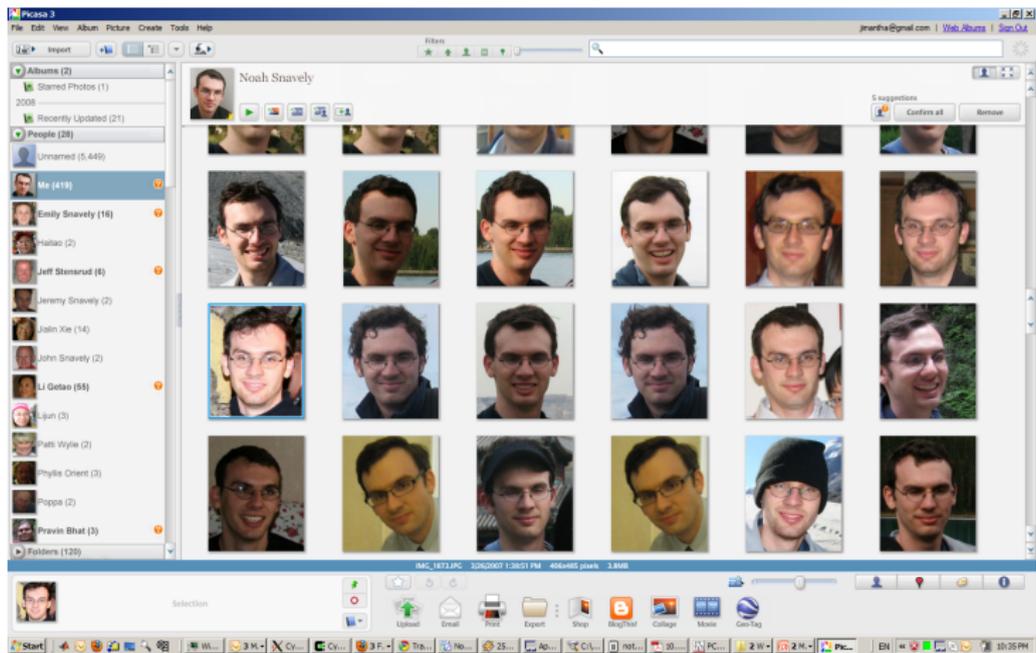
Applications: Face detection

- Many new digital cameras now detect faces: e.g., Canon, Sony, Fuji,



[Source: S. Seitz]

Applications: Face recognition

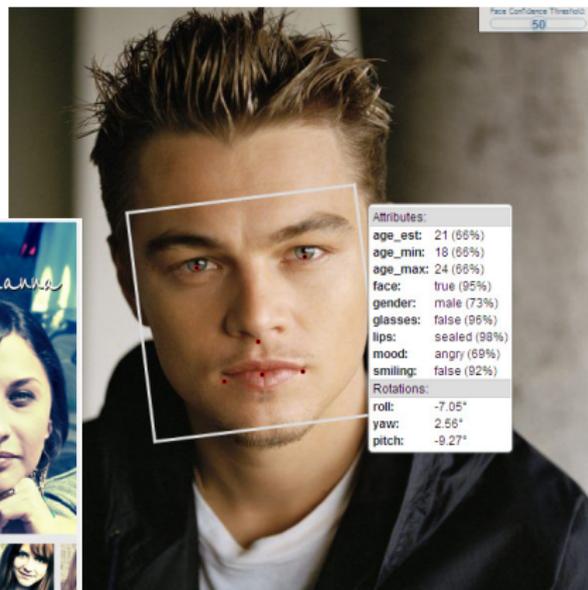
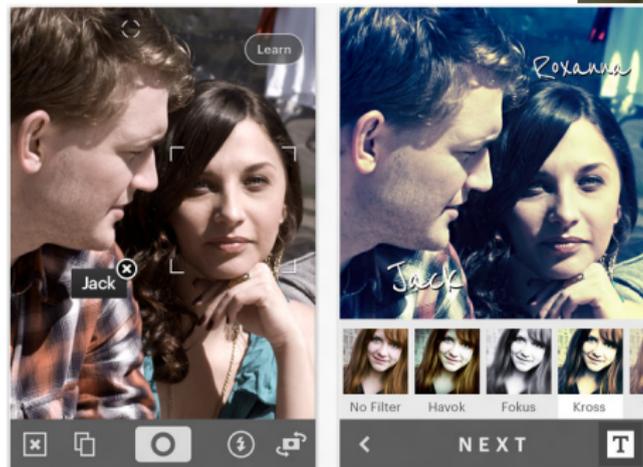


[Source: N. Snaveley]

Applications: Face recognition

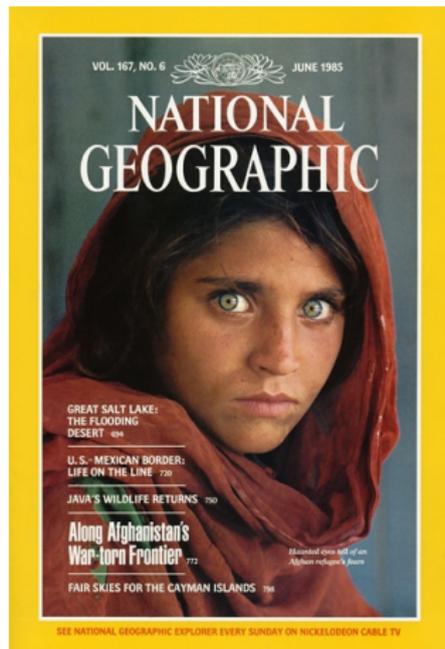


<http://developers.face.com/tools/>



[Source: N. Snavey]

Applications: Face Recognition

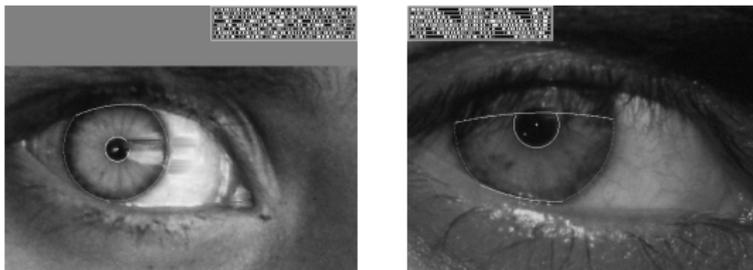


Who is she?

Applications: Vision-based biometrics



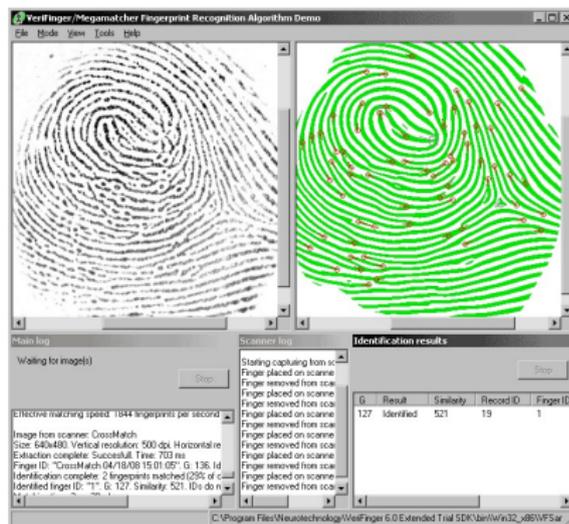
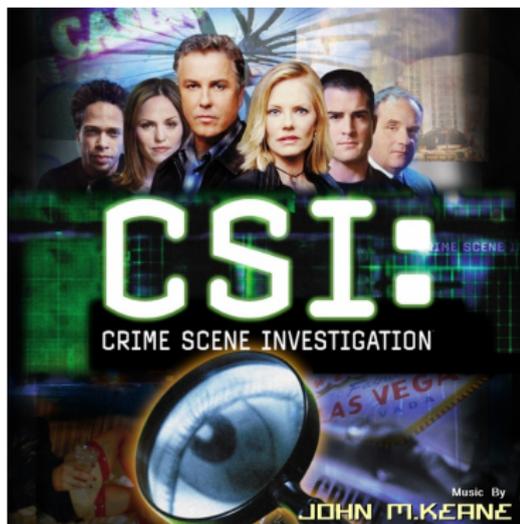
"How the Afghan Girl was Identified by Her Iris Patterns" Read the [story](#).



Source: S. Seitz

Click for the story

Applications: Fingerprint Recognition

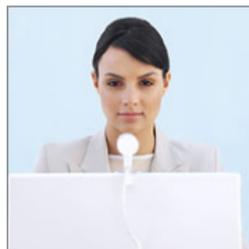


[Source: S. Lazebnik]

Applications: 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/>

Source: S. Seitz

Applications: 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... “

Source: S. Seitz

Applications: Object recognition (in mobile phones)



[Source: S. Seitz]

iPhone Apps: kooaba (www.kooaba.com)

MOBILE IMAGE RECOGNITION
TRY IT OUT NOW!!!



[Show another poster](#)

Movie data provided by:



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

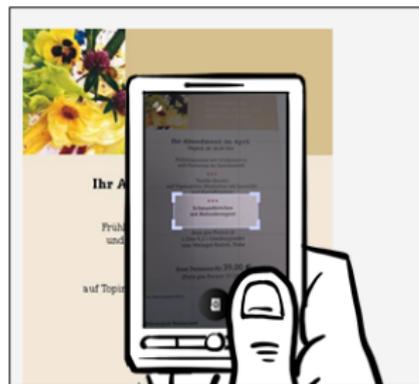
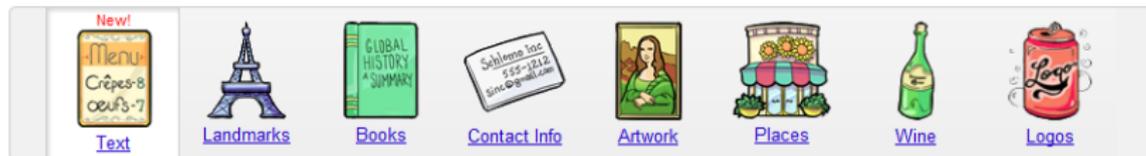
3. **FIND** ALL
RELEVANT INFOR-
MATION ABOUT THE
MOVIE ON YOUR
MOBILE PHONE

Source: S. Lazebnik

Applications: Google Goggles

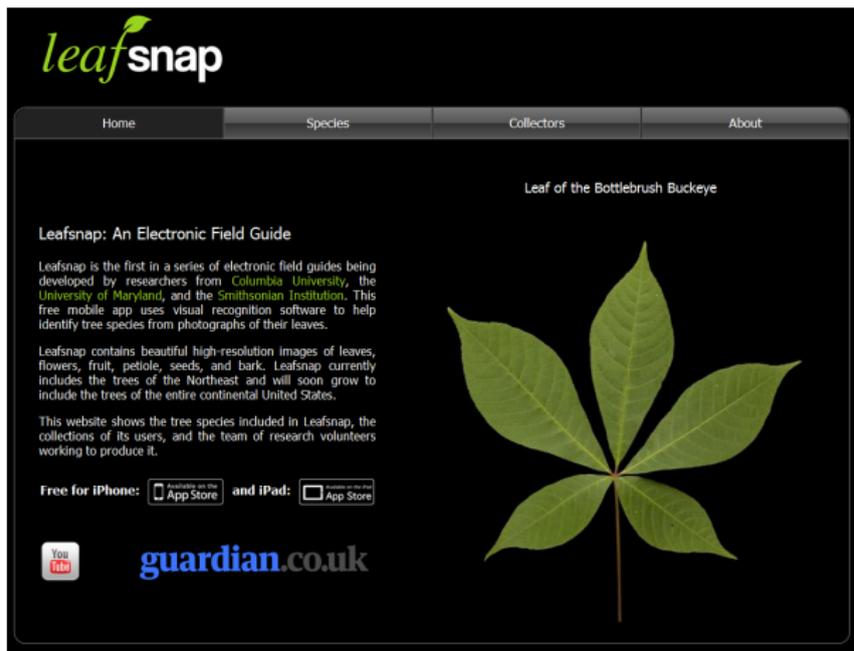
Google Goggles in action

Click the icons below to see the different kinds of objects and places you can search for using Google Goggles.



[Source: N. Snavely]

Applications: Google Search by Image



The screenshot shows the Leafsnap website interface. At the top left is the 'leafsnap' logo. Below it is a navigation bar with 'Home', 'Species', 'Collectors', and 'About' tabs. The main content area features the title 'Leaf of the Bottlebrush Buckeye'. Below the title is a sub-header 'Leafsnap: An Electronic Field Guide' followed by two paragraphs of text describing the app's purpose and content. To the right of the text is a large, detailed image of a green, palmately compound leaf with five leaflets. At the bottom left, there are links for 'Free for iPhone' and 'iPad' with 'Available on the App Store' logos, a 'YouTube' icon, and the 'guardian.co.uk' logo.

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

[Source: N. Snaveley]

Applications: Finding Similar Products

Google Product Search

http://www.google.com/shopping/offerdetails?docid=182608052530392970&sa=X&ei=JPL9Toe1HCs... fnger print recognition

Most Visited - Gmail - Inbox (2318... Getting Started Latest Headlines Apple TTIC - Calendar Gmail - Inbox (3556... Yahoo! Google Maps YouTube Wikipedia News

Google Product Search +

+You Web Images Videos Maps News Shopping Gmail More - Sign in

Google product search [input] Advanced Product Search

TORY BURCH
Tory Burch Envelope Clutch



\$375.00
Free shipping
[shopbop.com](#)

Add to Shopping List

This crocodile-embossed leather Tory Burch clutch features a logo medallion at the magnetic front flap. Lined interior features zip pocket. * 7"H x 10"L x 1.5"D. * Leather: Cowhide. * Weight: 15 oz / 0.42 kg. * Imported.

Visually Similar Items



Find: [input] Next Previous Highlight all Match case

Applications: Special effects, shape capture



Figure: The Matrix movies, ESC Entertainment, XYZRGB, NRC

[Source: S. Seitz]

Applications: Special effects, motion capture



Figure: Pirates of the Carribeon, Industrial Light and Magic

[Source: S. Seitz]

Applications: Sports



Day 5: Swimming - Men's 4X200M Final



Figure: Sportvision explanation

[Source: S. Seitz]

How football works

- Know the **orientation of the field with respect to the camera** so that it can paint the first-down line with the correct perspective from that camera's point of view.
- Know, in that same perspective framework, exactly **where every yard line is**.

How football works

- Know the **orientation of the field with respect to the camera** so that it can paint the first-down line with the correct perspective from that camera's point of view.
- Know, in that same perspective framework, exactly **where every yard line is**.
- Given that the cameraperson can move the camera, the system has to be able to **sense the camera's movement** (tilt, pan, zoom, focus) and **understand the perspective change** that results from that movement.

How football works

- Know the **orientation of the field with respect to the camera** so that it can paint the first-down line with the correct perspective from that camera's point of view.
- Know, in that same perspective framework, exactly **where every yard line is**.
- Given that the cameraperson can move the camera, the system has to be able to **sense the camera's movement** (tilt, pan, zoom, focus) and **understand the perspective change** that results from that movement.
- Given that the camera can pan while viewing the field, the system has to be able to **recalculate the perspective** at a rate of 30 frames per second as the camera moves.

How football works

- Know the **orientation of the field with respect to the camera** so that it can paint the first-down line with the correct perspective from that camera's point of view.
- Know, in that same perspective framework, exactly **where every yard line is**.
- Given that the cameraperson can move the camera, the system has to be able to **sense the camera's movement** (tilt, pan, zoom, focus) and **understand the perspective change** that results from that movement.
- Given that the camera can pan while viewing the field, the system has to be able to **recalculate the perspective** at a rate of 30 frames per second as the camera moves.
- A football field is not flat – it crests very gently in the middle to help rainwater run off. So the line calculated by the system has to appropriately **follow the curve of the field**.

How football works

- Know the **orientation of the field with respect to the camera** so that it can paint the first-down line with the correct perspective from that camera's point of view.
- Know, in that same perspective framework, exactly **where every yard line is**.
- Given that the cameraperson can move the camera, the system has to be able to **sense the camera's movement** (tilt, pan, zoom, focus) and **understand the perspective change** that results from that movement.
- Given that the camera can pan while viewing the field, the system has to be able to **recalculate the perspective** at a rate of 30 frames per second as the camera moves.
- A football field is not flat – it crests very gently in the middle to help rainwater run off. So the line calculated by the system has to appropriately **follow the curve of the field**.
- A football game is filmed by **multiple cameras** at different places in the stadium, so the system has to do all of this work for several cameras.

How football works

- Know the **orientation of the field with respect to the camera** so that it can paint the first-down line with the correct perspective from that camera's point of view.
- Know, in that same perspective framework, exactly **where every yard line is**.
- Given that the cameraperson can move the camera, the system has to be able to **sense the camera's movement** (tilt, pan, zoom, focus) and **understand the perspective change** that results from that movement.
- Given that the camera can pan while viewing the field, the system has to be able to **recalculate the perspective** at a rate of 30 frames per second as the camera moves.
- A football field is not flat – it crests very gently in the middle to help rainwater run off. So the line calculated by the system has to appropriately **follow the curve of the field**.
- A football game is filmed by **multiple cameras** at different places in the stadium, so the system has to do all of this work for several cameras.
- The system has to be able to **sense when players, referees or the ball crosses the first-down line** so it does not paint the line right on top of them.

How football works

- Know the **orientation of the field with respect to the camera** so that it can paint the first-down line with the correct perspective from that camera's point of view.
- Know, in that same perspective framework, exactly **where every yard line is**.
- Given that the cameraperson can move the camera, the system has to be able to **sense the camera's movement** (tilt, pan, zoom, focus) and **understand the perspective change** that results from that movement.
- Given that the camera can pan while viewing the field, the system has to be able to **recalculate the perspective** at a rate of 30 frames per second as the camera moves.
- A football field is not flat – it crests very gently in the middle to help rainwater run off. So the line calculated by the system has to appropriately **follow the curve of the field**.
- A football game is filmed by **multiple cameras** at different places in the stadium, so the system has to do all of this work for several cameras.
- The system has to be able to **sense when players, referees or the ball crosses the first-down line** so it does not paint the line right on top of them.
- The system also has to **be aware of superimposed graphics** that the network might overlay on the scene.

How football works

- Know the **orientation of the field with respect to the camera** so that it can paint the first-down line with the correct perspective from that camera's point of view.
- Know, in that same perspective framework, exactly **where every yard line is**.
- Given that the cameraperson can move the camera, the system has to be able to **sense the camera's movement** (tilt, pan, zoom, focus) and **understand the perspective change** that results from that movement.
- Given that the camera can pan while viewing the field, the system has to be able to **recalculate the perspective** at a rate of 30 frames per second as the camera moves.
- A football field is not flat – it crests very gently in the middle to help rainwater run off. So the line calculated by the system has to appropriately **follow the curve of the field**.
- A football game is filmed by **multiple cameras** at different places in the stadium, so the system has to do all of this work for several cameras.
- The system has to be able to **sense when players, referees or the ball crosses the first-down line** so it does not paint the line right on top of them.
- The system also has to **be aware of superimposed graphics** that the network might overlay on the scene.

Applications: 3D Pose Estimation with Depth Sensors



[Source: Microsoft Kinect]

Applications: 3D Reconstruction from Photo Collections

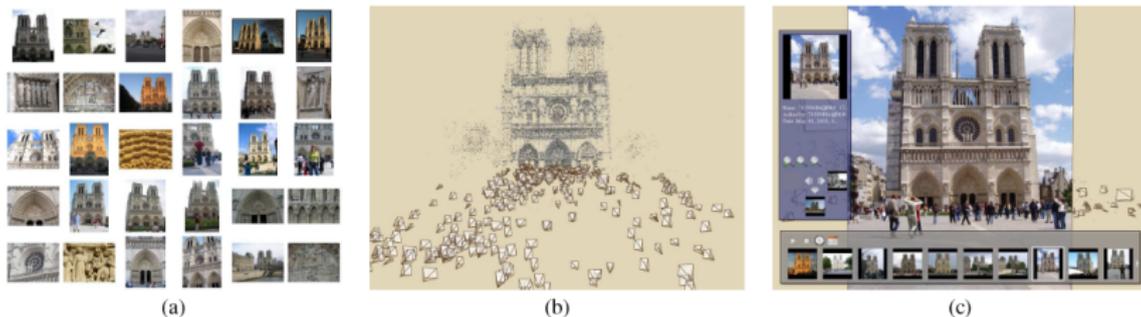


Figure 1: Our system takes unstructured collections of photographs such as those from online image searches (a) and reconstructs 3D points and viewpoints (b) to enable novel ways of browsing the photos (c).

[N. Snavely et al. Siggraph 2006]

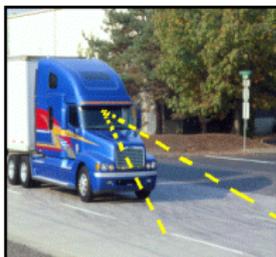
Applications: 3D Reconstruction from Depth Cameras



[Izadi et al. Siggraph 2011]

Applications: Assisted Driving

- Pedestrian and car detection
- Lane detection and lane departure warning
- Collision warning systems with adaptive cruise control



[Source: R. Fergus]

Applications: Smart cars

- **Mobileye:** Vision systems currently in high-end BMW, GM, Volvo models

The image is a screenshot of the Mobileye website. At the top, there are navigation tabs for "manufacturer products" and "consumer products". The main header reads "Our Vision. Your Safety." Below this is a top-down view of a car with three camera fields of view highlighted in yellow: "rear looking camera" at the back, "forward looking camera" at the front, and "side looking camera" on the side. Below the car view are three product highlights: "EyeQ Vision on a Chip" with an image of a chip, "Vision Applications" with an image of a pedestrian and text "Road, Vehicle, Pedestrian Protection and more", and "AWS Advance Warning System" with an image of a car on a road display. On the right side, there is a "News" section with two articles: "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", followed by an "Events" section with two items: "Mobileye at Equip Auto, Paris, France" and "Mobileye at SEMA, Las Vegas, NV".

manufacturer products consumer products

Our Vision. Your Safety.

rear looking camera forward looking camera

side looking camera

• **EyeQ** Vision on a Chip

• **Vision Applications**
Road, Vehicle, Pedestrian Protection and more

• **AWS** Advance Warning System

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

[Source: A. Shashua and S. Seitz]

Applications: Smart cars

- Part of my own research focus on smart cars



Applications: Vision in Space

Vision systems (JPL) uses for several tasks

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



The Heights of Mount Sharp

http://www.nasa.gov/mission_pages/msl/multimedia/pia16077.html

Panorama captured by Curiosity Rover, August 18 (Sol 12)

[Source: N. Snavely]

Applications: Robotics



NASA's Mars Curiosity Rover (Mars Science Laboratory)

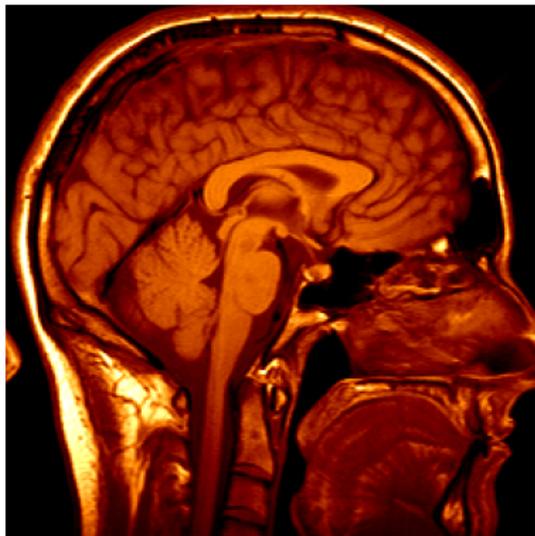
http://en.wikipedia.org/wiki/Spirit_rover



Autonomous RC Car

<http://www.cs.cornell.edu/~asaxena/rccar/>

Applications: Medical Imaging



3D imaging
MRI, CT



Image guided surgery
[Grimson et al., MIT](#)

Source: S. Seitz

Current state of the art

- You just saw examples of current systems, many less than 5 years old
- This is a very active research area, and rapidly changing
- To learn more about vision applications and companies David Lowe maintains an excellent overview of vision companies

<http://www.cs.ubc.ca/spider/lowe/vision.html>

[Source: N. Snavely]

Let's talk about the inevitable ...

- One final exam, NO midterm
- Exercises every week
 - Theoretical
 - Summary and critic of papers
 - Programming
- Project?

- Every day late, the max grade is divided by 2.
- Strict submission time!
- All submissions in latex. A template will be provided.

In this class ...

We will "hopefully" cover the following topics:

- Image processing
- Reconstruction
- Grouping
- Tracking
- Recognition: only an intro. See full class next year if you are interested.

Questions?