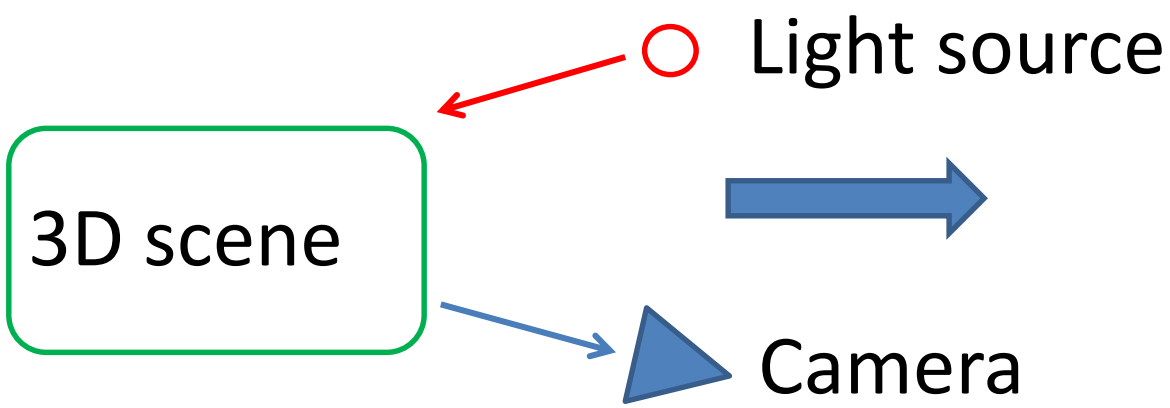
Introduction to Computer Vision

Srikumar Ramalingam
School of Computing
University of Utah
srikumar@cs.utah.edu

Course Website

- http://www.eng.utah.edu/~cs6320/

What is computer vision?

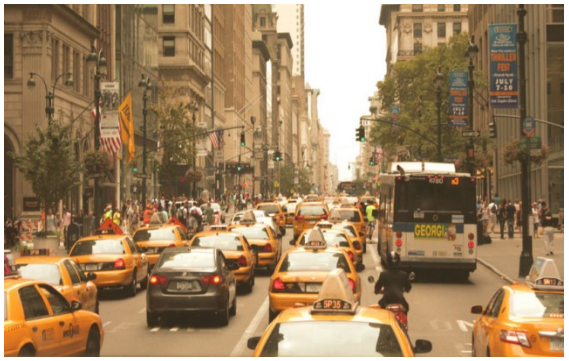


- Same image can not be captured twice
  - viewpoint, scene, and lighting changes
- What led to this image? (some kind of detective work!)
  - geometry, objects (humans, cars), lighting, camera pose
- “Vision is putting the toothpaste back into the tube” – John Mayhew

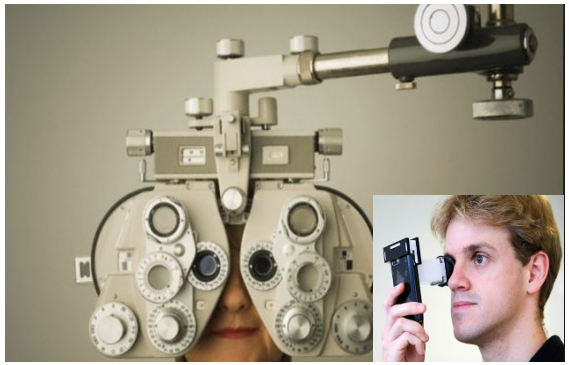
What can you infer?

- Where is this photo taken?
- When was this photo taken?
- What is the height of the photographer?

Applications of computer vision



Autonomous driving



Health

Factory automation

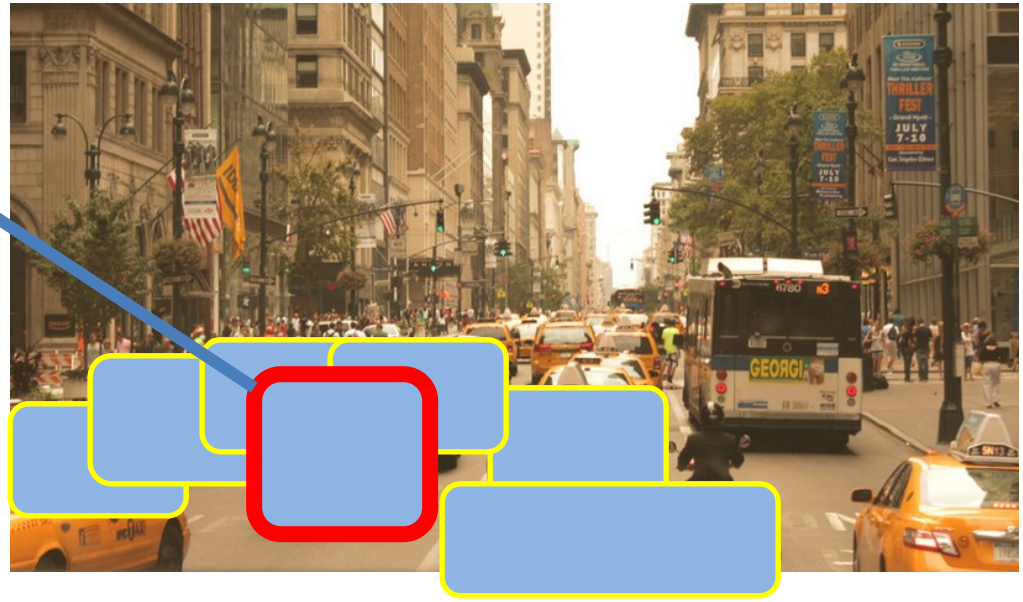
Gaming

Home appliances
(camera + vacuum cleaner)

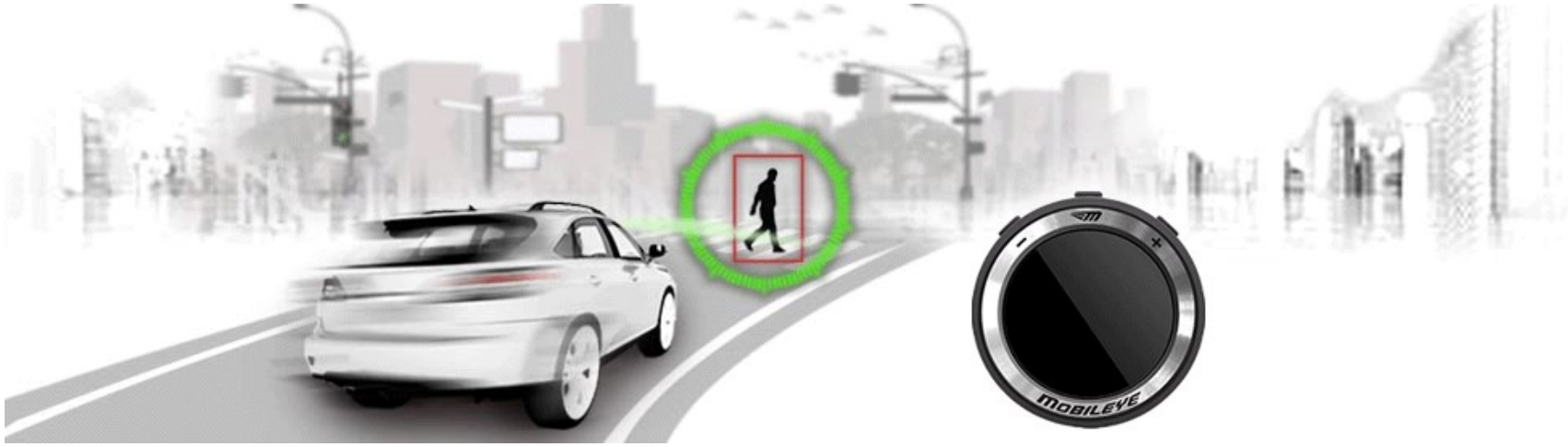
Surveillance

Is autonomous driving hard?

- Location
- Path planning
- Surroundings
- Predicting the behavior of neighbors
  – recursive!



Mobileye

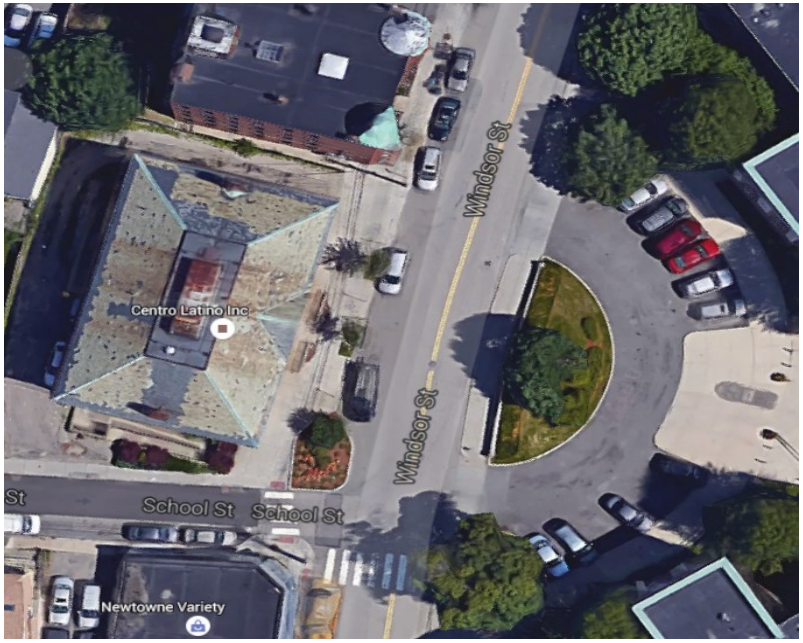


[Image courtesy: Mobileye]

Provides several functionalities like pedestrian collision warning, lane departure warning, etc.

Google Earth

High resolution 3D model



Aerial image with roads marked

[Courtesy: Google Earth and Maps]

Face detection

[Image courtesy: Szeliski]

- Most digital cameras can detect faces now.

[Paul Viola and Michael Jones, 2001]

Reconstructing the interiors of museums for Google Earth

Jianxiong Xiao and Yasutaka Furukawa

Exploring Image Collections in 3D

Microsoft Photo Tourism Project
Noah Snavely, Steve Seitz, and Rick Szeliski

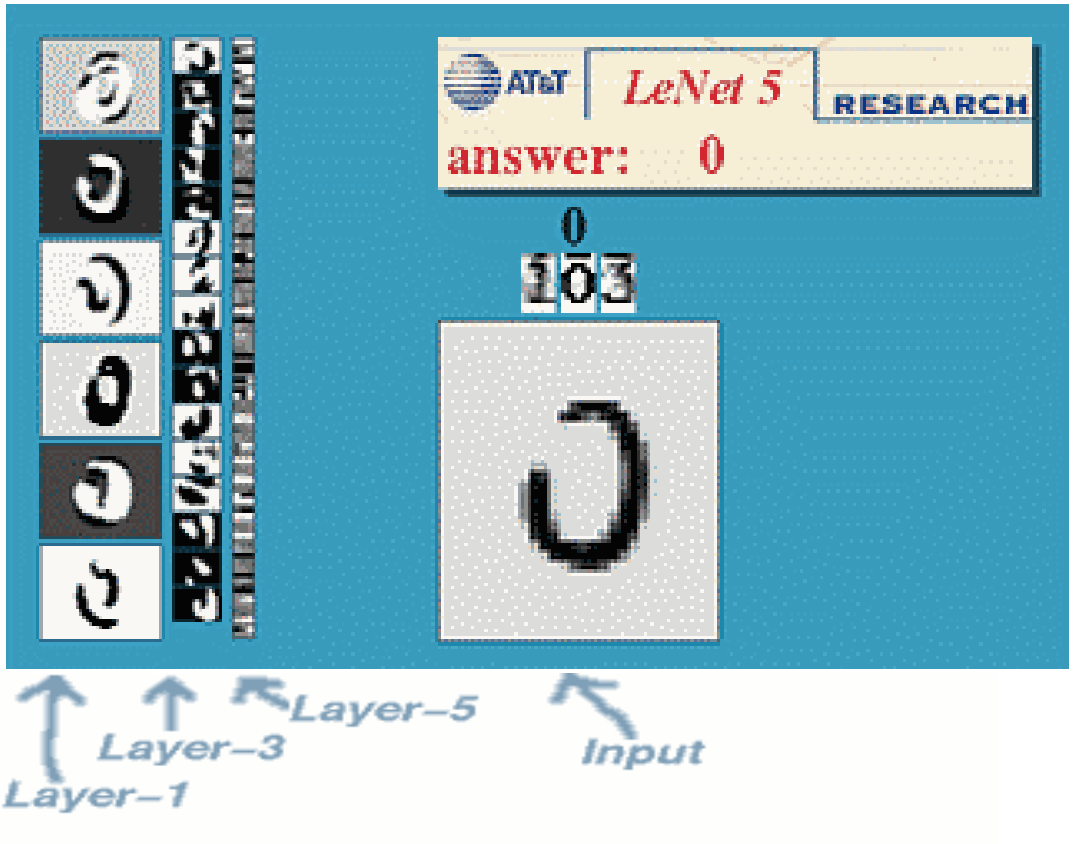
Microsoft Kinect: Gaming and beyond

Depth image

Segmented body parts

[Shotton et al. 2011]

Optical character recognition (OCR)

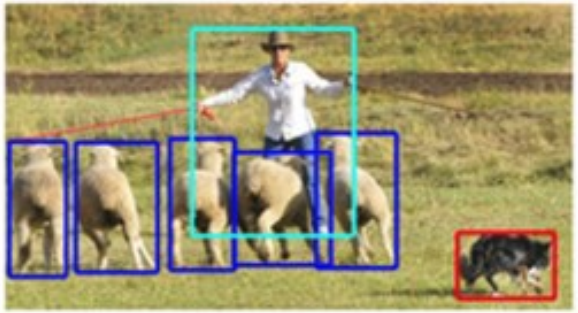


[Image courtesy: http://yann.lecun.com/exdb/lenet/ Digit Recognition, AT & T Labs]

Applications of OCR: reading checks, postal addresses, assisting blind people, reading number plates, etc.

Interpretation of images from social networks

(a) classification



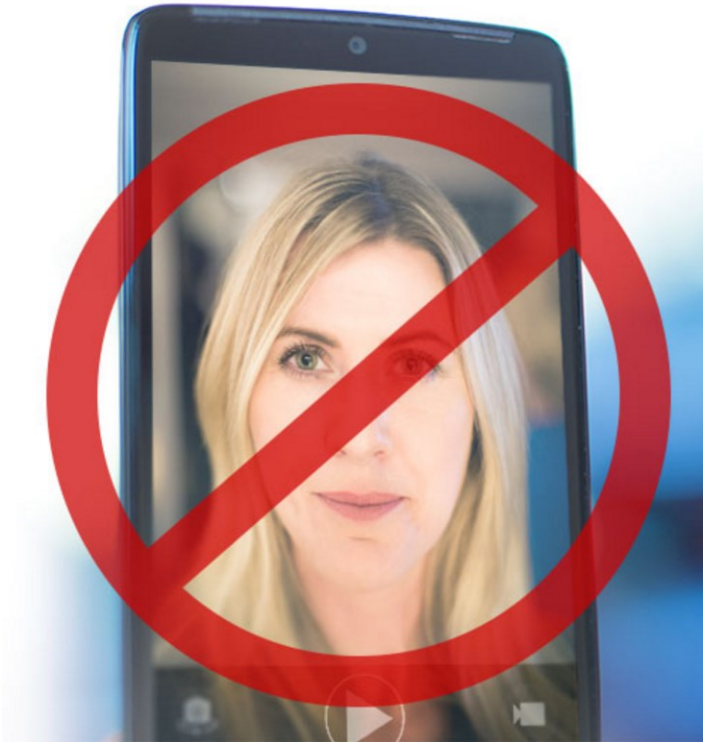
(b) detection

(c) segmentation

[Image courtesy: Facebook]

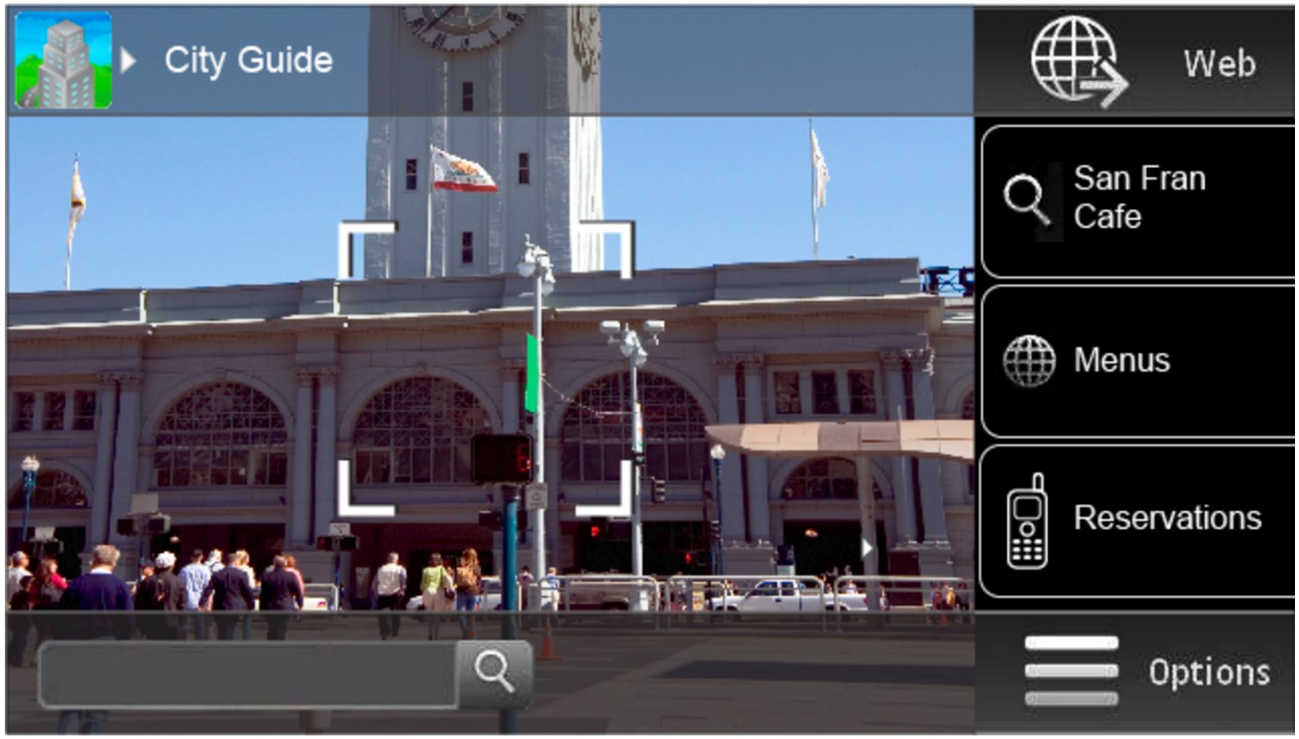
~1.6 billion people and growing fast!

Face authentication can not be fooled



[www.sensiblevision.com]

Object recognition (in mobile phones)



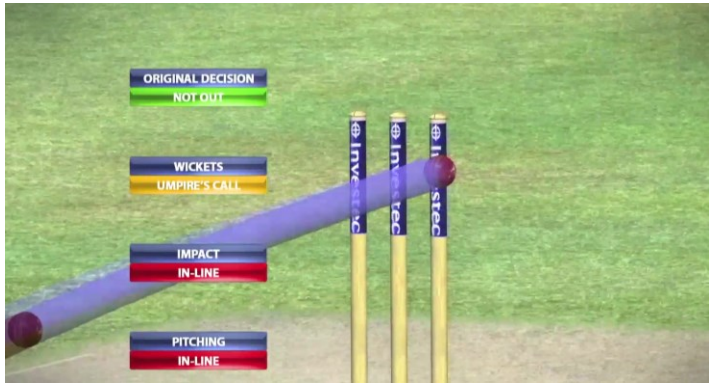
Point and Find App, Nokia

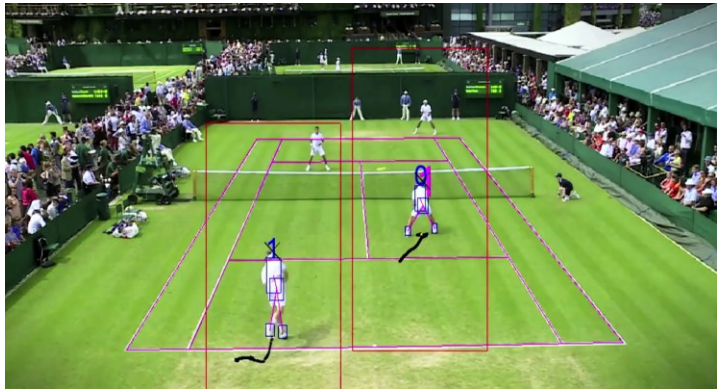
3D Modeling from a collection of Images

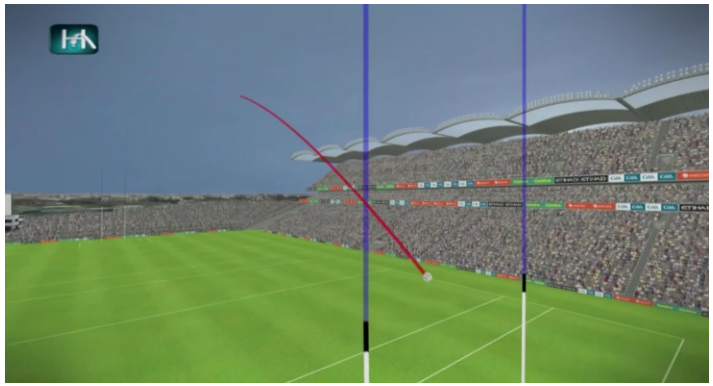
[From Debevec's PhD Thesis: The Campanile Movie]

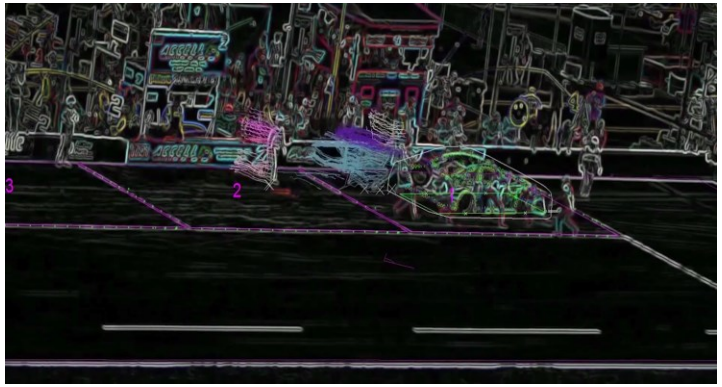
The Bullet Time: The Matrix

Sports Analysis









- Some kind of tracking of players and balls!

[http://www.hawkeyeinnovations.co.uk/]

StreetScore

[https://www.media.mit.edu/projects/streetscore/overview/]

Amazon Go checkout-free convenience store opens to public

- Uses computer vision and deep learning technologies

Facebook's efforts in fighting video piracy

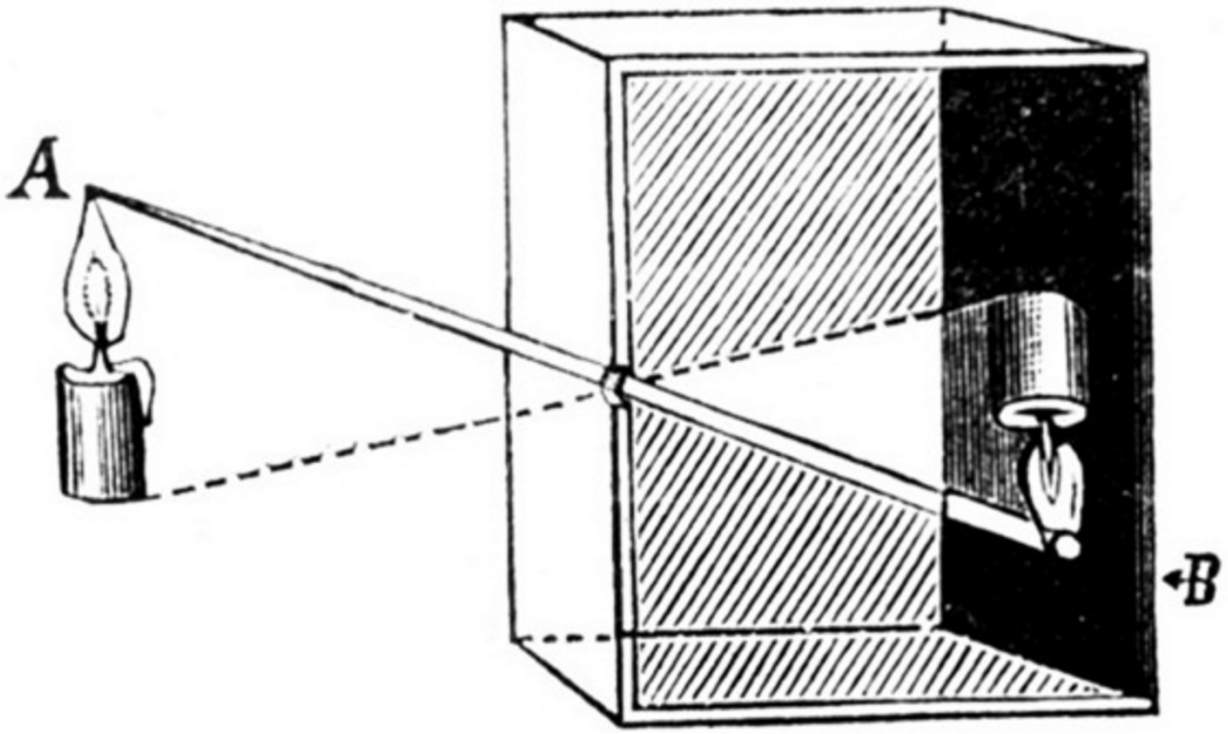
- How to automatically identify the right owners of media content shared without permission?

Fight against fake news – Google and Facebook

Charlie Hebdo march in Paris: Some people were edited out of the photo.
Source: blogs.mathworks.com

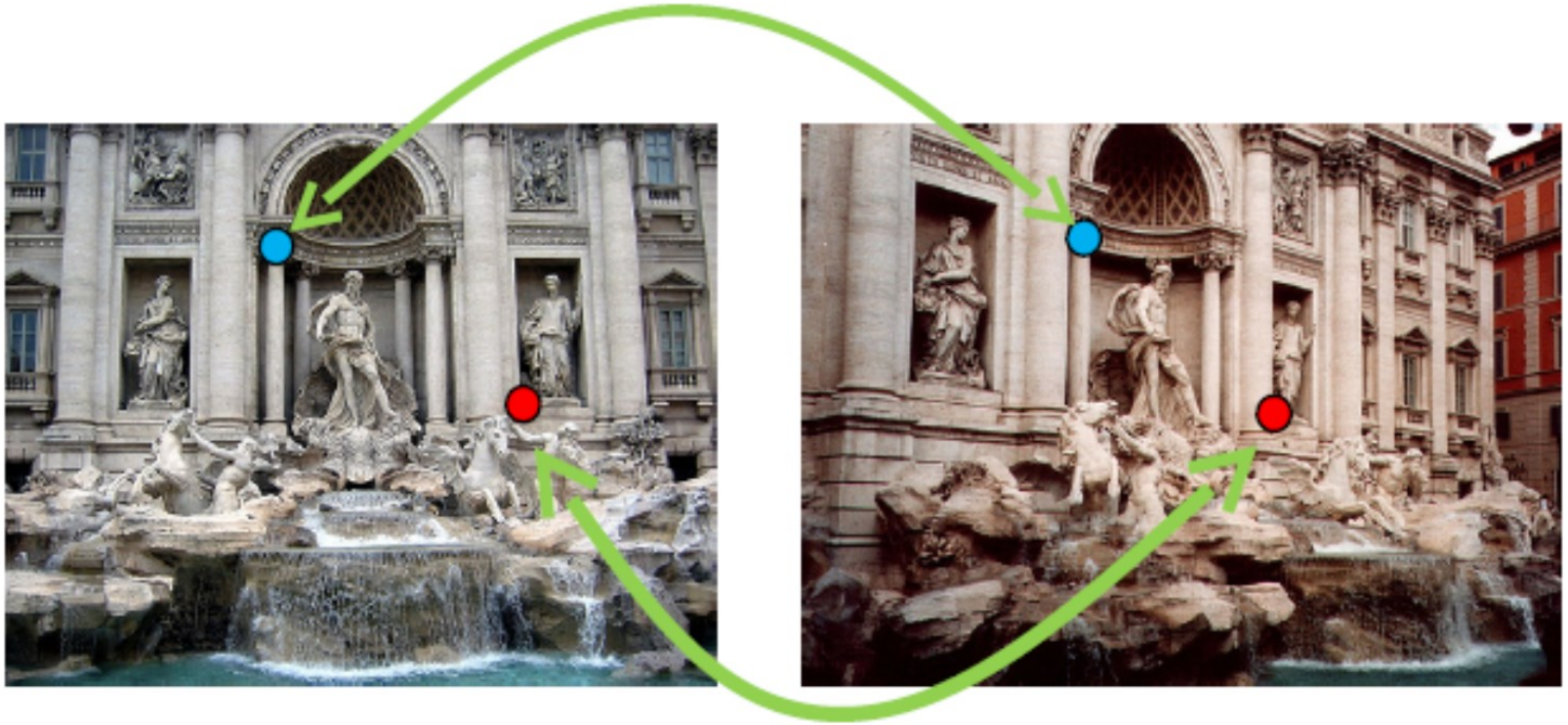
- How to detect image tampering?

Camera Models and Image Formation



- Camera obscura, pinhole model, projection of 3D scene on 2D images, etc.

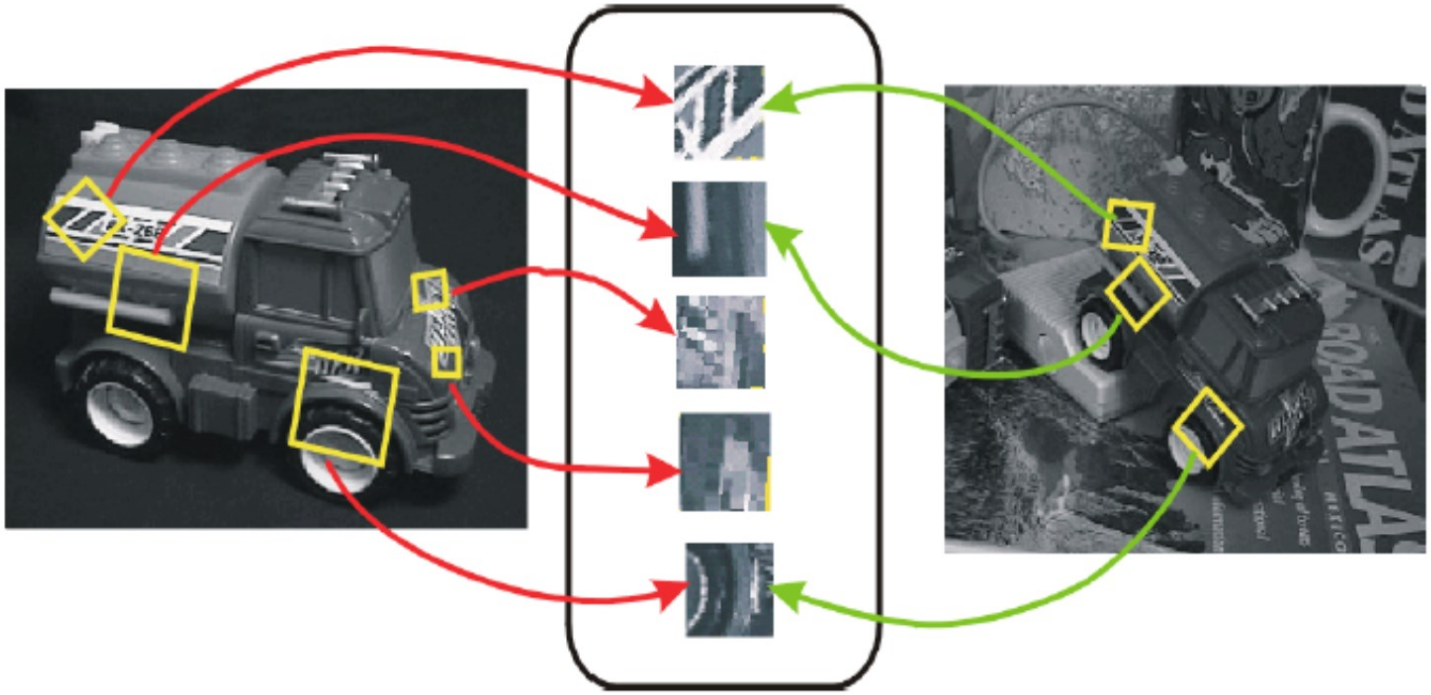
Feature detectors and matching



Keypoint matching under varying illumination

- Basic understanding of keypoints, feature descriptors, and matching

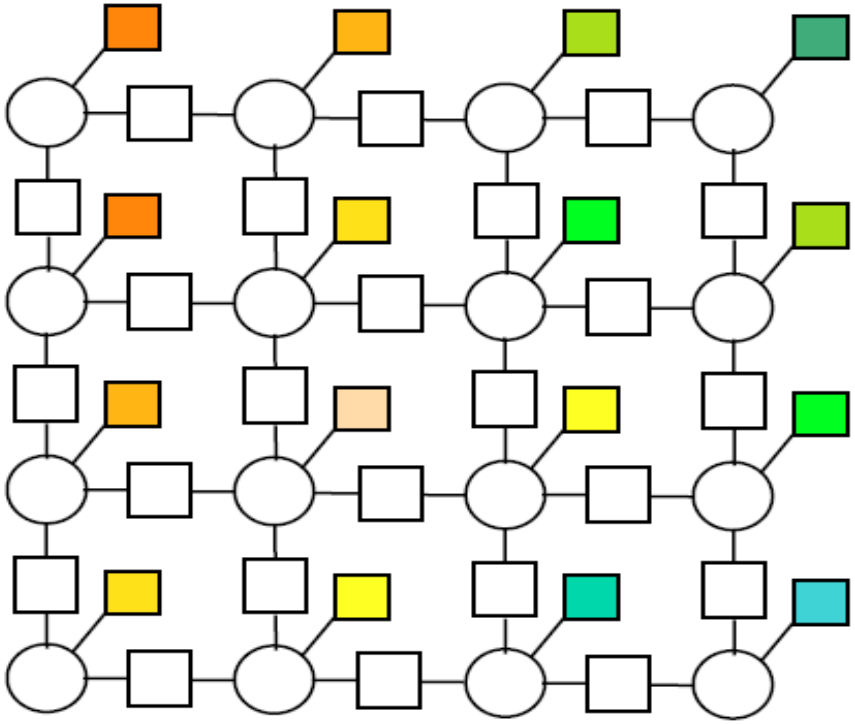
Feature detectors and matching



Keypoint matching under varying viewpoints

- Basic understanding of keypoints, feature descriptors, and matching

Inference problems



Goal: find most probable interpretation of scene

[Courtesy: J. Yedidia]

- Inference on Markov Random Fields

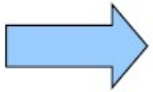
Simple Segmentation problems with 2-labels

[Boykov and Jolly'2001, Rother et al. 2004]

Multi-label problems

Left Camera Image

Right Camera Image



Dense Stereo Result

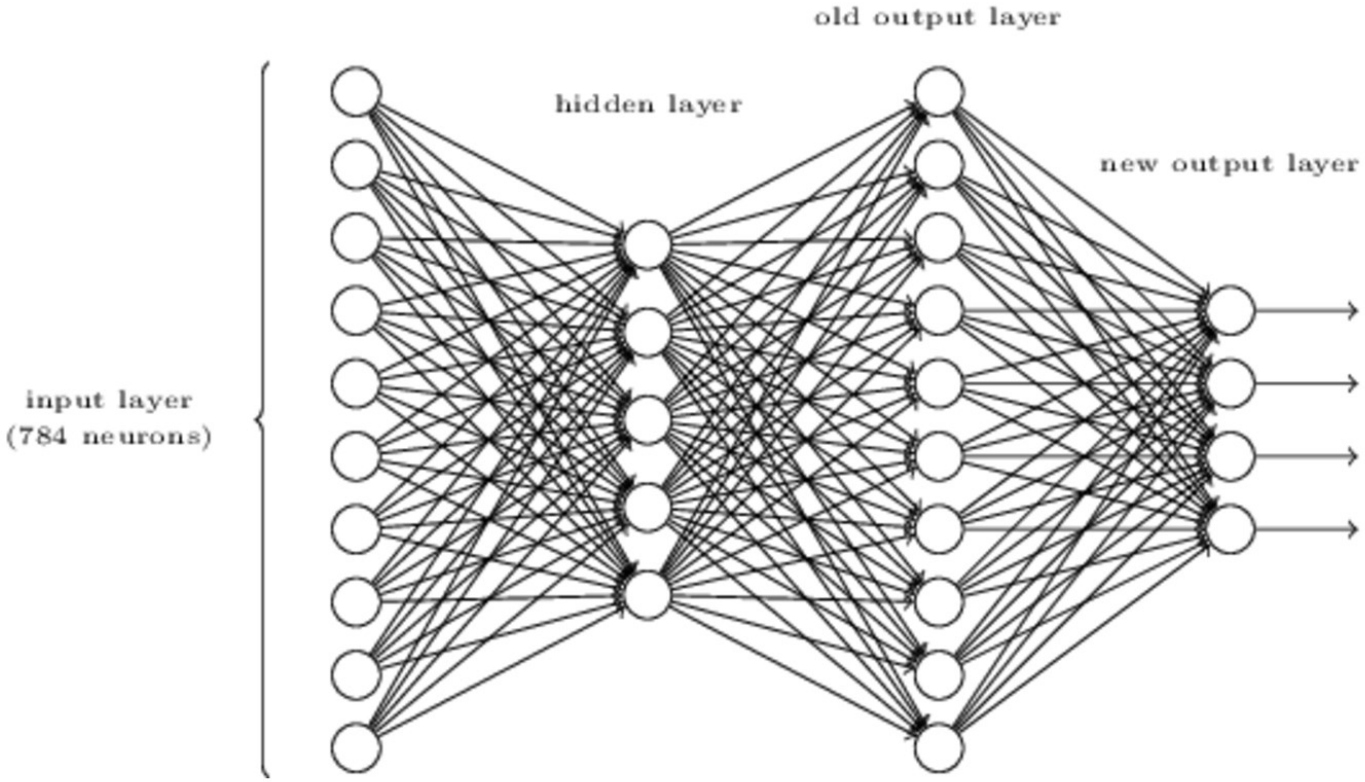
- Choose the disparities from the discrete set: $(1,2,\ldots,L)$

Human Detection

[Dalal and Triggs, 2005]

Object Class Segmentation

Deep Neural Networks



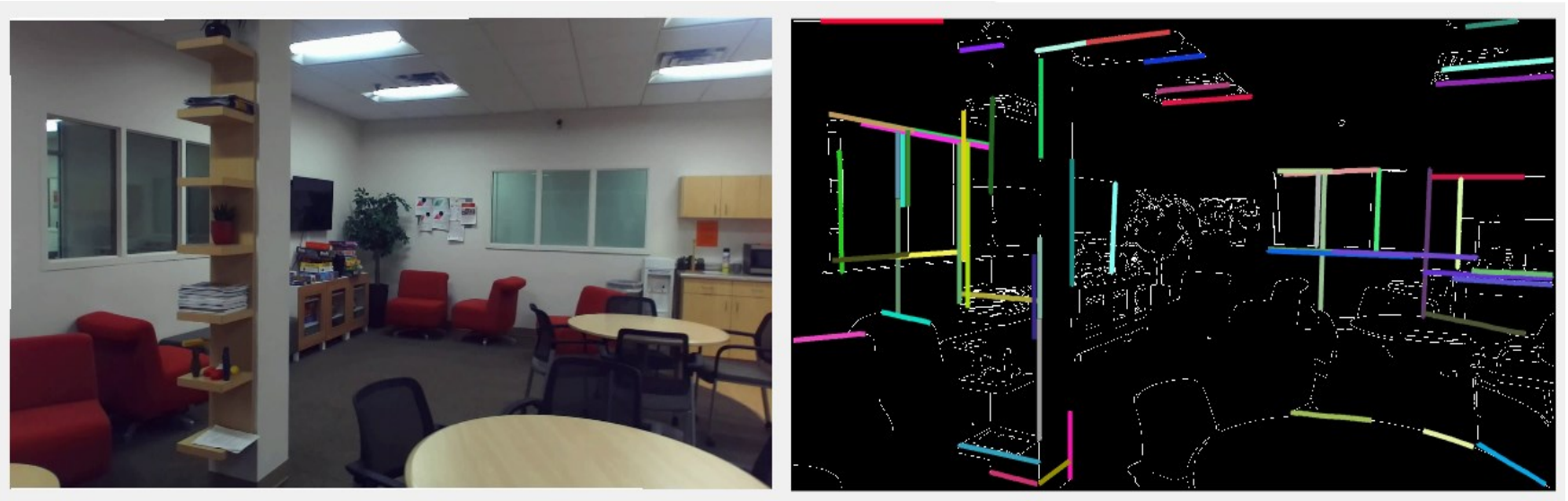
http://neuralnetworksanddeeplearning.com/

- We will discuss some of the basics and learn to use the existing deep learning packages such as caffe or tensorflow.

Homework 1

- Line detection
- Segmentation
- Stereo Reconstruction

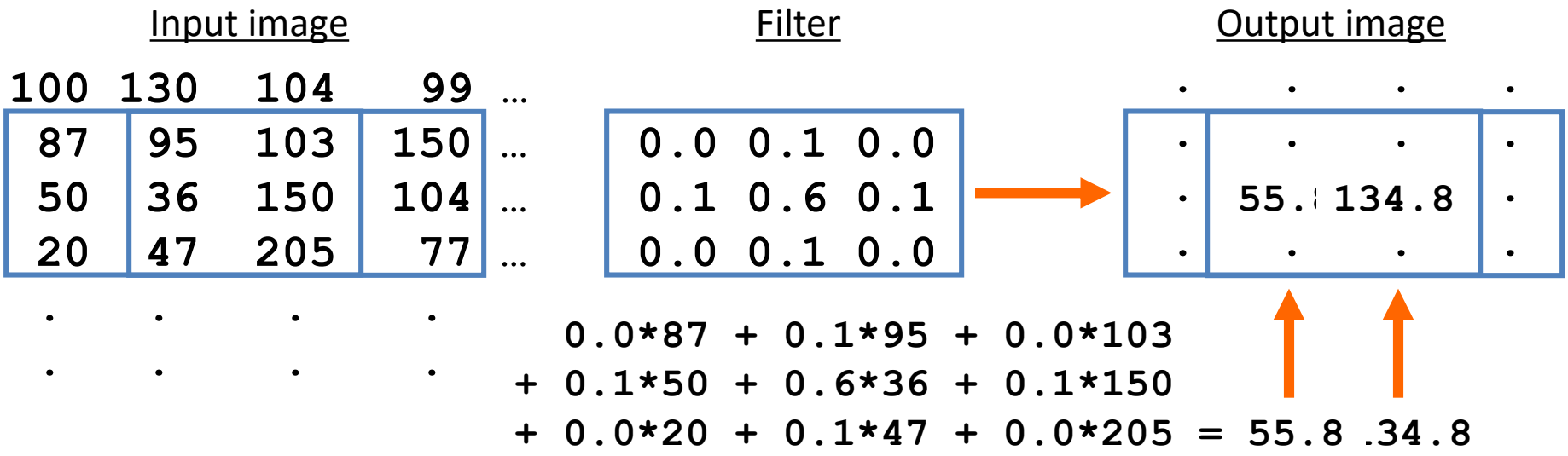
Line Detection



Goal : Given an image, extract the line segments from it.

Image Filtering

- Operation on image neighborhood and small …
  - "mask", "filter", "stencil", "kernel"
- Linear operations within a moving window



Examples 1

$$\begin{matrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{matrix}$$

$$1/9 * \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix}$$

Examples 2

$$\frac{1}{9} * \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

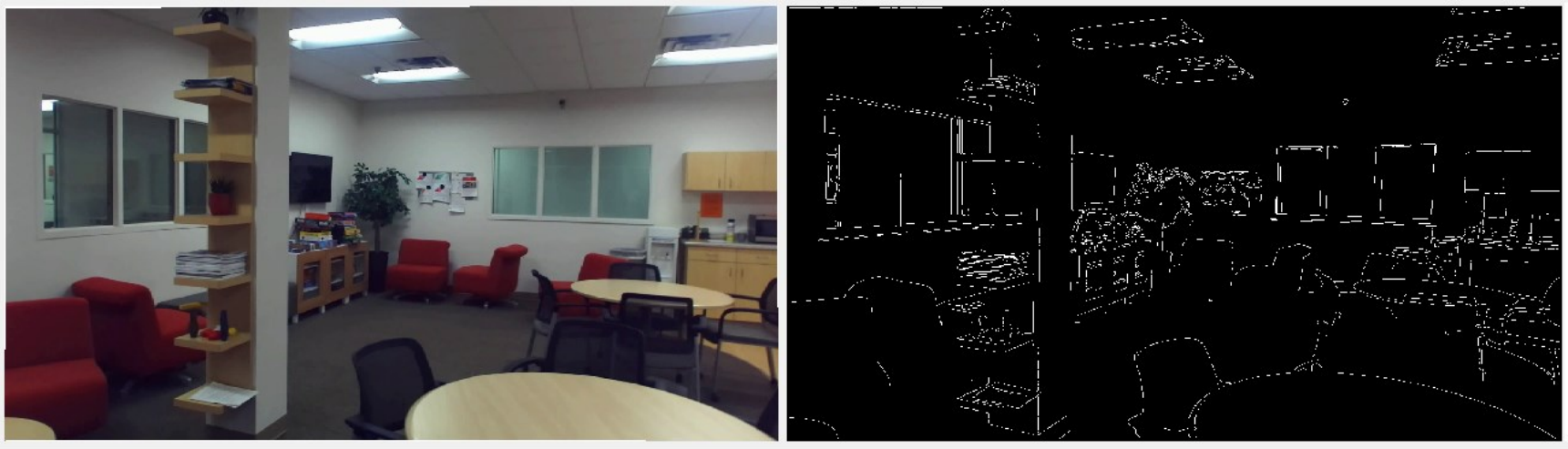
$$\frac{1}{25} * \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Derivative Example

$$\begin{matrix} 0 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{matrix}$$

$$\begin{matrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{matrix}$$

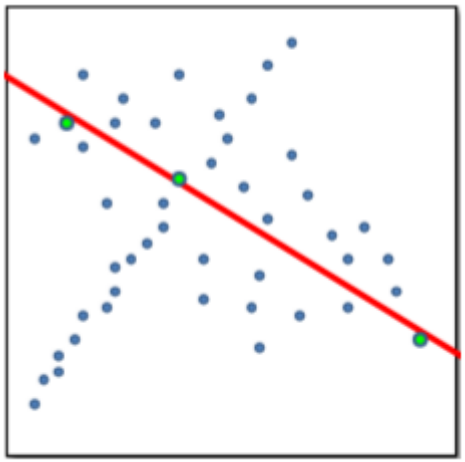
Canny edge detection

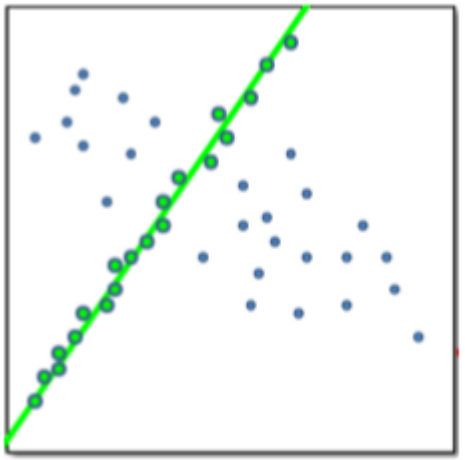


- Single command in Matlab:

```
Img = imread('lineDetectionImage.bmp');
Iedge = edge(rgb2gray(Img),'canny',0.1);
```

Line Fitting





- Select 2 points randomly and obtain a line equation. Check the # of points that lie on this line. Iteratively find line segments that have sufficient # of points.

Segmentation

Left: Original Image, Right: Segmentation mask for the sky (White corresponds to Sky, and Black corresponds to non-sky region).

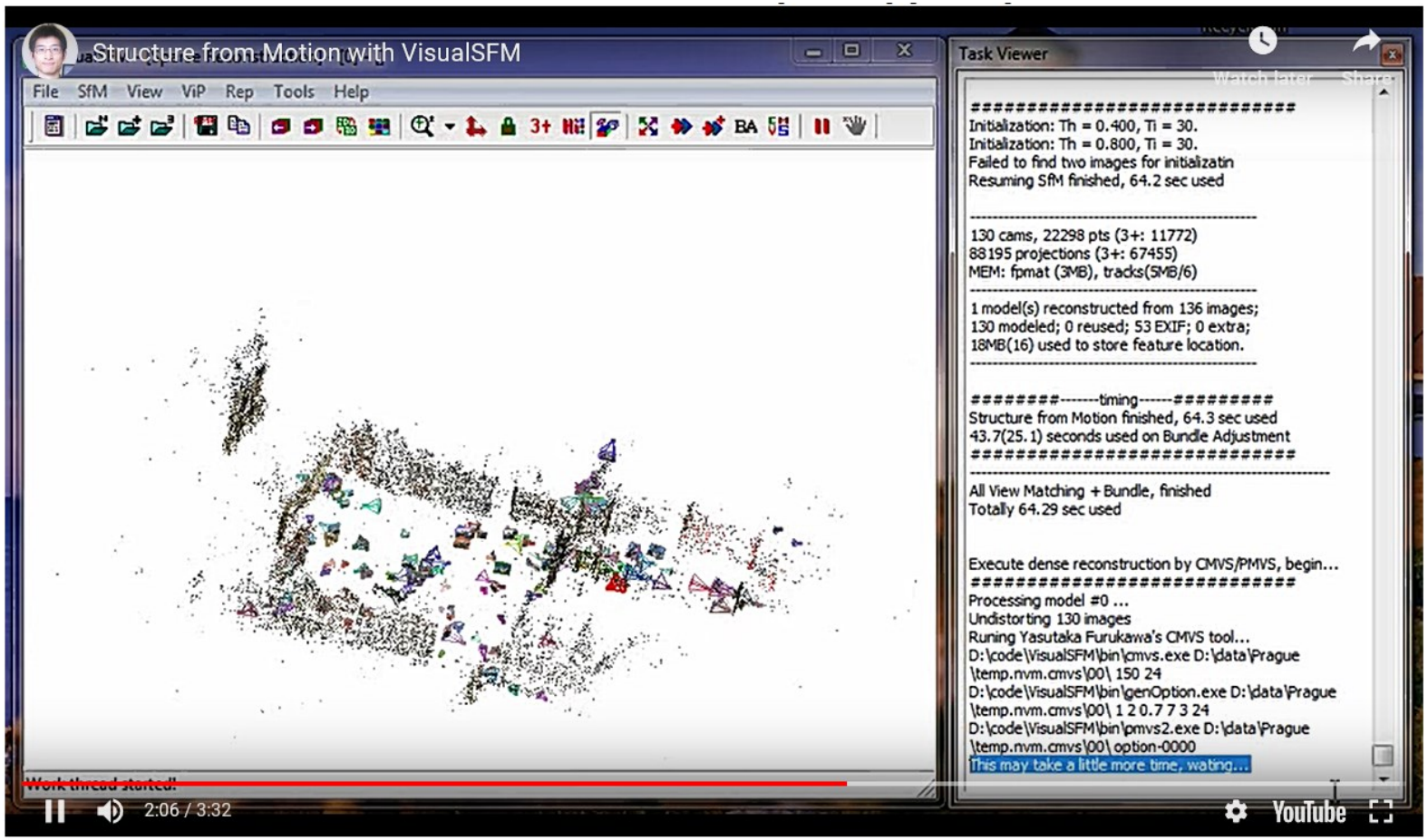
- Loop over every pixel in the image and if the three components (R,G,B) lie within the specified range, treat it as sky and give White color (255,255,255). Otherwise treat it as non-sky and give black color (0,0,0).

Stereo Reconstruction

Left: left image, Middle: right image, Right: Disparity map (white corresponds to large disparity and darker gray regions correspond to small disparity values). We searched for disparity values ranging from 1 to 50 for each pixel.

Simple Stereo : For every pixel in the left image, we search for the corresponding pixel in the right image at the same vertical height that provides the best NCC score.

VisualSFM : A Visual Structure from Motion System



Source: http://ccwu.me/vsfm/

Thank You!