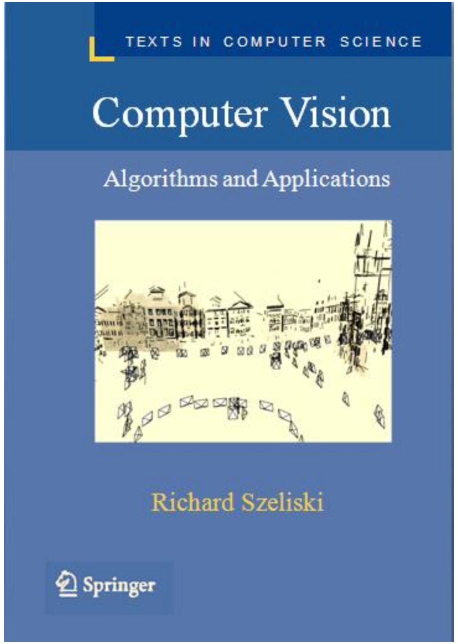
Microsoft® Research
FacultySummit

Vision-based Natural User Interfaces

Richard Szeliski
Principal Researcher
Microsoft Research

FUTURE / WORLD
2011 2031

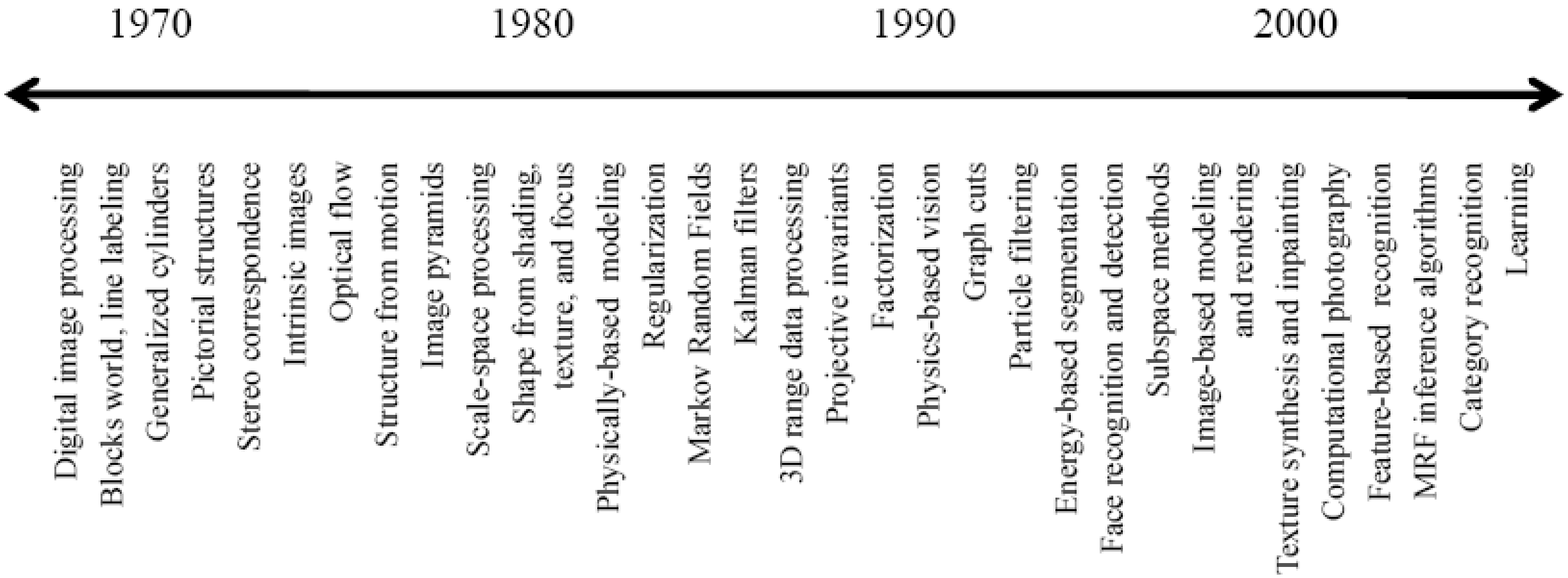
A Brief History of Computer Vision



A Brief History of Computer Vision

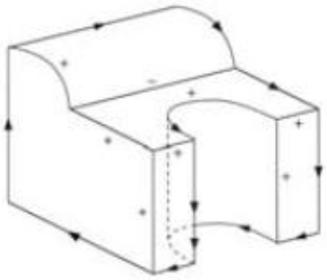
1970s. When computer vision first started out in the early 1970s, it was viewed as the visual perception component of an ambitious agenda to mimic human intelligence and to endow robots with intelligent behavior. At the time, it was believed by some of the early pioneers of artificial intelligence and robotics (at places such as MIT, Stanford, and CMU) that solving the "visual input" problem would be an easy step along the path to solving more difficult problems such as higher-level reasoning and planning. According to one well-known story, in 1966, Marvin Minsky at MIT asked his undergraduate student Gerald Jay Sussman to "spend the summer linking a camera to a computer and getting the computer to describe what it saw" (Boden 2006, p. 781).[5] We now know that the problem is slightly more difficult than that.[6]

A Brief History of Computer Vision

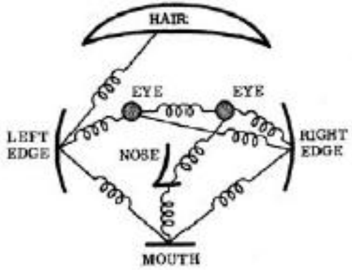


A Brief History of Computer Vision

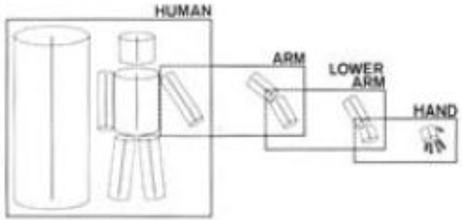
1970s



(a)

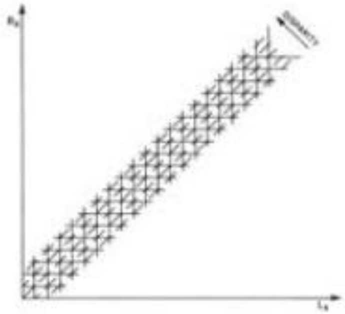


(b)

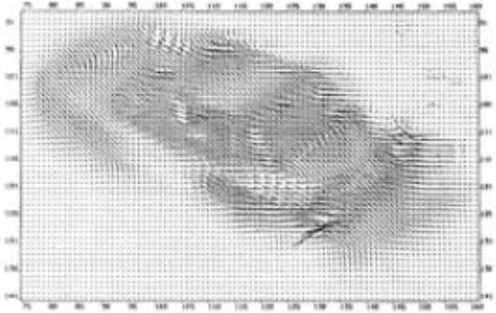


(c)

(d)



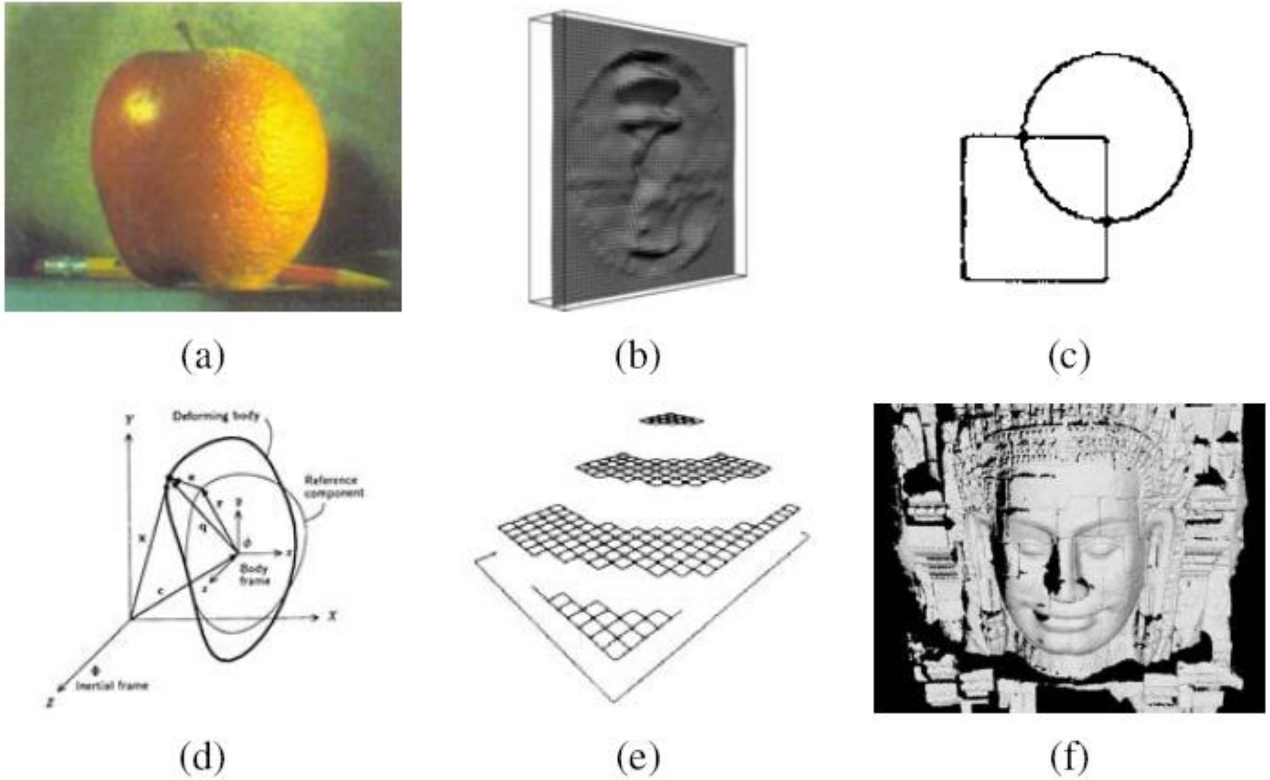
(e)



(f)

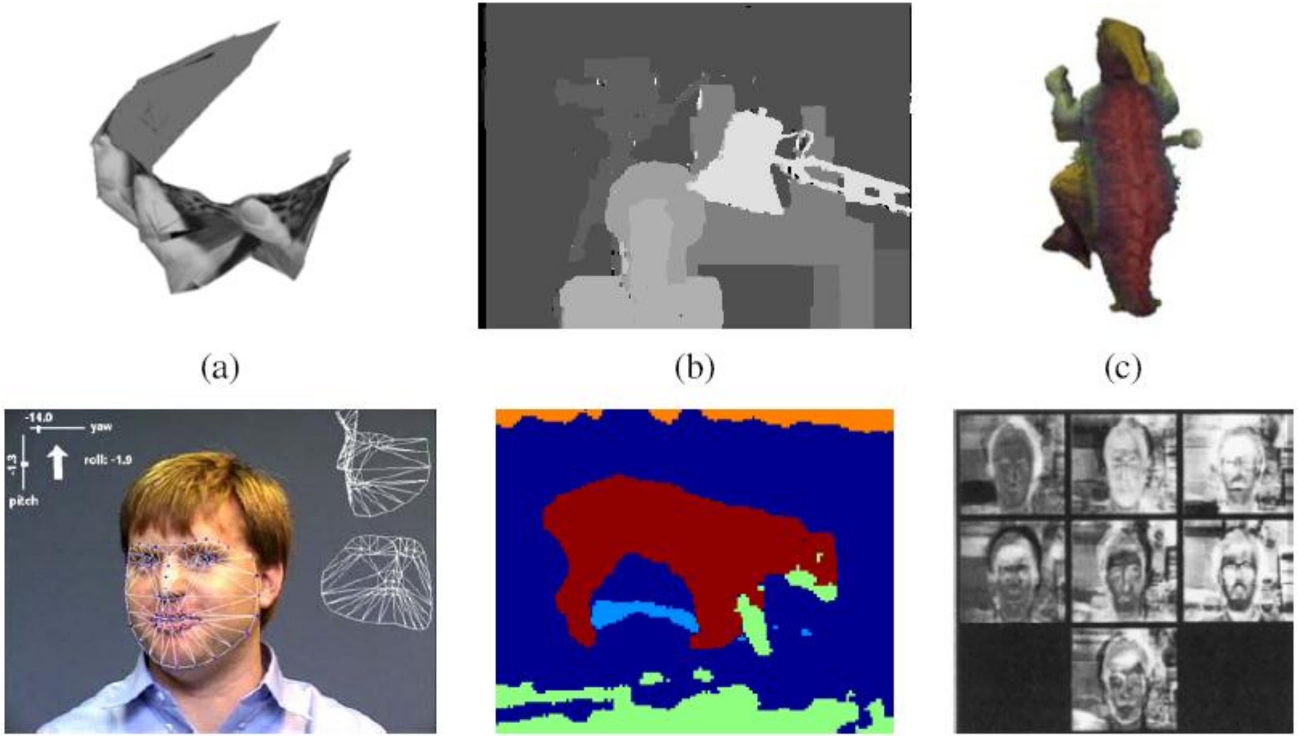
A Brief History of Computer Vision

1980s



A Brief History of Computer Vision

1990s

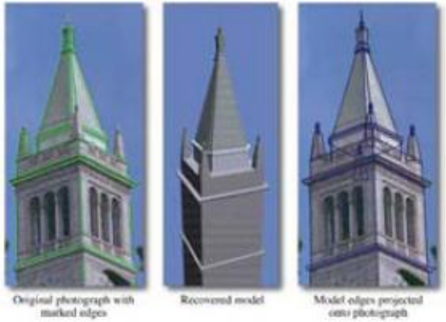


(a) (b) (c) (d) (e) (f)

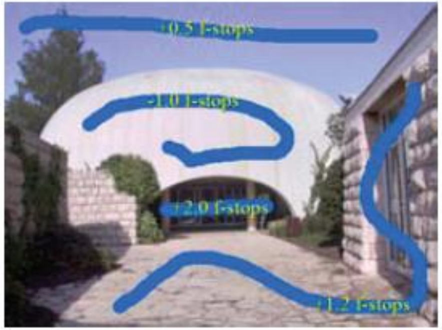
A Brief History of Computer Vision

2000s

(a)



(b)



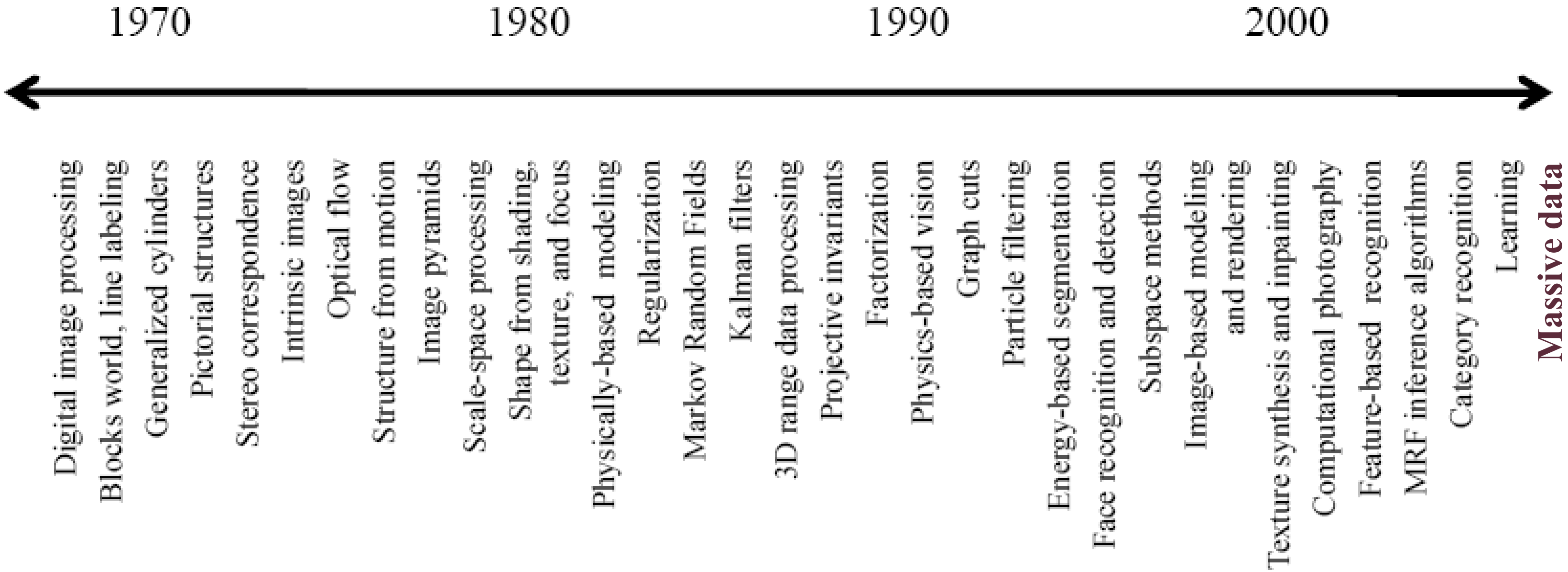
(c)

(d)

(e)

(f)

A Brief History of Computer Vision



Outline

Computer vision and machine learning techniques are maturing and having major impact:

1. 3D body tracking [Kinect]
2. Medical image segmentation [Amalga]
3. Object (product) recognition [Bing Vision]
4. Multi-image matching and navigation [Photosynth]

Massive (Internet) data is playing a key role

Microsoft® Research
FacultySummit

1. Body part recognition for Kinect

Jamie Shotton
Microsoft Research Cambridge
FG 2011 & CVPR 2011

FUTURE/WORLD 2011 2031

Microsoft Research FacultySummit

2. Medical Image Segmentation

Antonio Criminisi
Microsoft Research Cambridge

FUTURE WORLD 2011 2031

To find out more…

9:00–10:30	**Breakout Sessions**	
	Session: Medical Visualization Medical Imaging on the Microsoft Platform **Session Chair:** Rick Benge, Microsoft Research **Presentations:** ▪ Advanced Medical Imaging Research at Microsoft and its Applications on Product Groups—Khan Siddiqui, Microsoft ▪ Inner Eye: Toward a Computational Platform for Imaging Metadata—Steve White, Microsoft ▪ Applications of Advanced Semantic Tagging in Clinical Settings—David Haynor, University of Washington Analysis and metadata extraction and from medical image data represent significant computational challenges, but current open source efforts in the field of medical imaging focus on sharing code rather than sharing information. A common platform enabling researchers to benchmark and integrate very different analysis techniques in a common environment, and exchange both data and analyses on the web, would greatly accelerate research in this area. In this session, the speakers will present three different aspects of how Microsoft and its partners are addressing these challenges in terms of research, development, and real-world deployment.	Rainier

Microsoft Research
FacultySummit

3. (Mobile) Object Recognition

Larry Zitnick
Microsoft Research Redmond

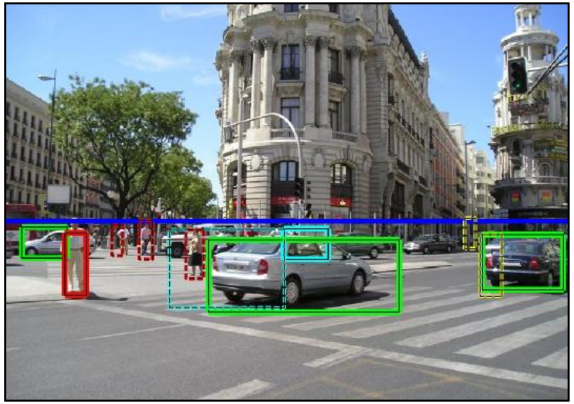
David Nister
Bing Vision

FUTURE/WORLD 2011 2031

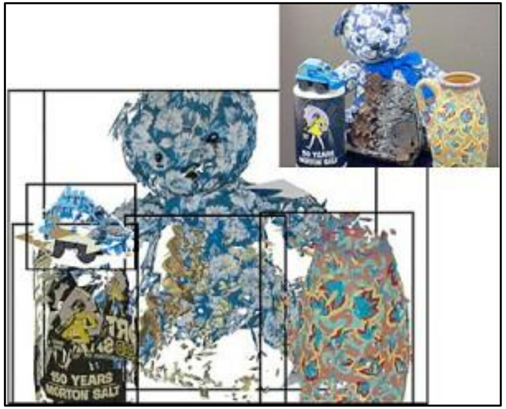
Object Recognition

- Detecting and localizing objects in images

Lowe IJCV 04

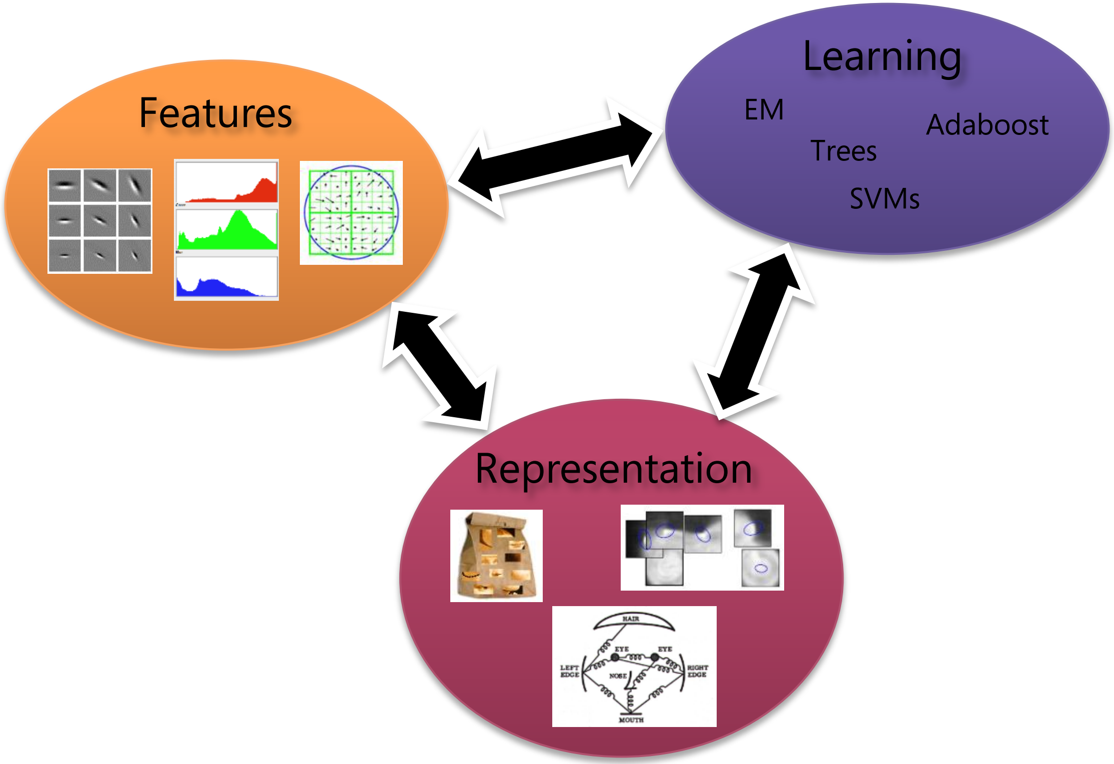


Hoiem et al. CVPR 06

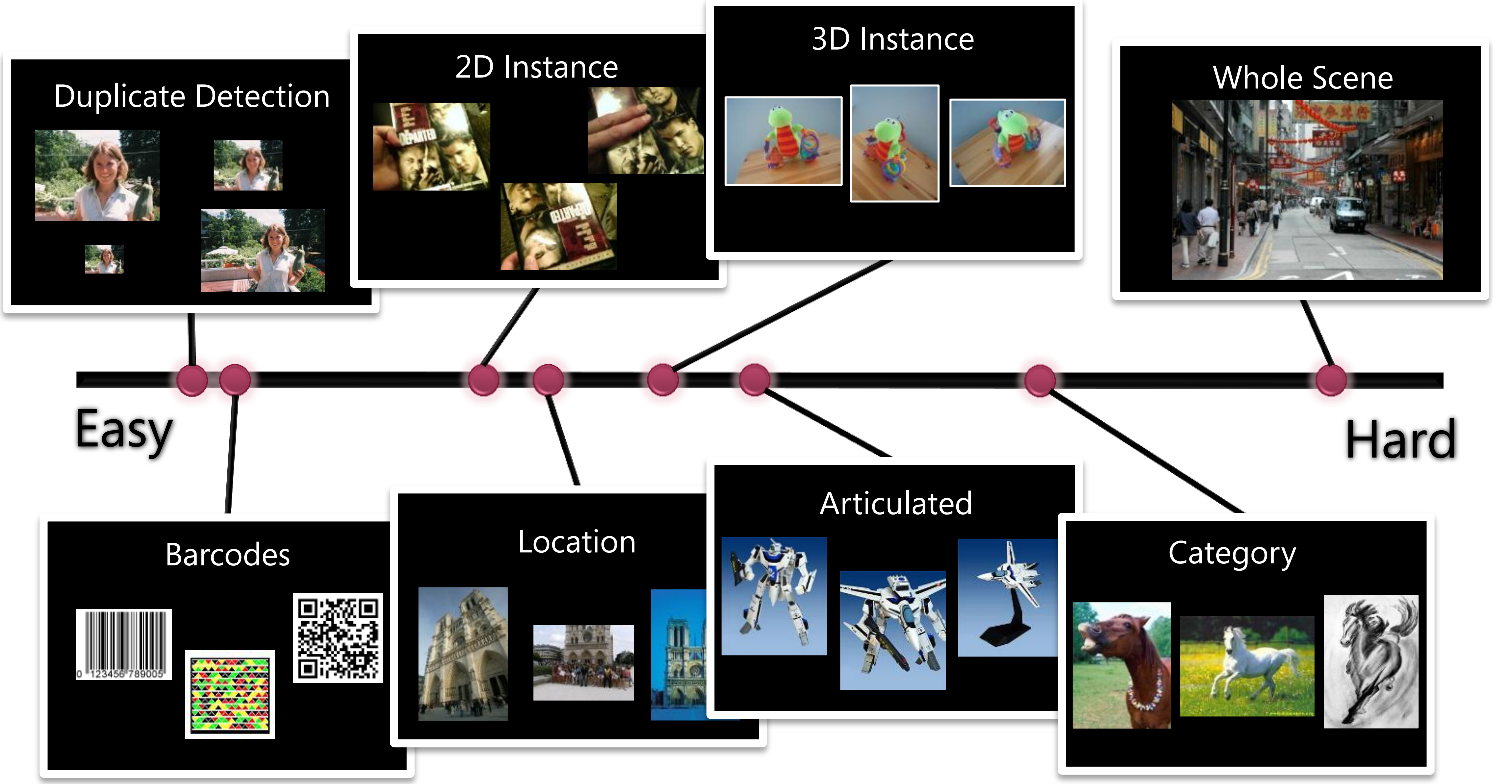


Rothganger et al. IJCV 06

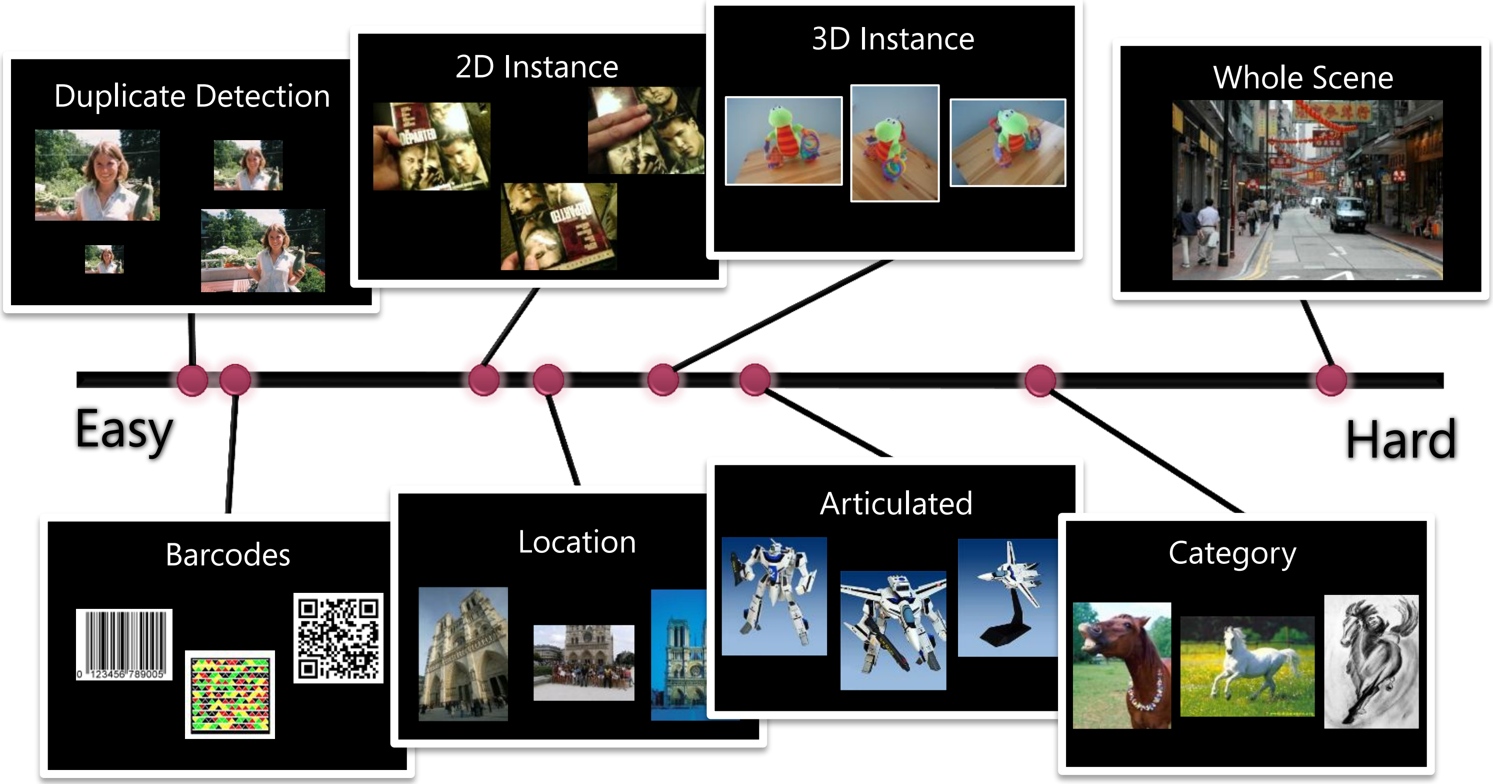
Problems


Features
Learning
EM
Adaboost
Trees
SVMs
Representation
HAIR
EYE
EYE
LEFT EDGE
RIGHT EDGE
NOSE
MOUTH

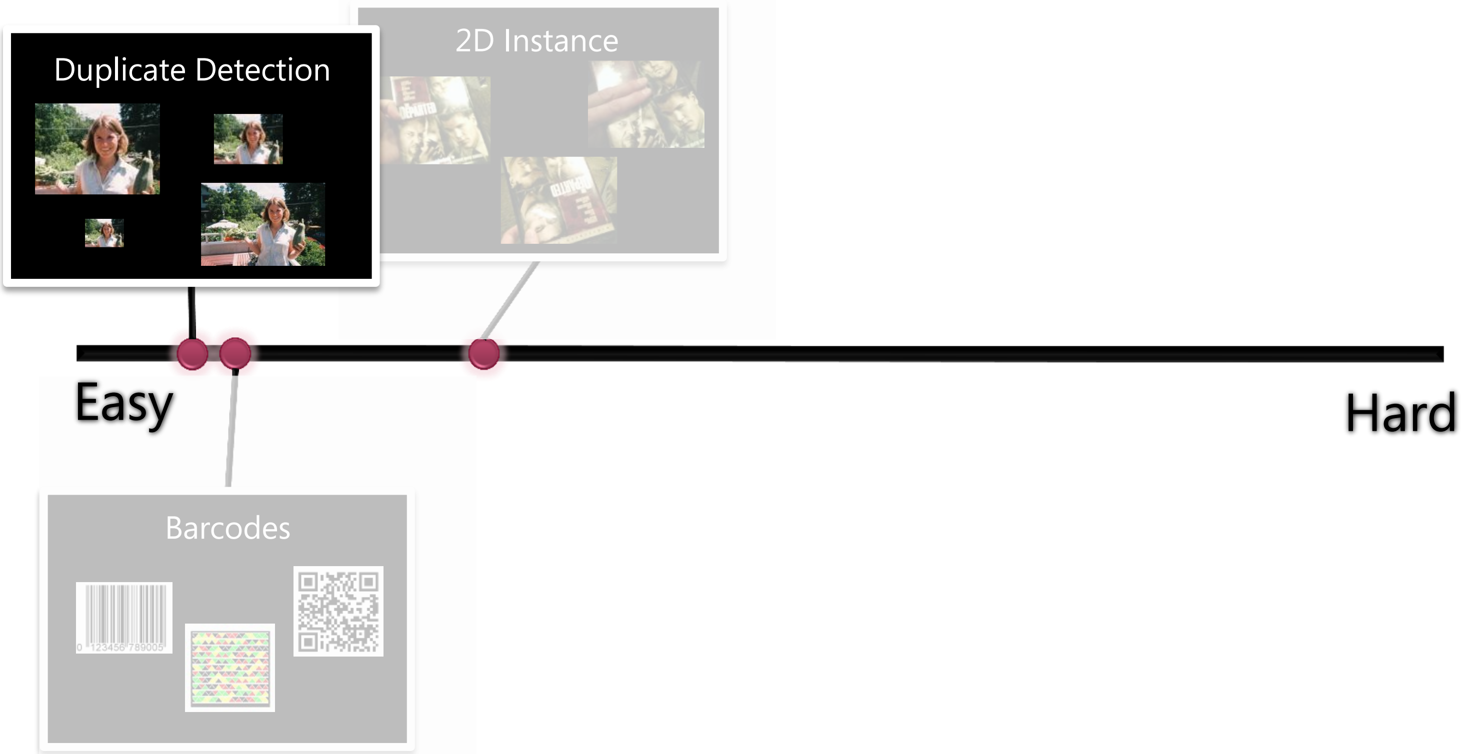
Spatial Complexity



Lincoln (Live Labs → Bing Mobile)

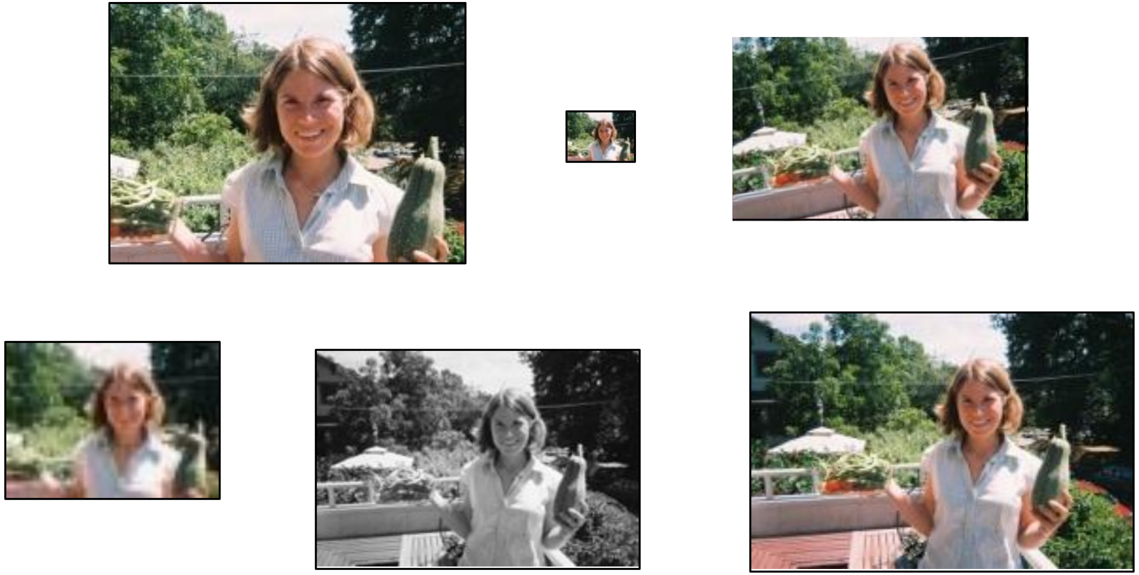


Near Duplicate Detection



Near Duplicate Detection

Are these images the same?



Near Duplicate Detection

Why is duplicate detection important?

- Increase search relevance
- Copyright search
- Remove illegal images






PhotoDNA

In 2009, Microsoft, working with Dartmouth College, developed PhotoDNA, a technology that aids in finding and removing some of the "worst of the worst" images of child sexual exploitations from the Internet. Microsoft donated the PhotoDNA technology to the National Center for Missing & Exploited Children (NCMEC), who established a PhotoDNA-based program for online service providers to help disrupt the spread of child pornography online. Over the next year, Microsoft, working with NCMEC, implemented a gradual rollout of PhotoDNA on Bing, SkyDrive and Hotmail services. In early 2011, Facebook joined Microsoft in sublicensing the technology for use on its network. It is our hope that other online service providers will follow Microsoft and Facebook's lead in adopting this game-changing technology.

Facebook Implements Microsoft's PhotoDNA Technology

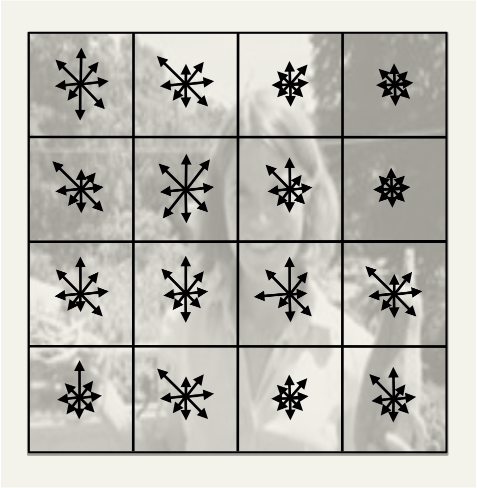
May 19, 2011

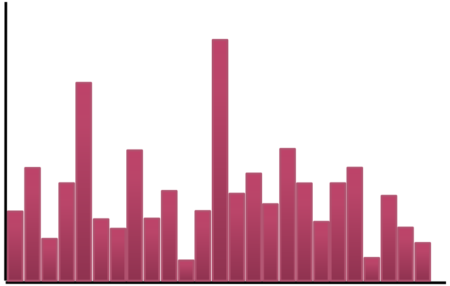
Facebook adopts PhotoDNA and joins Microsoft and The National Center for Missing & Exploited Children to disrupt the proliferation of online child exploitation.

- **Blog:** 500 Million Friends Against Child Exploitation
- **Blog:** Facebook To Use Microsoft's PhotoDNA Technology to Combat Child Exploitation
- **Video:** The Next Chapter in Protecting Children Online
- **Interactive:** Join the Facebook live event, May 20, 3 pm EDT

Approach

- Speed is most important

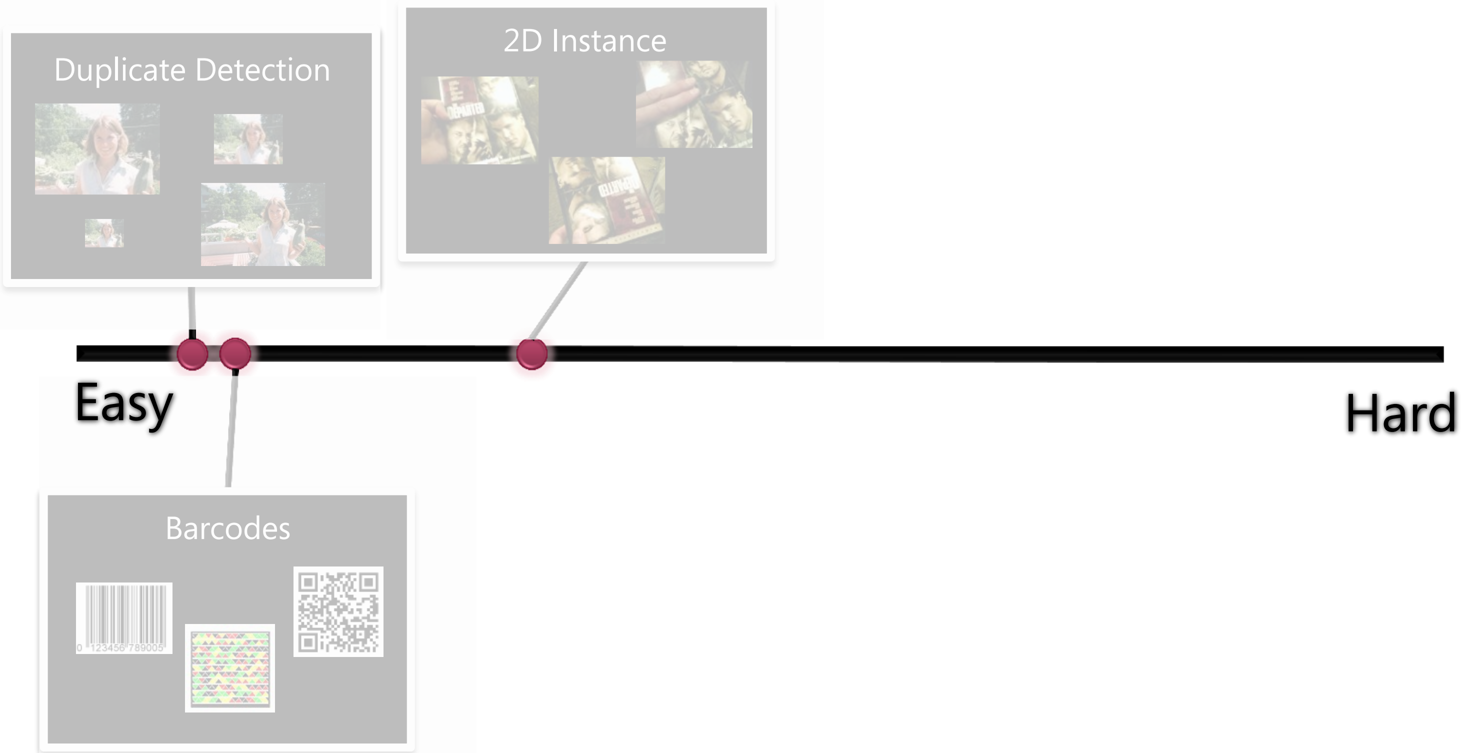




Fuzzy Hash

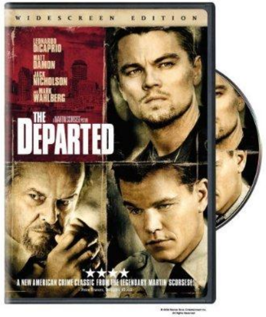
95% found
1 in 100 million false positives

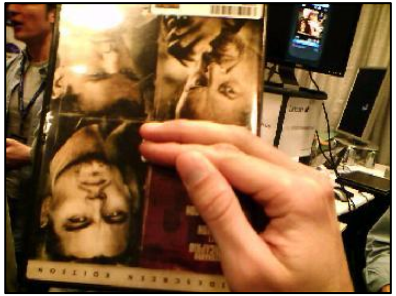
2D Object Instance Recognition

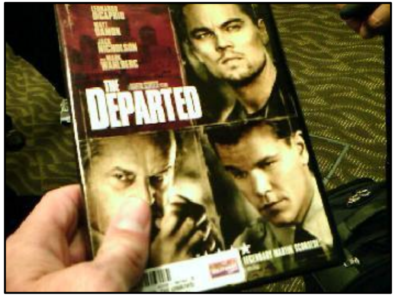


2D Object Instance Recognition

Is this the same planar object?







2D Object Instance Recognition

Why is this interesting?

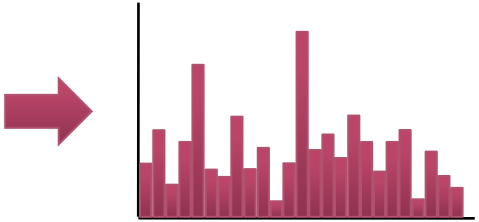
- Recognize real world objects
- Search using images
- Add metadata to images

Approach

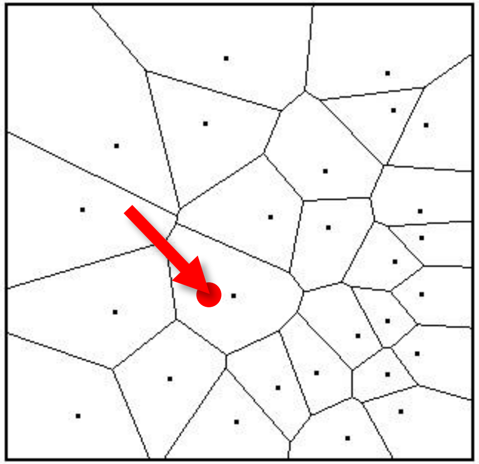
Find interest points

Extract patches

Approach



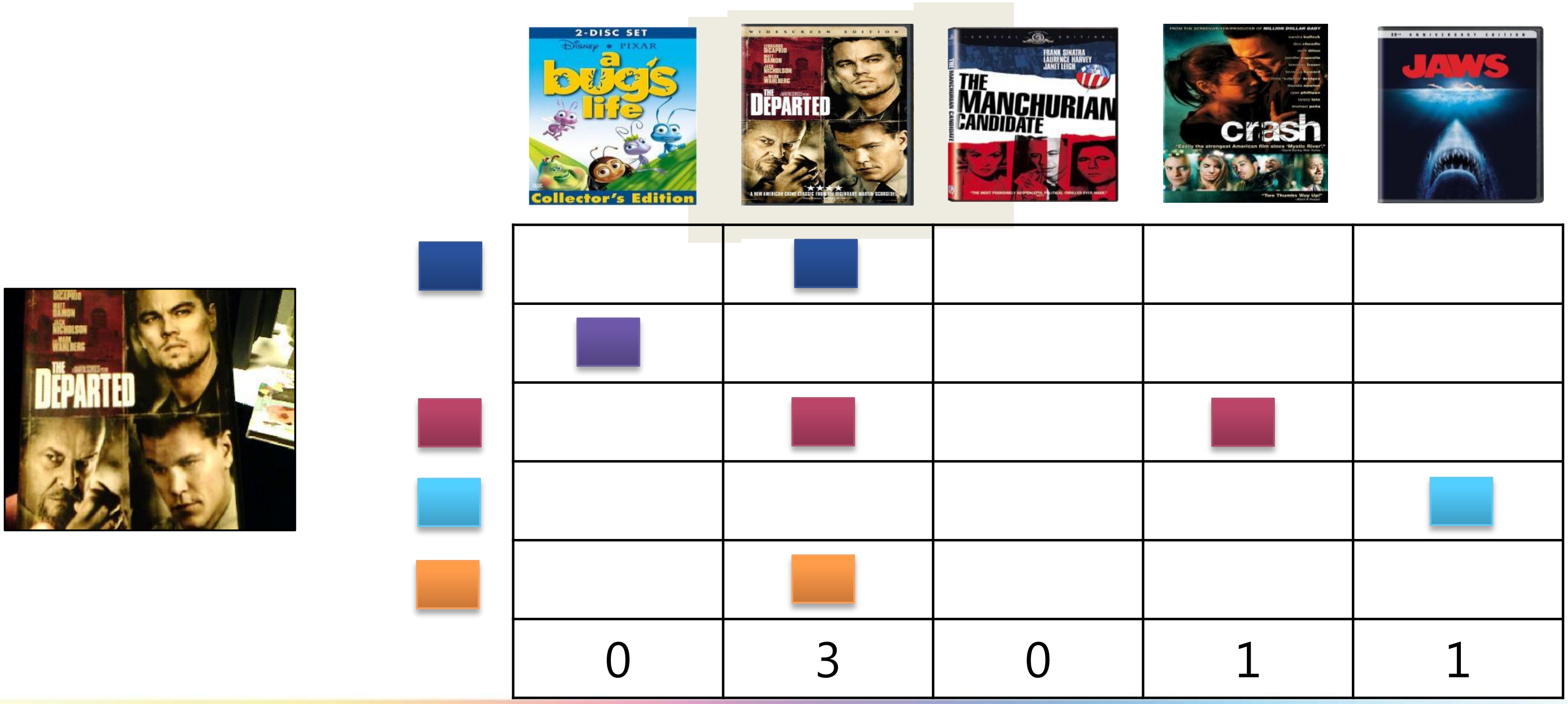
Compute descriptors



Quantize

kd-tree
vocabulary tree

Approach


2-DISC SET
a bug's life
Collector's Edition
THE DEPARTED
THE MANCHURIAN CANDIDATE
crash
JAWS
0
3
0
1
1

Windows Phone Mango and Bing Vision hands-on

By Tim Stevens posted May 24th 2011 12:43PM

HANDS-ON

Object Recognition Landscape

Microsoft Research FacultySummit

4. Internet Images

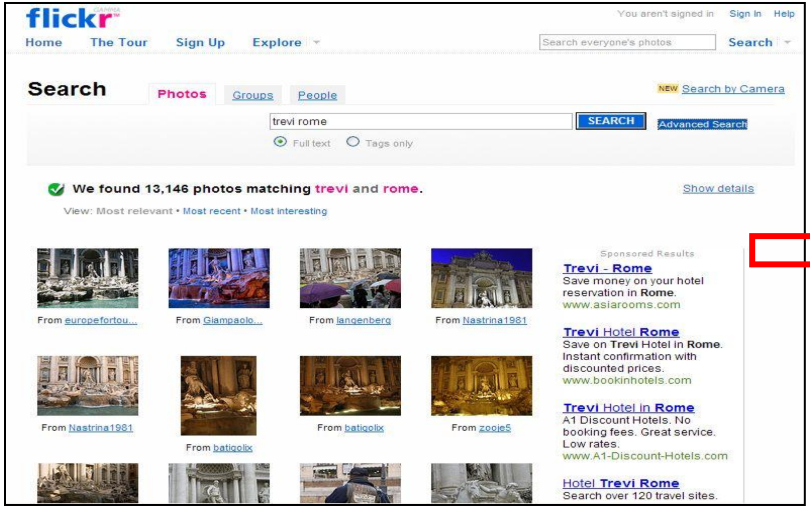
Noah Snavely
Cornell

+ …

FUTURE WORLD 2011 2031

Photo Tourism

[Snavely, Seitz, Szeliski, SIGGRAPH 2006]



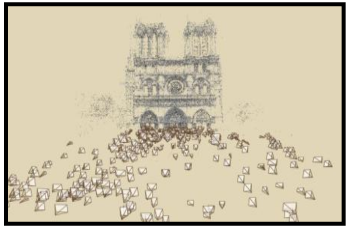
Images on the Internet

Computed 3D structure

Photo Tourism overview

Input photographs

Scene reconstruction



Relative camera positions and orientations

Point cloud

Sparse correspondence

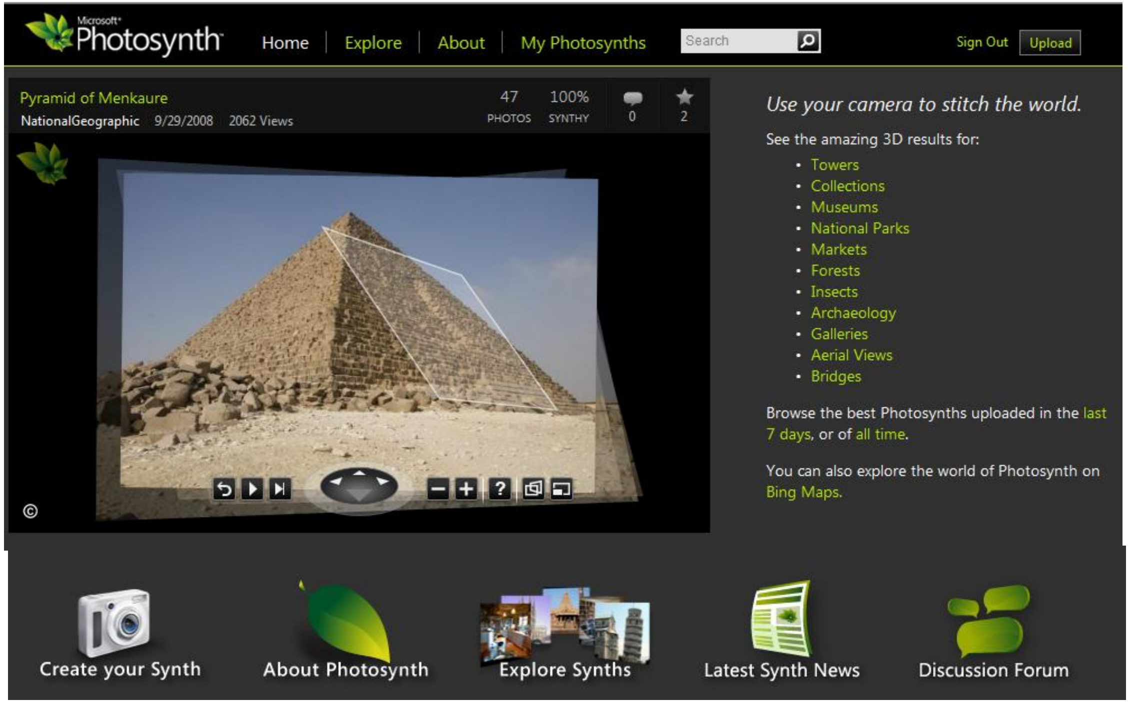
Photo Explorer

Incremental structure from motion

Navigation: Prague Old Town Square

Annotations: Notre Dame

Microsoft Photosynth



http://photosynth.net/

Microsoft Photosynth

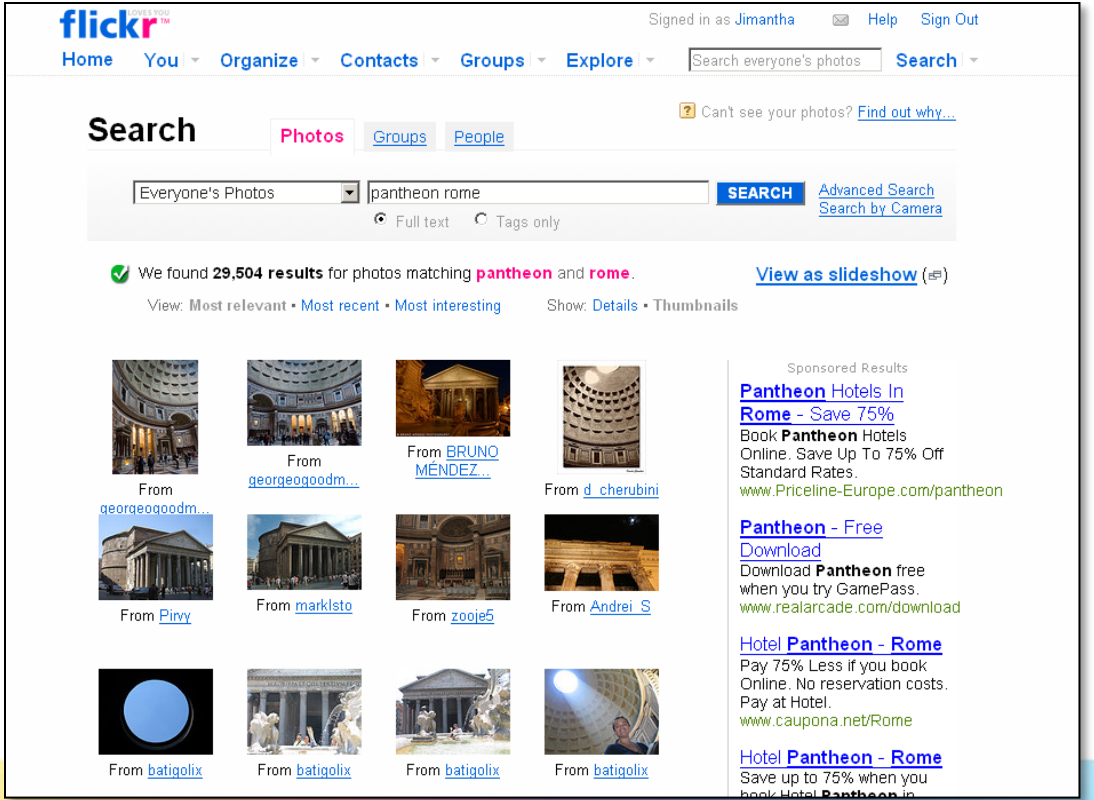
- 3D reconstruction
- Multi-resolution streaming & zooming
- Quad-based exploration
- Community photo sharing

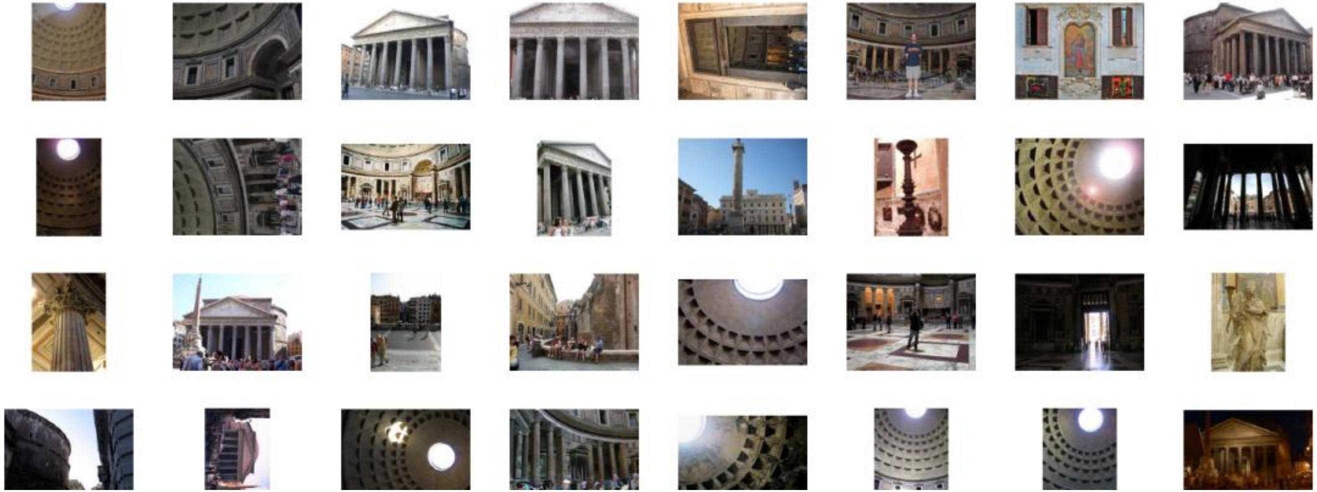
http://photosynth.net/

How well does this (Internet) scale?

Scene summarization for online photo collections

[Simon, Snavely, Seitz, ICCV 2007]





All photos on the Wikipedia page for the Pantheon (April, 2007)

Computed summary

Building Rome in a Day

Sameer Agarwal, Noah Snavely,
Ian Simon, Steven M. Seitz,
Richard Szeliski
ICCV'2009

Results: Dubrovnik

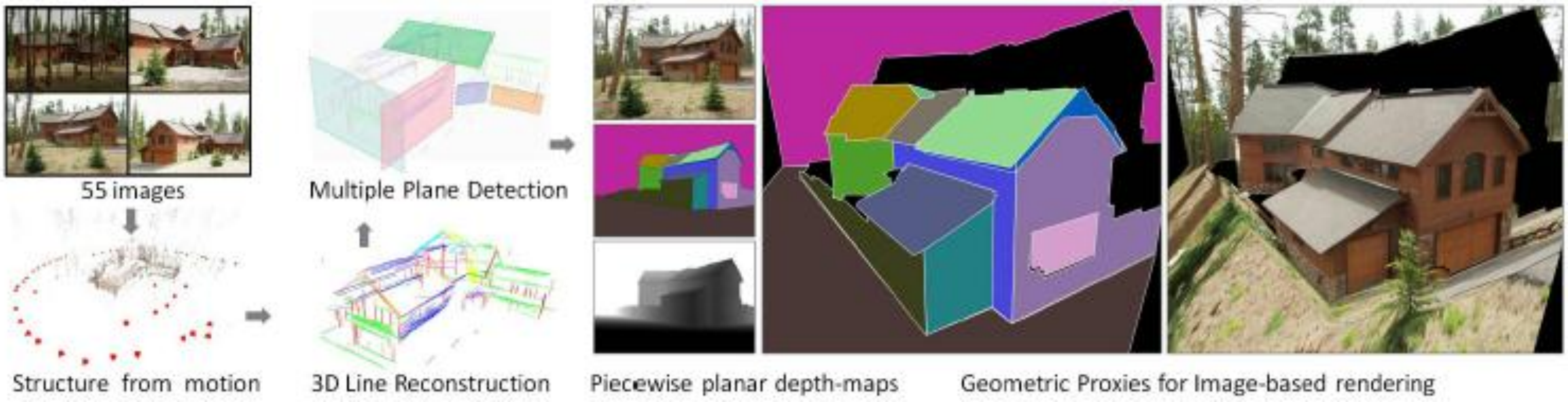
(a) Dubrovnik: Four different views and associated images from the largest connected component. Note that the component captures the entire old city, with both street-level and roof-top detail. The reconstruction consists of 4,585 images and 2,662,981 3D points with 11,839,682 observed features.

What about “real” 3D?

Piecewise Planar Stereo for Image-based Rendering

[Sinha, Steedly, and Szeliski. ICCV 2009]

Per-image piecewise-planar proxies



View Interpolation

Examples

Blogs

TechFest 2011: 3D Scanning with a regular camera or phone!

Posted: Mar 09, 2011 at 6:39 PM

By: Laura Foy

(3) | 44,866 Views | 14 Comments | Tweet 57 | Share 73

Avg Rating: 5

< > embed

Download

Right click "Save as..."

- High Quality WMV (PC, XBox, MCE)
- MP3 (Audio only)
- WMA (Audio only)
- Mid Quality WMV (Lo-band, Mobile)
- High Quality MP4 (iPad, WP7)
- MP4 (iPod, Zune HD)

Related posts


Expert to Expert: Helen Wang and Alex Moshchuk -...


Project JSMeter: JavaScript Performance Analysis...


E2E: Erik Meijer and Leslie Lamport - Mathematical...

3-D television is creating a huge buzz in the consumer space, but the generation of 3-D content remains a largely professional endeavor. Our research demonstrates an easy-to-use system for creating photorealistic, 3-D-image-based models simply by walking around

nit

What about "regular" Internet Images?

IMAGENET

12,184,113 images, 17624 synsets indexed

Explore New! Download New! **Challenge** People Publication About

Not logged in. Login | Signup

ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures.

Click here to learn more about ImageNet, Click here to join the ImageNet mailing list.

SEARCH

What do these images have in common? *Find out!*

Internet Computer Vision

COMP 790-096: Con

Fall 2007, Tuesdays 3:30-4:30, S

Instructor: Svetlana Lazebnik (l

Quick links: presentation sched

Second IEEE Workshop on Intern
CVPR 2009)

Program
- IV Home Page
- Workshop Home
- Areas of Interest
- Program

Workshop Home

General Chairs

Thomas S. Huang, *UIUC*
Harry Shum, *MSR*

Program Chairs

Shai Avidan, *Adobe Research*
Simon Baker, *MSR*
Ying Shan, *Microsoft AdCenter Labs*

Computer Vision and the Internet (09w5126)

Arriving Sunday, August 30 and departing Friday September 4, 2009

Proceedings OF THE IEEE

AUGUST 2010 / VOL. 98 / NO. 8

CONTENTS

SPECIAL ISSUE

INTERNET VISION

Edited by S. Avidan, S. Baker, and Y. Shan

DEPARTMENTS

What *else* can we do with these photos?

ShadowDraw: Real-Time User Guidance for Freehand Drawing

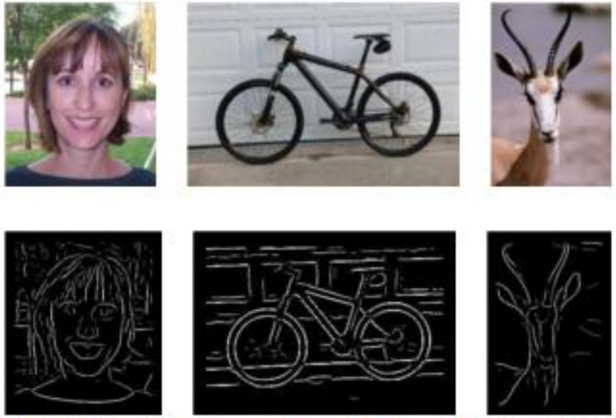
Yong Jae Lee, Larry Zitnick, and Michael Cohen

SIGGRAPH 2011

Computer Assisted Drawing

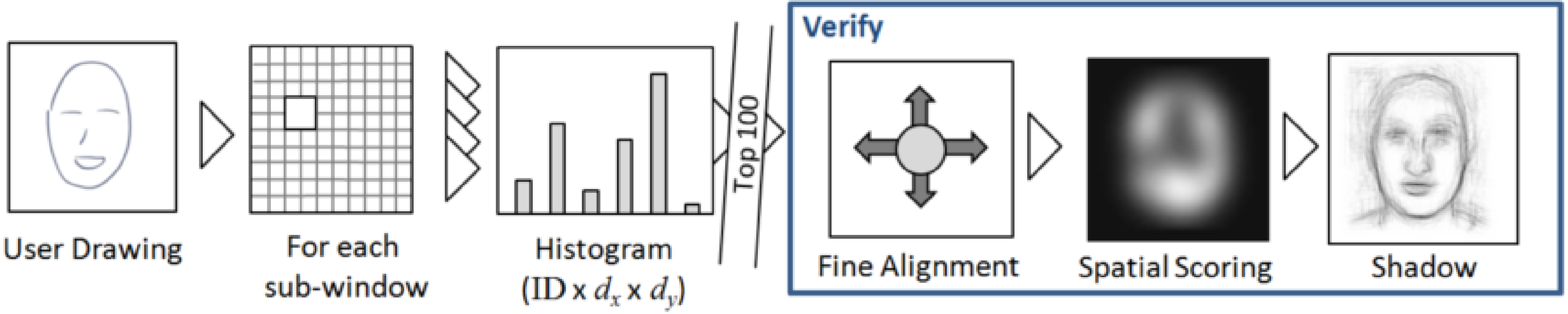
Main idea:

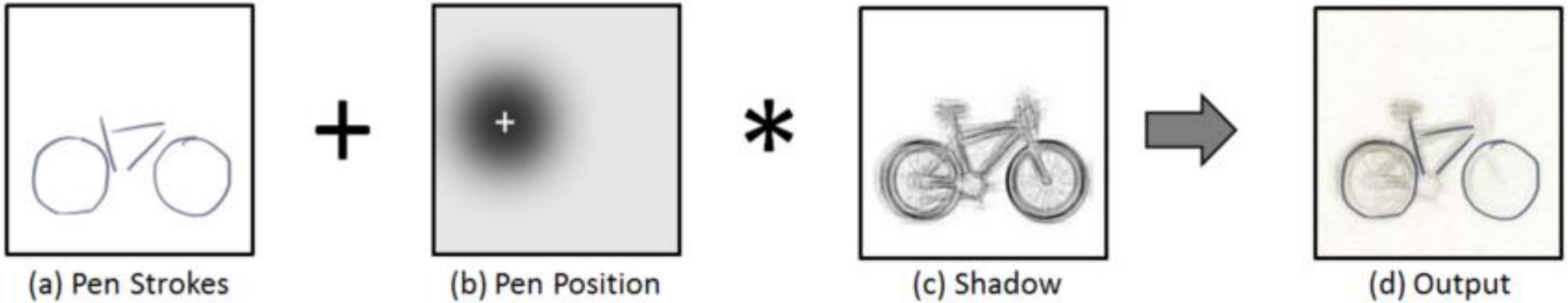
- Use a large database of images to suggest good contours for your drawing



Examples of database images and corresponding edge maps

Computer Assisted Drawing





[Video](#)

Wrapping up

- Exciting time to be working in computer vision
- Machine learning and massive (Internet) data are having a huge impact
  - Body tracking
  - Medical imaging
  - Object recognition
  - Internet-scale reconstruction
- General image recognition is still an open problem

Question?


XBOX 360


Final Cut Pro HD for
Final Cut Pro HD for
Final Cut Pro HD for

URLs for Web content

- http://www.engadget.com/2011/05/24/windows-phone-mango-and-bing-vision-hands-on/
- http://channel9.msdn.com/posts/TechFest-2011-3D-Scanning-with-a-regular-camera-or-phone
- http://www.microsoft.com/presspass/presskits/photodna/
- https://webspace.utexas.edu/yl3663/~ylee/shadowdraw/shadowdraw.html

Microsoft® Research
FacultySummit

2031
FUTURE/WORLD
2011