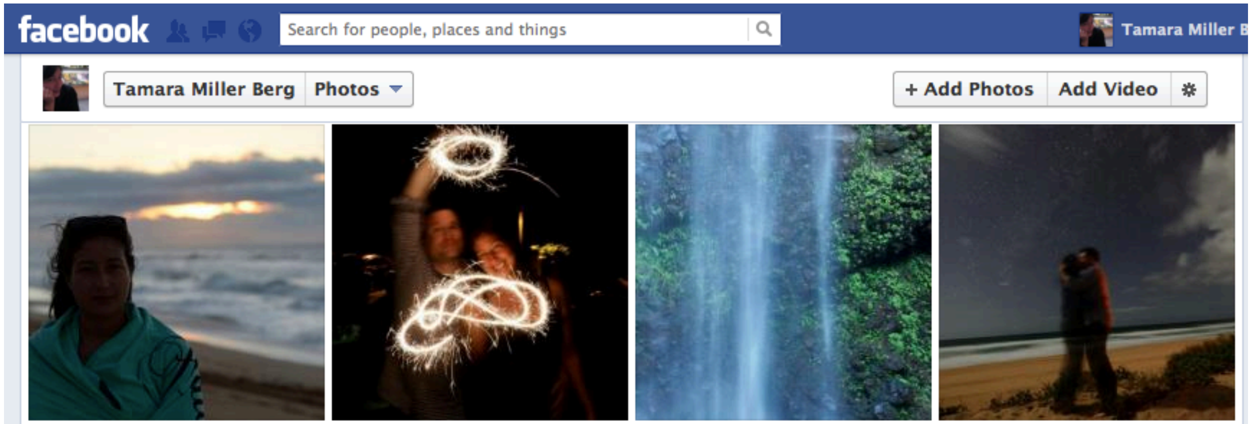
Computer Vision

Tamara Berg

Why computer vision?



100 billion – [Estimated](#) number of photos on Facebook, mid-2011.

40 million – Photos [uploaded](#) per day to Instagram.

6 billion – Photos [hosted](#) on Flickr (August 2011).

4.5 million – Number of [photos](#) uploaded to Flickr each day.

The goal of computer vision

- To perceive the story behind the picture

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

The goal of computer vision

- To perceive the story behind the picture

- What exactly does this mean?
  - Vision as a source of metric 3D information
  - Vision as a source of semantic information

Vision as measurement device

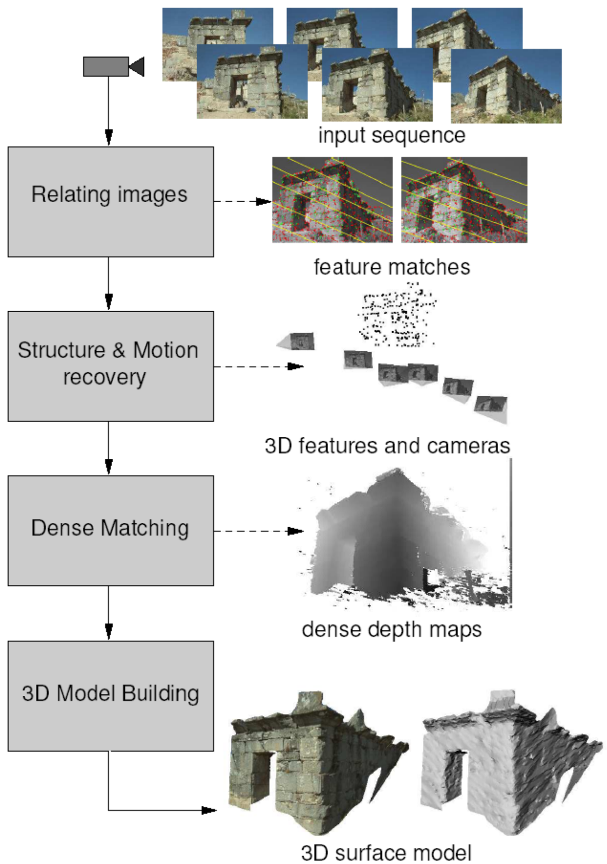


Photo Tourism

Vision as a source of semantic information

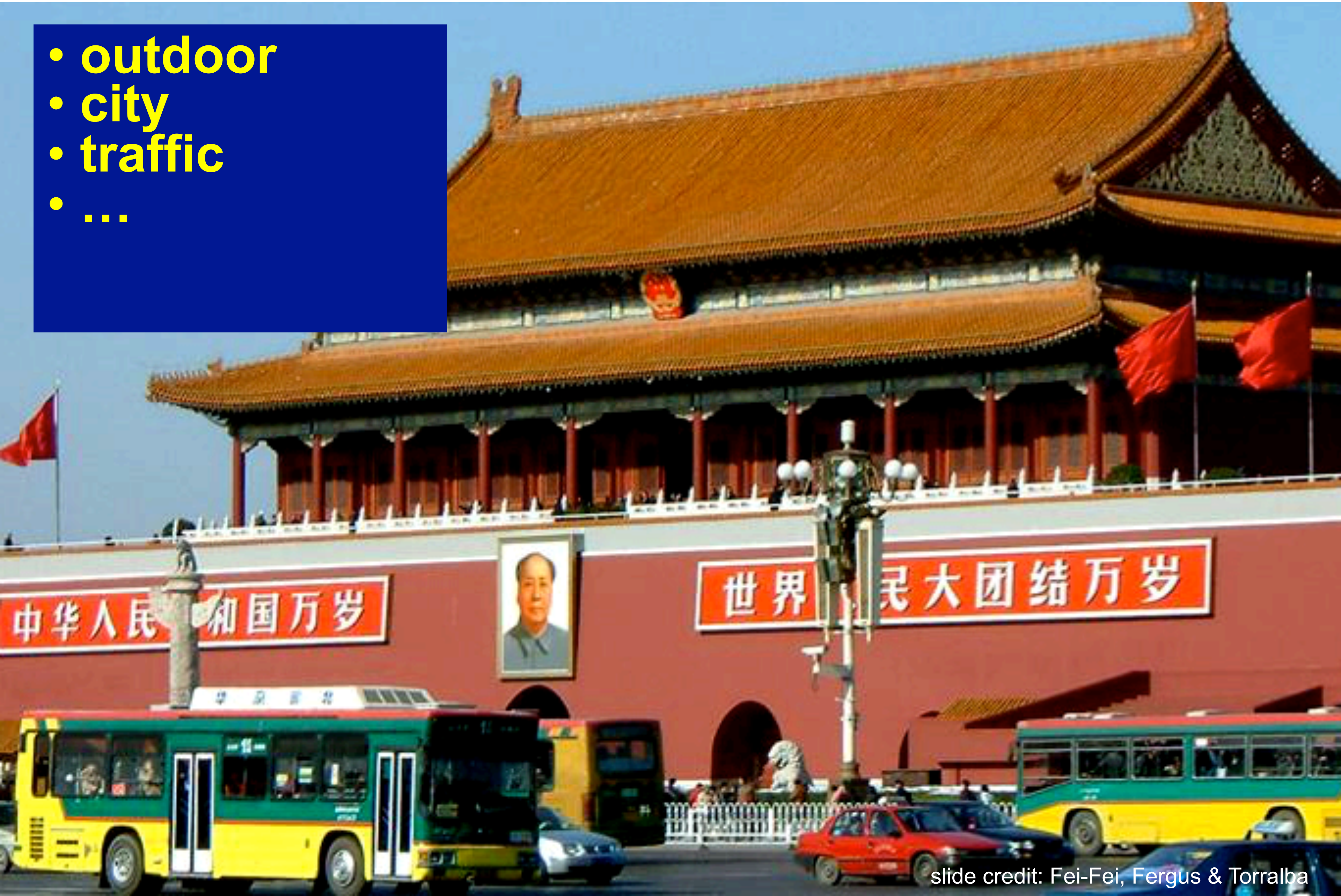
slide credit: Fei-Fei, Fergus & Torralba

Object categorization

slide credit: Fei-Fei, Fergus & Torralba

Scene and context categorization

- **outdoor**
- **city**
- **traffic**
- **…**



slide credit: Fei-Fei, Fergus & Torralba

Why study computer vision?

- Vision is useful
- Vision is interesting
- Vision is difficult
  - Half of primate cerebral cortex is devoted to visual processing
  - Achieving human-level visual perception is probably “AI-complete”

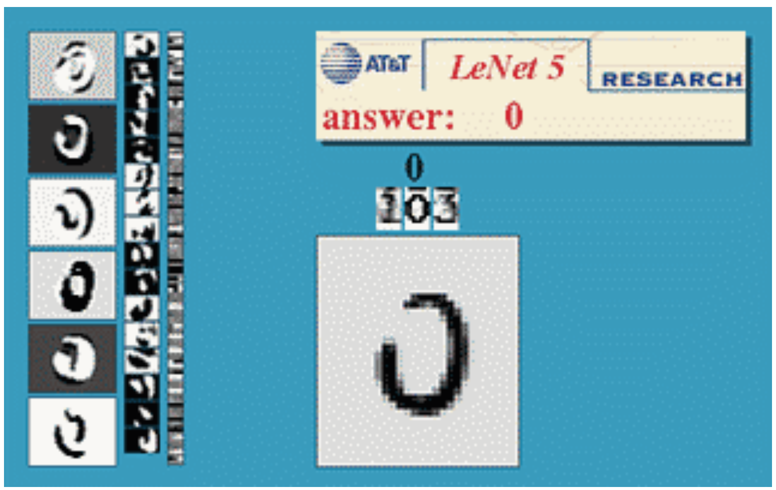
Progress to date

- The next slides show some examples of what current vision systems can do

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

Also used for zipcode reading by the USPS

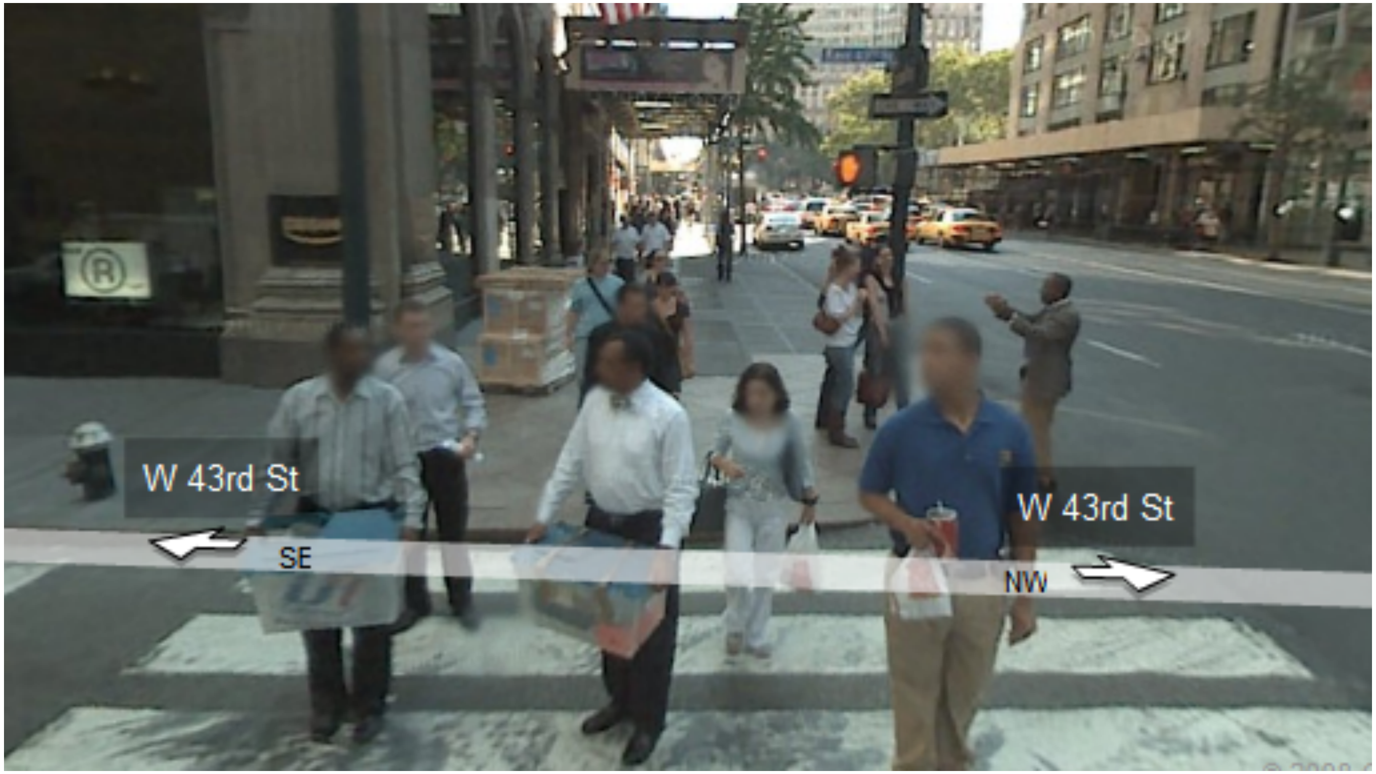
Source: S. Seitz

Face detection

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

Source: S. Seitz

Face Detection for Privacy



Face Blurring for Google Streetview

Face Detection for Privacy

Face Blurring for Google Streetview

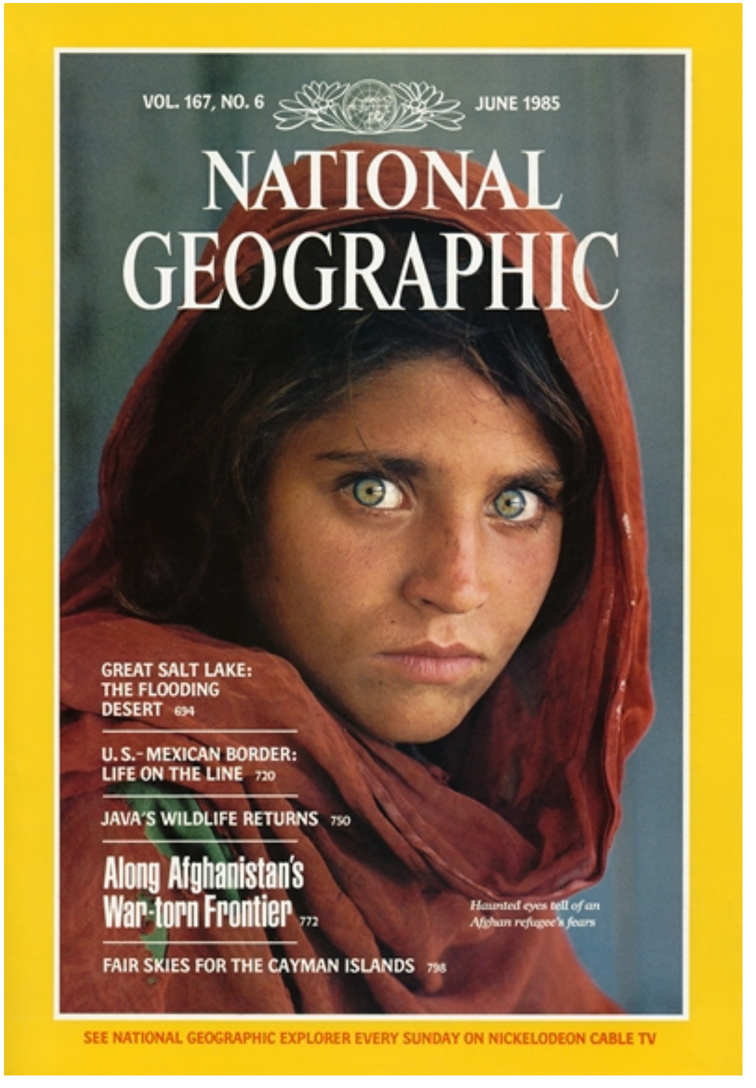
Object recognition (in supermarkets)

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

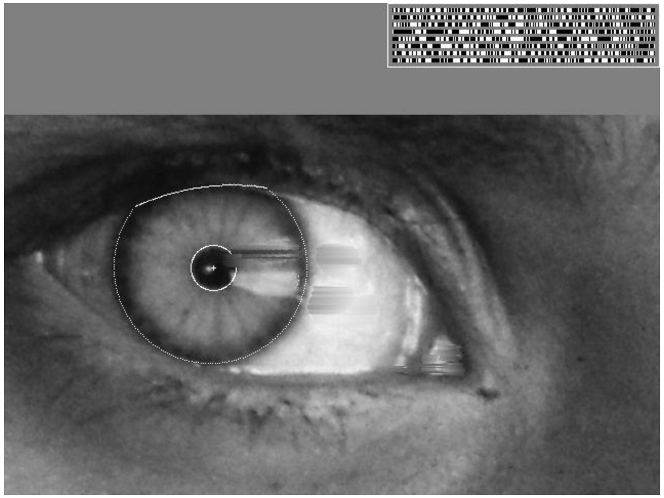
Who is she?

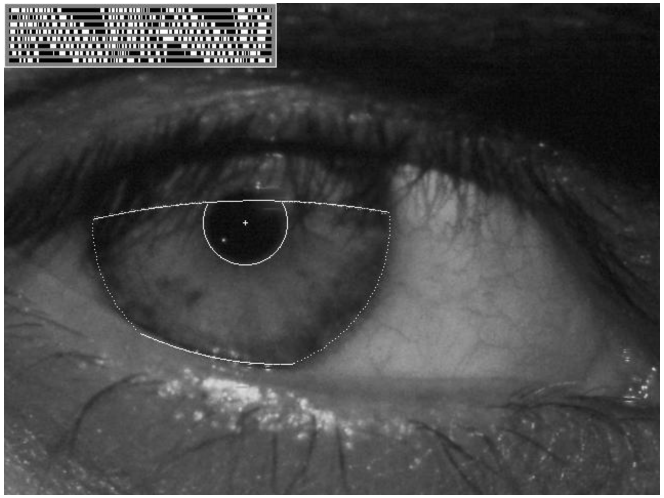


Source: S. Seitz

Vision-based biometrics

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

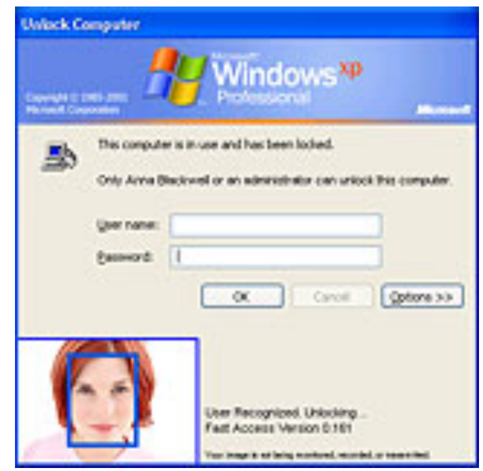




Source: S. Seitz

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

Object recognition (in mobile phones)

- This is becoming real:
  - Lincoln Microsoft Research
  - Point & Find
  - Google goggles

Source: S. Seitz

Sports

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

Smart cars

Slide content courtesy of Amnon Shashua

- [Mobileye]
  - Vision systems currently in high-end BMW, GM, Volvo models

Source: S. Seitz

Vision-based interaction (and games)

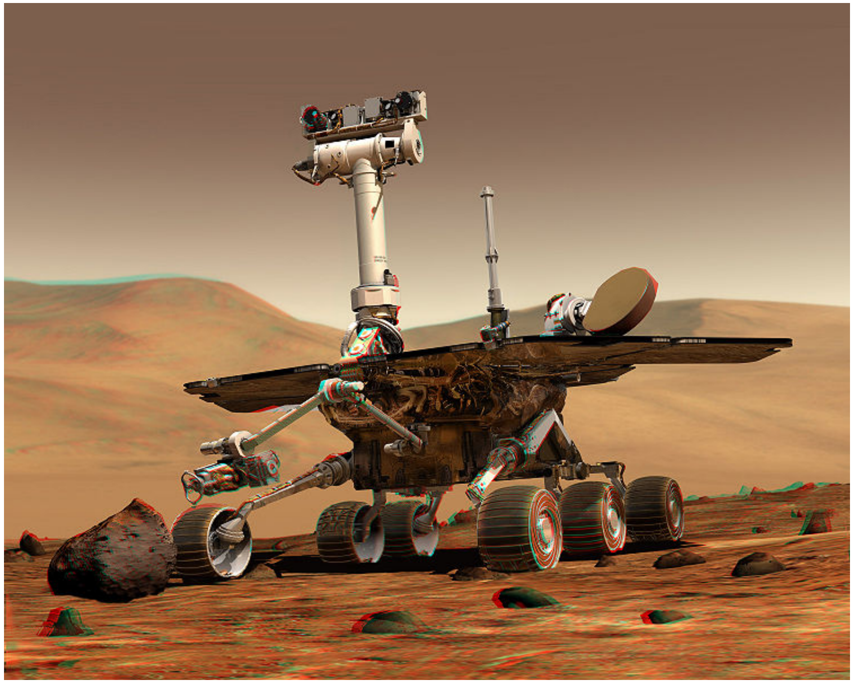
Nintendo Wii has camera-based IR tracking built in. See [Lee's work at CMU](#) on clever tricks on using it to create a [multi-touch display](#)!

Kinect – projector+camera for depth

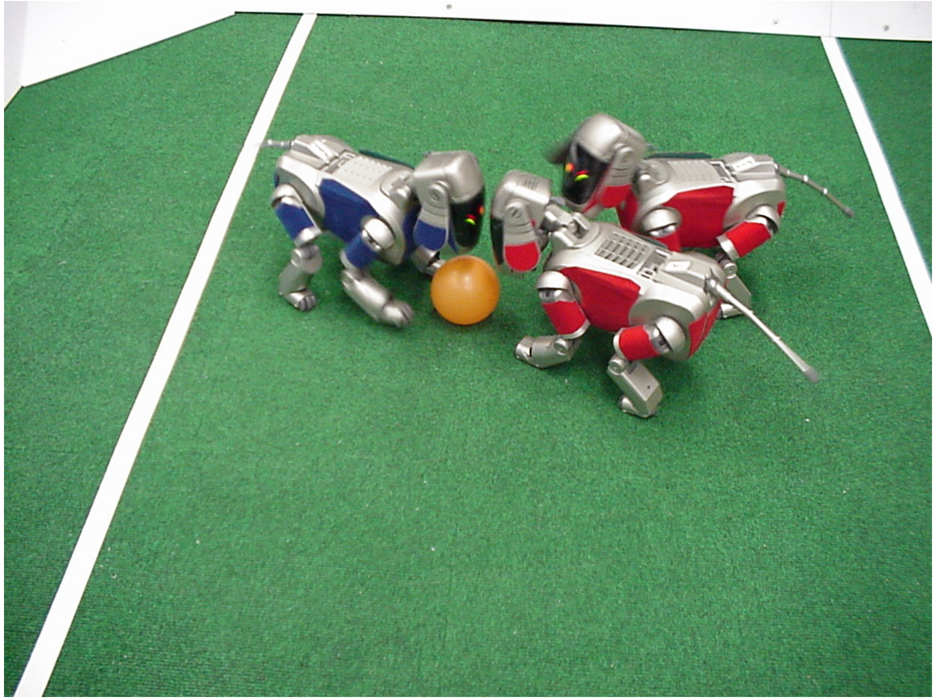
Assistive technologies

Source: S. Seitz

Robotics



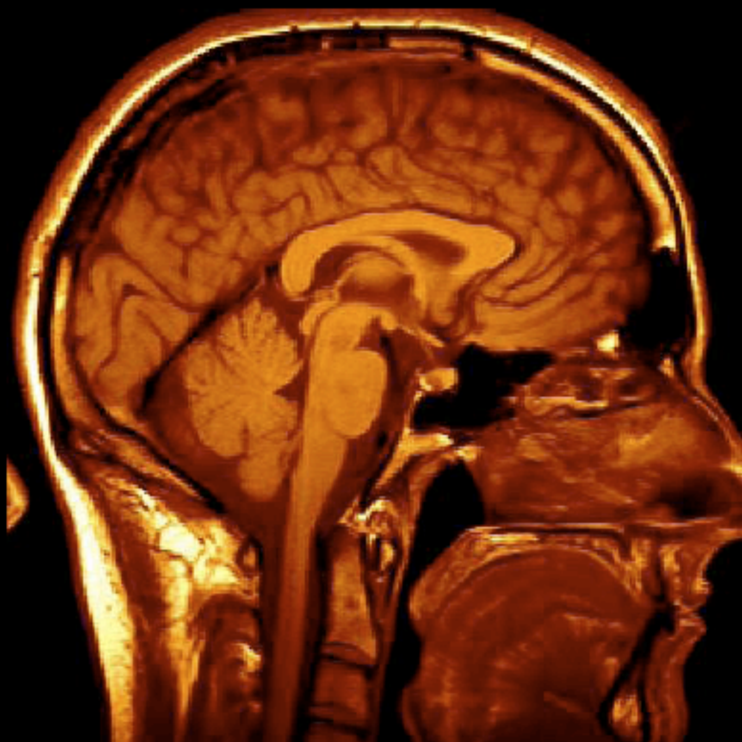
NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover



http://www.robocup.org/

Source: S. Seitz

Medical imaging



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

Source: C. Fowlkes

What I work on…

Two Example Projects

Moving recognition outputs toward human-like predictions

Extracting socio-identity information from pictures

Descriptive Text

“It was an arresting face, pointed of chin, square of jaw. Her eyes were pale green without a touch of hazel, starred with bristly black lashes and slightly tilted at the ends. Above them, her thick black brows slanted upward, cutting a startling oblique line in her magnolia-white skin–that skin so prized by Southern women and so carefully guarded with bonnets, veils and mittens against hot Georgia suns”

Scarlett O’Hara described in Gone with the Wind.

More Nuance than Traditional Recognition…

person

shoe

car

Human centric recognition outputs

car

Human centric recognition outputs

pink car

Human centric recognition outputs

car on road

Human centric recognition outputs

Little pink smart car parked on the side of a road in a London shopping district.

Telling the *"story of an image"*

generating image descriptions, a first attempt...

Baby Talk: Understanding and Generating Simple Image Descriptions

Girish Kulkarni, Visruth Premraj, Sagnik Dhar, Siming Li, Yejin Choi, Alexander C Berg, Tamara L Berg (CVPR 2011)

CooKies
Lemonade 50¢
Free CooKie's

“This picture shows one person,

“This picture shows one person, one grass,

“This picture shows one person, one grass, one chair,

“This picture shows one person, one grass, one chair, and one potted plant.

“This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass,

“This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair.

“This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair,

"This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair, and near the potted plant."

“This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair, and near the potted plant.”

Methodology

- Vision – find and identify objects
- Text -- statistics from parsing lots of descriptive text
- Model predicts best image labeling given vision and text based estimates
- Generation algorithms to compose natural language

Some good results

This is a picture of one sky, one road and one sheep. The gray sky is over the gray road. The gray sheep is by the gray road.

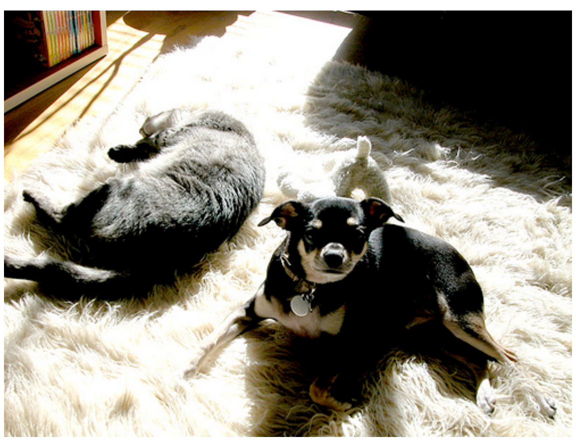
Here we see one road, one sky and one bicycle. The road is near the blue sky, and near the colorful bicycle. The colorful bicycle is within the blue sky.

This is a picture of two dogs. The first dog is near the second furry dog.

Some bad results

Missed detections:

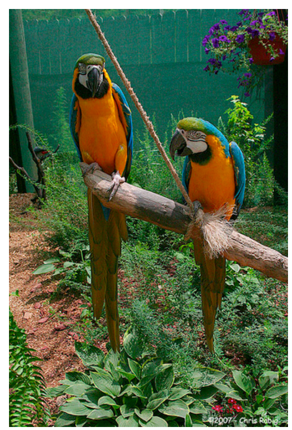
Here we see one potted plant.



This is a picture of one dog.

False detections:

There are one road and one cat. The furry road is in the furry cat.



This is a picture of one tree, one road and one person. The rusty tree is under the red road. The colorful person is near the rusty tree, and under the red road.

Incorrect attributes:

This is a photograph of two sheeps and one grass. The first black sheep is by the green grass, and by the second black sheep. The second black sheep is by the green grass.

This is a photograph of two horses and one grass. The first feathered horse is within the green grass, and by the second feathered horse. The second feathered horse is within the green grass.

Our Generation Algorithm vs Humans

"This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair, and near the potted plant."

↕

H1: A Lemonaide stand is manned by a blonde child with a cookie.
H2: A small child at a lemonade and cookie stand on a city corner.
H3: Young child behind lemonade stand eating a cookie.

Sounds unnatural!

Caption
guessing game

a) monkey playing in the tree canopy, Monte Verde in the rain forest

b) capuchin monkey in front of my window

c) monkey spotted in Apenheul Netherlands under the tree

d) a white-faced or capuchin in the tree in the garden

e) the monkey sitting in a tree, posing for his picture

Caption
guessing game

a) monkey playing in the tree canopy, Monte Verde in the rain forest

b) capuchin monkey in front of my window

c) monkey spotted in Apenheul Netherlands under the tree

d) a white-faced or capuchin in the tree in the garden

e) the monkey sitting in a tree, posing for his picture

Data Driven Generation

“Im2Text: Describing Images Using 1 Million Captioned Photographs”

Vicente Ordonez, Girish Kulkarni, Tamara L. Berg

NIPS 2011

“Collective Generation of Natural Image Descriptions”

Polina Kuznetsova, Vicente Ordonez,

Alexander C. Berg,Tamara L. Berg and Yejin Choi

ACL 2012

Through the smoke

Duna Portrait #5

Mirror and gold

the cat lounging in the sink

Data exists, but buried in junk!

Captions in the Wild

http://tamaraberg.com/sbucaptions

The Egyptian cat statue by the floor clock and perpetual motion machine in the pantheon

Little girl and her dog in northern Thailand. They both seemed interested in what we were doing

Interior design of modern white and brown living room furniture against white wall with a lamp hanging.

Man sits in a rusted car buried in the sand on Waitarere beach

Our dog Zoe in her bed

Emma in her hat looking super cute

Ordonez et al, NIPS 2011

Harness the Web

Ordonez et al, NIPS 2011

Captioned Photo Dataset
1 *million* captioned images!

Global Matching
(GIST + Color)

The bridge over the lake on Suzhou Street.

Bridge to temple in Hoan Kiem lake.

A walk around the lake near our house with Abby.

Smallest house in paris between red (on right) and beige (on left).

Hangzhou bridge in West lake.

The daintree river by boat.

...

Transfer Whole Caption(s)

e.g. “The bridge over the lake on Suzhou Street.”

Transfer pieces of Captions

Object appearance → NP: the dirty sheep

Object pose → VP: meandered along a desolate road

Scene appearance → PP: in the highlands of Scotland

Region appearance & relationship → PP: through frozen grass

Example Composed Description: the dirty sheep meandered along a desolate road in the highlands of Scotland through frozen grass

Kuznetsova et al, ACL 2012

Two Example Projects

Moving recognition outputs toward human-like predictions

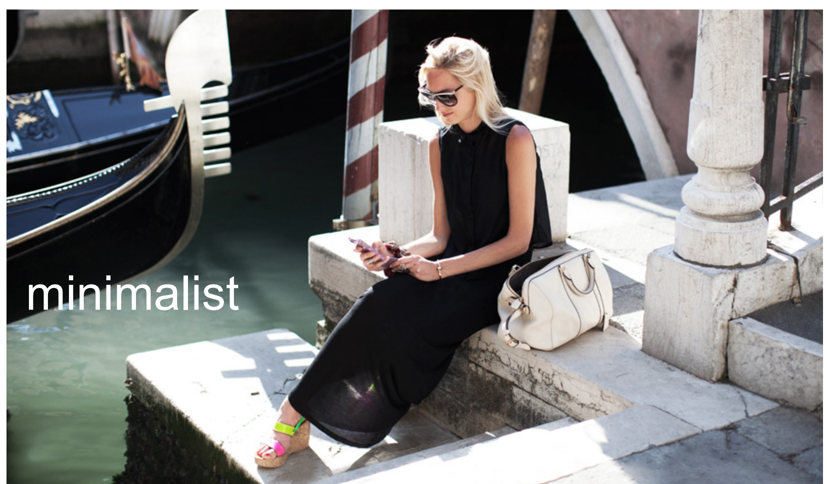
Extracting socio-identity information from pictures

Clothing & Identity…

Wealth

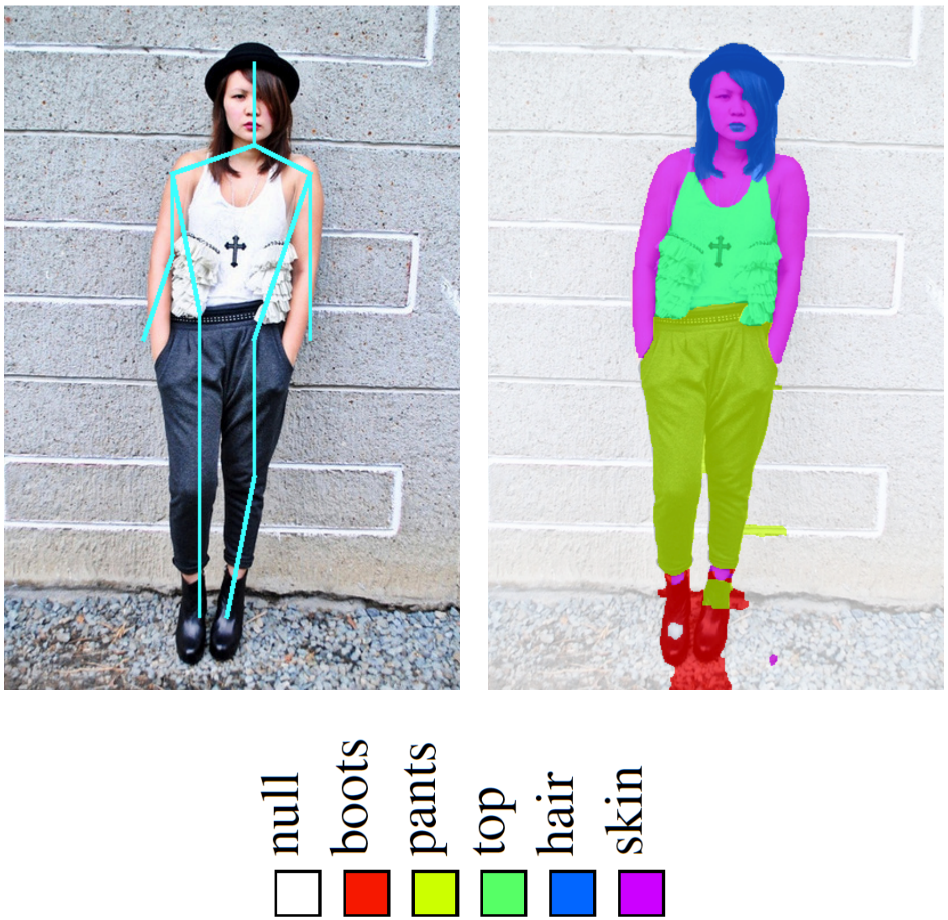
Occupation

Social Tribe

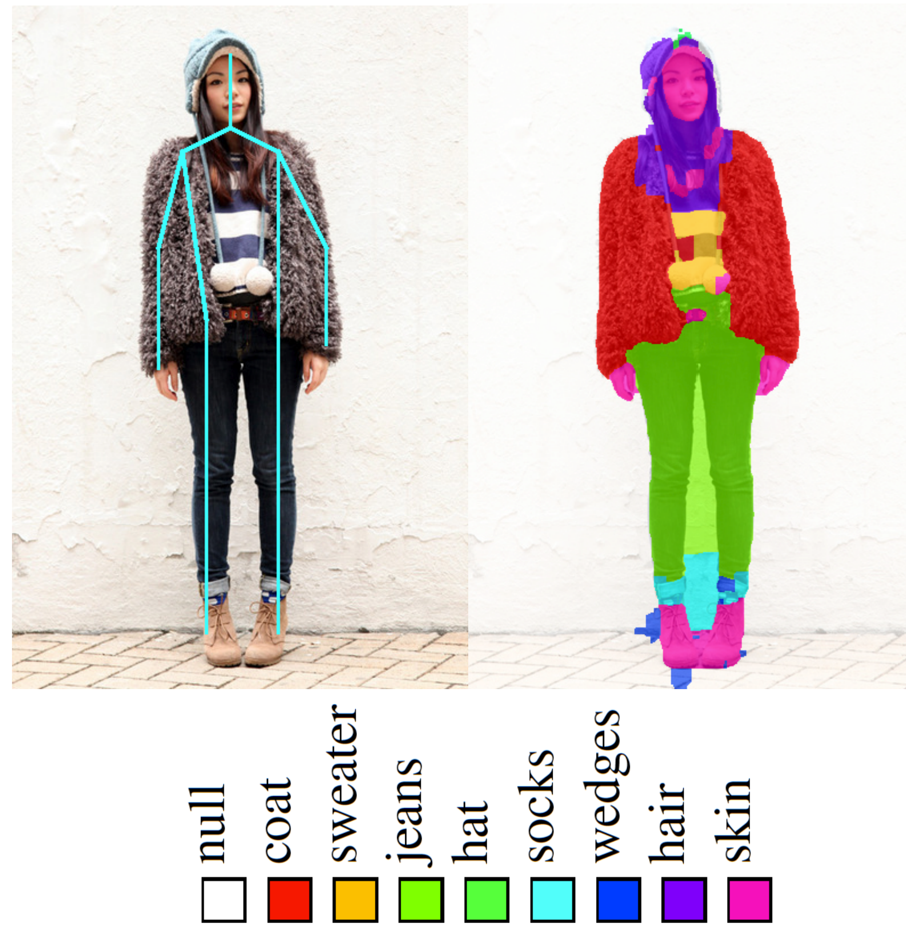


Clothing Recognition

"Parsing Clothing in Fashion Photographs"
Kota Yamaguchi, Hadi Kiapour, Luis E Ortiz Tamara L. Berg,
CVPR 2012.



"Paper Doll Parsing: Retrieving Similar Styles to Parse Clothing"
Kota Yamaguchi, Hadi Kiapour, Tamara L. Berg,
ICCV 2013.



Locally ambiguous

Large Variation in appearance









shirts

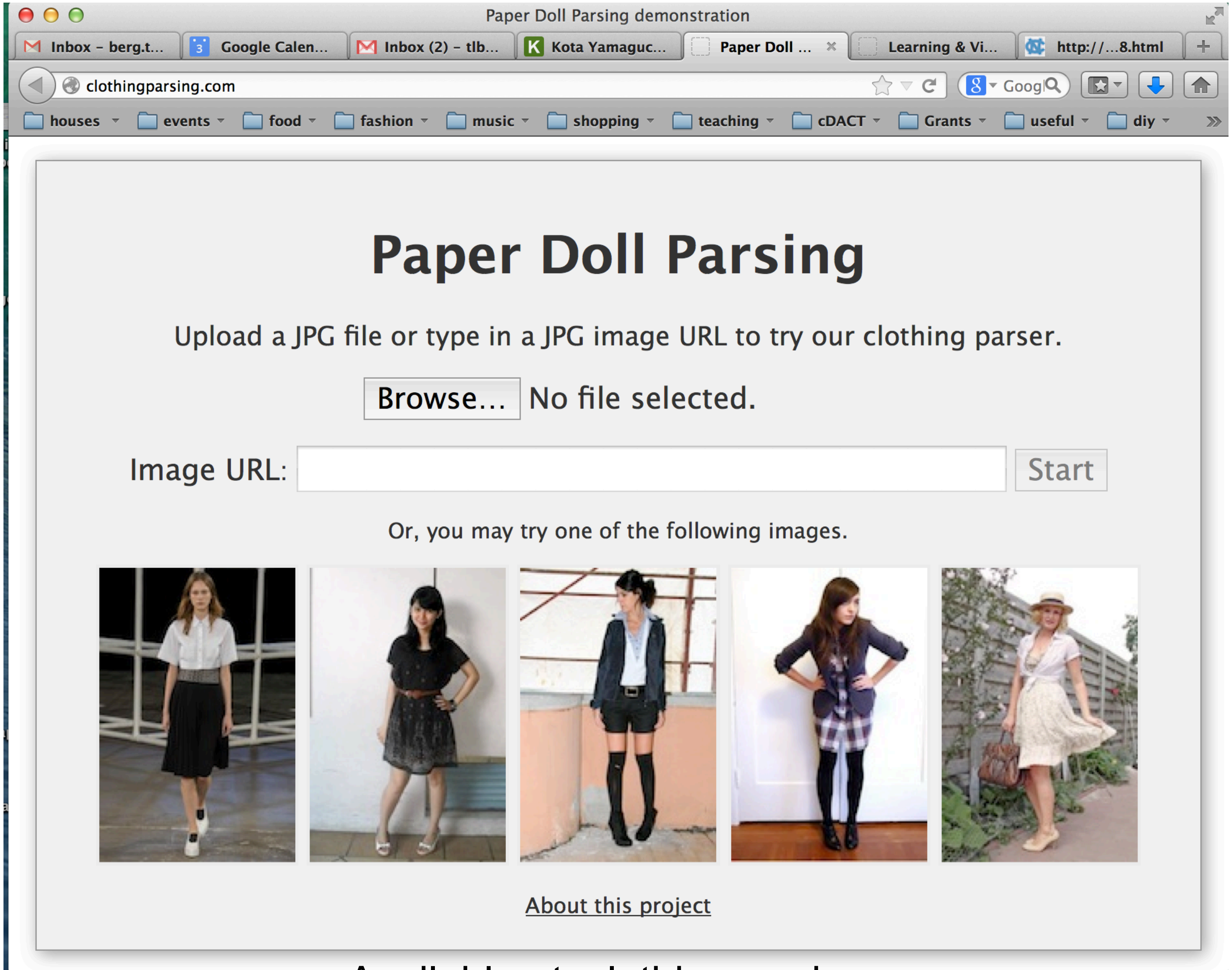
Layering

Key Intuition: Pose and Clothing are Related

POSE ESTIMATION

CLOTHING RECOGNITION

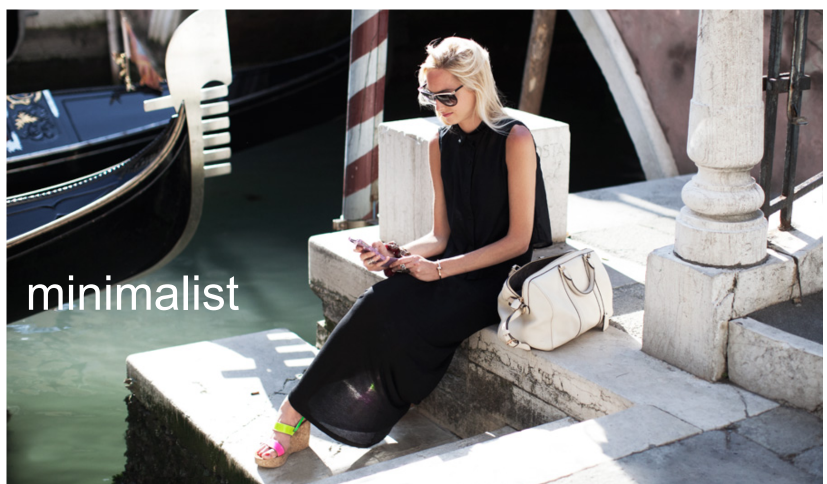
Recognizing Clothing - Demo



Available at: clothingparsing.com

Yamaguchi et al, ICCV 2013

Social Tribe



Collecting Labels

- How?

How hipster is he?

◯ 1 ◯ 2 ◯ 3 ◯ 4 ◯ 5

Collecting Labels

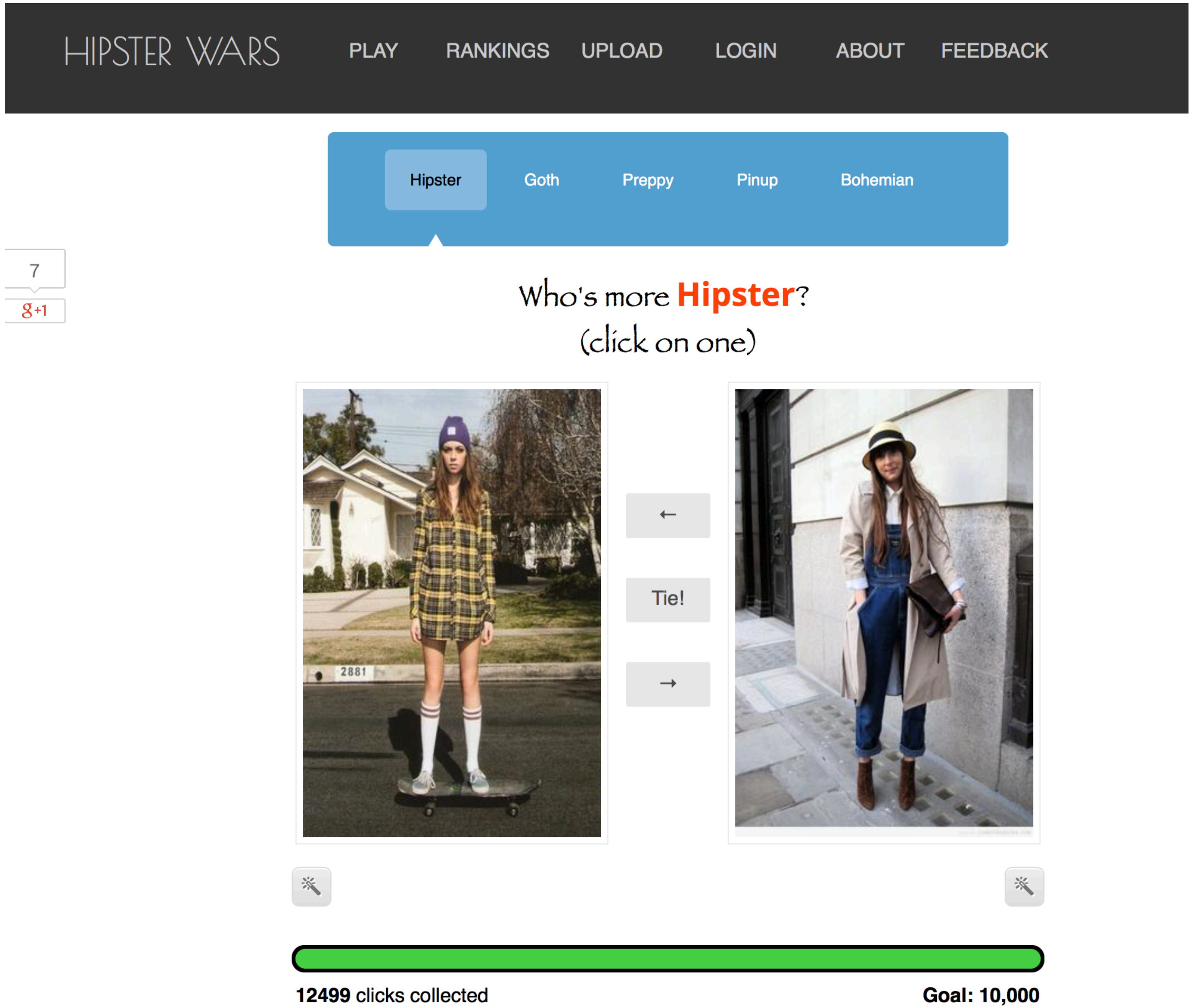
- How?

How [illegible] he?

People have different internal scales for rating.

Asking about individual images can produce unstable results.

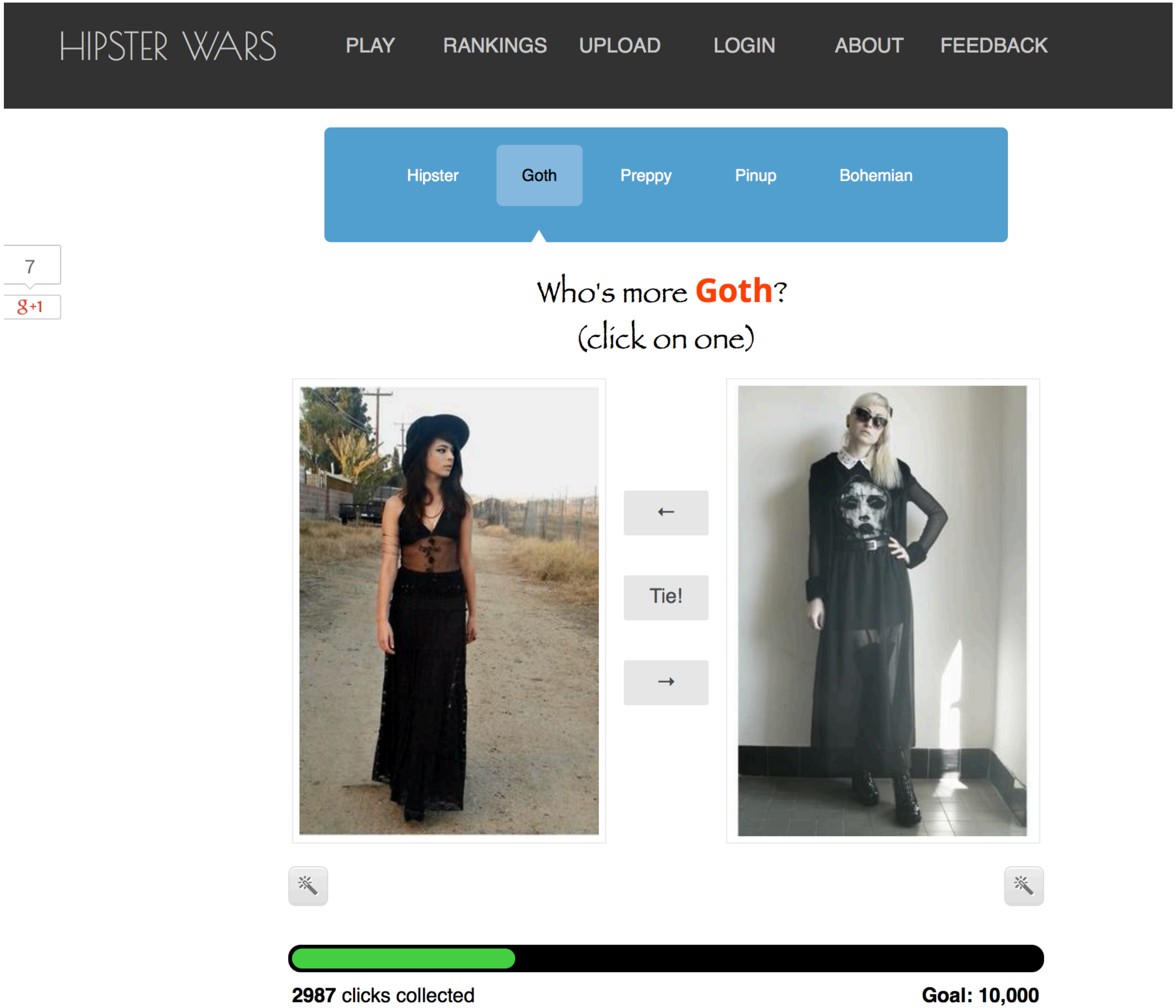
Collecting style ratings



Hipsterwars game!

Available at:
hipsterwars.com

Collecting style ratings

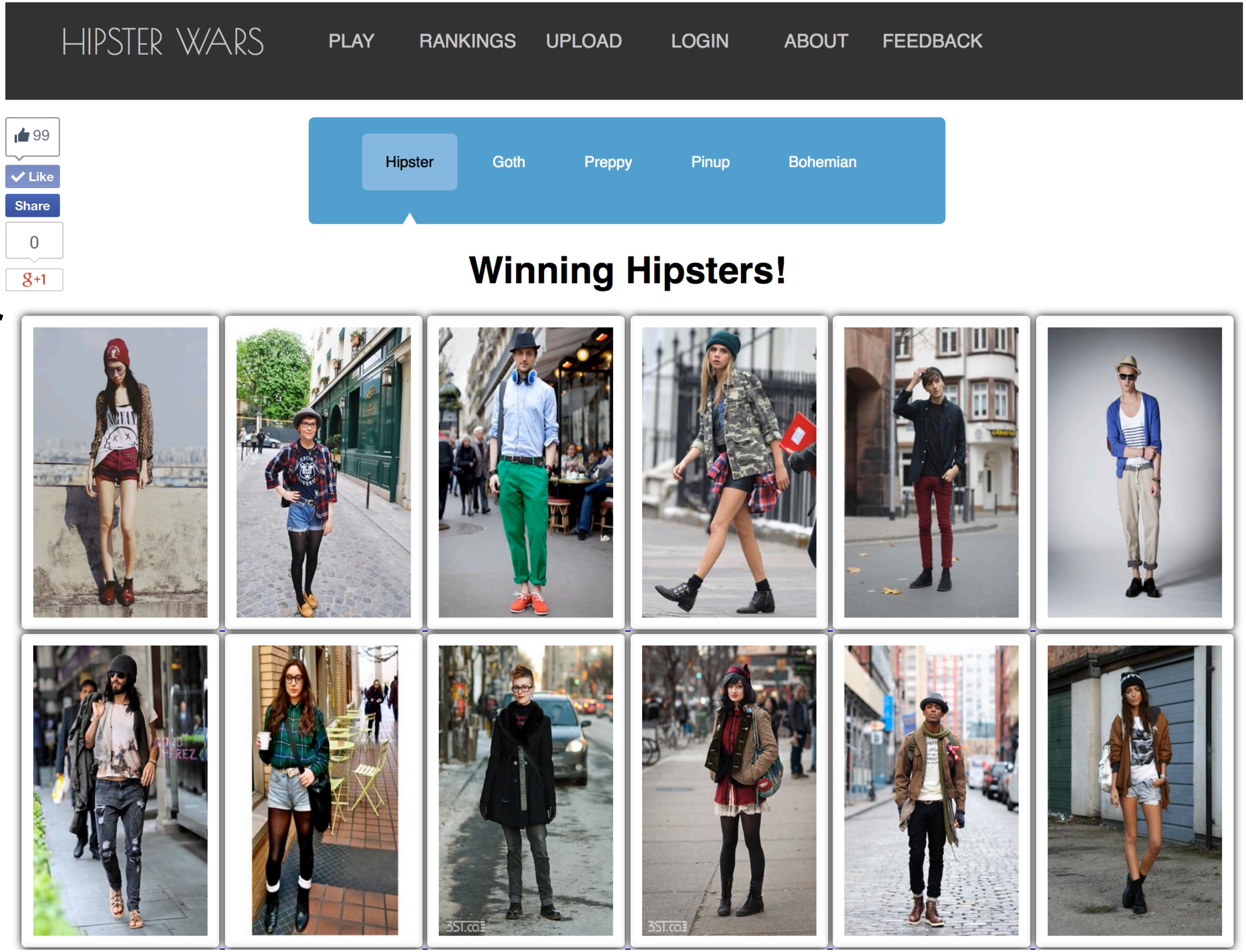


Range of styles:

hipster, goth, preppy, pinup, bohemian

Available at:
hipsterwars.com

Most hipster



… Least hipster

How hipster are you?

Thank you!

Vicente Ordonez, Kota Yamaguchi, Hadi Kiapour, Xufeng Han, Polina Kuznetsova, Siming Li, Girish Kulkarni, Visruth Premraj, Sagnik Dhar

Alex Berg, Yejin Choi, Luis Ortiz