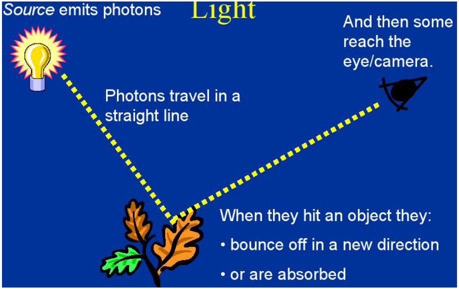
Introduction to Computer Vision

Mårten Björkman

Division of Robotics, Perception and Learning
School of Electrical Engineering and Computer Science

September 23, 2020

What does it mean to see?


Source emits photons
Light
And then some reach the eye/camera.
Photons travel in a straight line
When they hit an object they:
• bounce off in a new direction
• or are absorbed

- Vision is an active process for deriving efficient symbolic representations of the world from the light reflected from it.
- Computer vision: Computational models and algorithms to solve visual tasks and interact with the world.

Why is vision relevant?

Safety

Health

Security

Comfort

Fun

Access

Many applications where vision is the only good solution.

Computer vision examples

Figure: Google self-driving cars

Computer vision examples

Figure: Tracking in 1000 Hz (Tokyo Uni)

Computer vision examples

Figure: Fast book scanning (Tokyo Uni)

Some companies with people from RPL

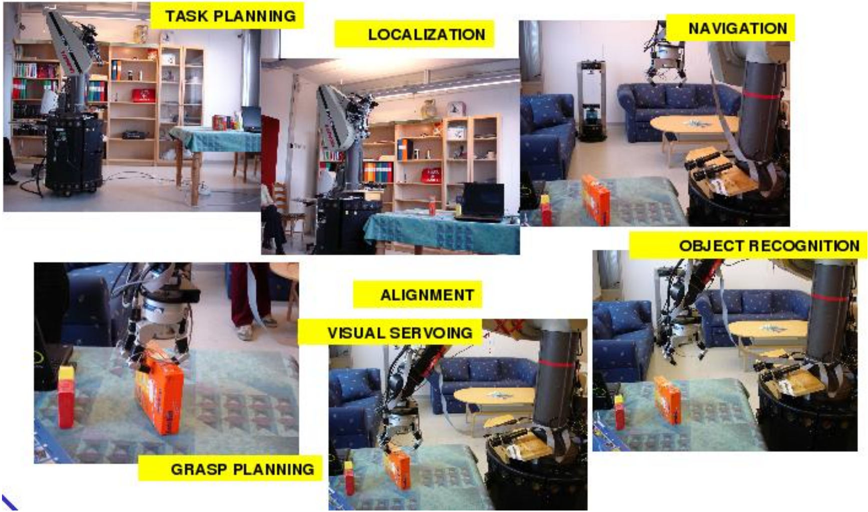
13th Lab (Facebook)

Tracab (Chyron Hego)

Volumental

Univrses

Example - A robot task

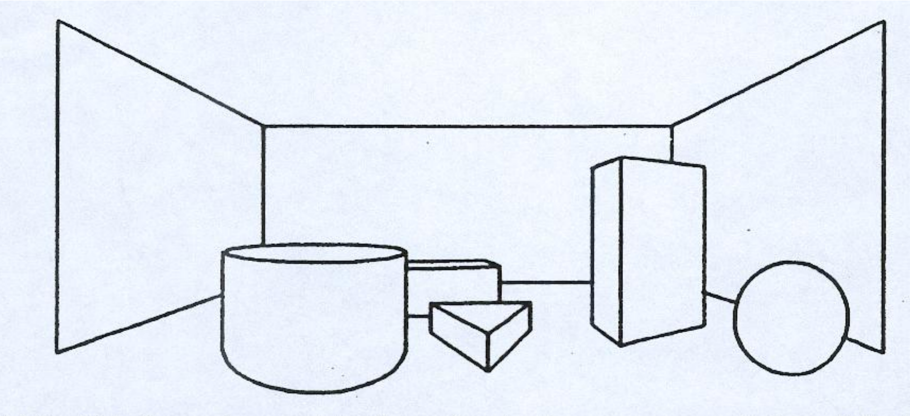


Why is vision interesting?

- Intellectually interesting
  - How do we figure out what objects are and where they are?
  - Harder to go from 2D to 3D (vision), than from 3D to 2D (graphics).
- Psychology:
  - After all, about 50% of cerebral cortex is for vision.
  - Vision is (to a large extent) how we experience the world.
- Engineering:
  - Intelligent machines that interact with the environment.
  - Computer vision opens up for multi-disciplinary work.
  - Digital images are everywhere.

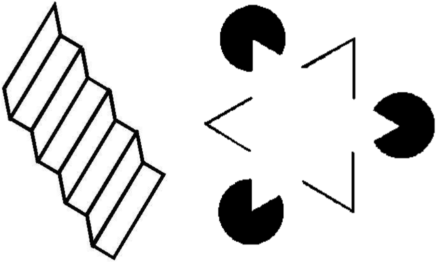
The importance of discontinuities

What are the explanations for the discontinuies you see?



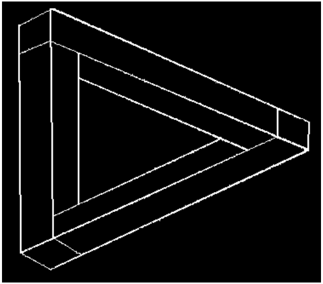
Human vision is not perfect!

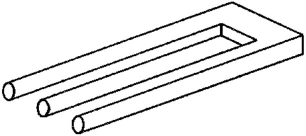
Reversing staircase illusion and subjective contours:



- Our perceptual organization process continues after providing a (first) interpretation. Continue viewing the reversing staircase illusion and you will see it flip into a second staircase.

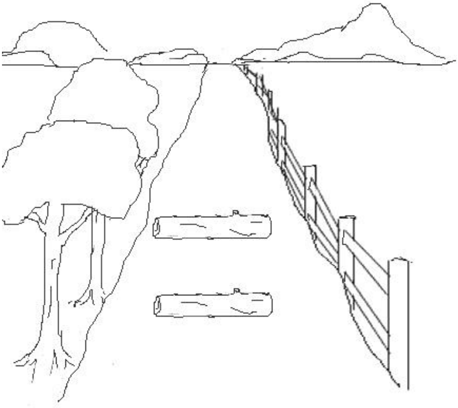
Impossible objects





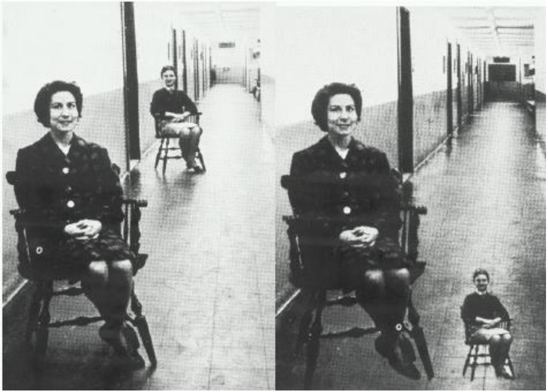
Another example that vision is an ongoing process.

Depth illusion - size constancy

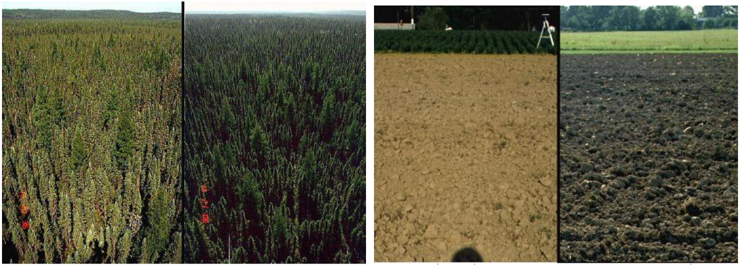


We tend to “normalize” things, such as size, shape and colors.

Depth illusion - size constancy



Light source examples



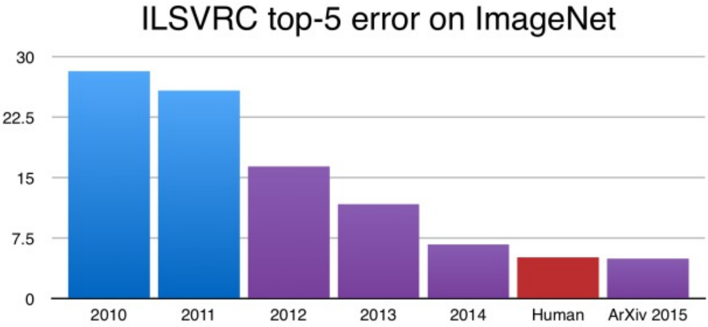
Left: sun behind observer, Right: sun opposite observer

Multi-disciplinarity

- Neuroscience / Cognition: how do animals do it?
- Philosophy: why do we consider something an object? (Hard!)
- Physics: how does an image become an image?
- Geometry: how does things look under different orientations?
- Signal processing: how do you work on images?
- Probability / Statistics: deal with noise, develop models.
- Numerical methods / Scientific computing: do this efficiently.
- Machine learning / AI: how to draw conclusions from lots of data?

Deep Networks for image classification

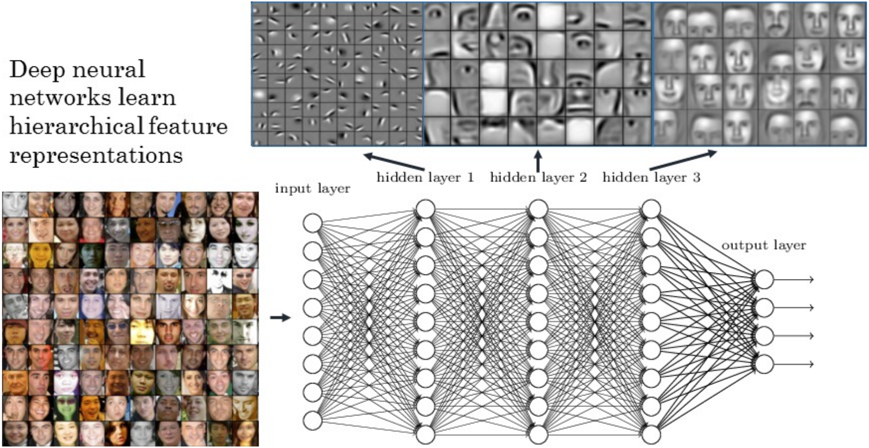
- Neural networks were long forgotten in computer vision.
- Recently, deep neural networks have become state-of-the-art.
- Superior on most challenging benchmarks (1K+ classes)



The best methods today have a top-5 error of about 3%.

Fully Connected Neural Network (FCN)

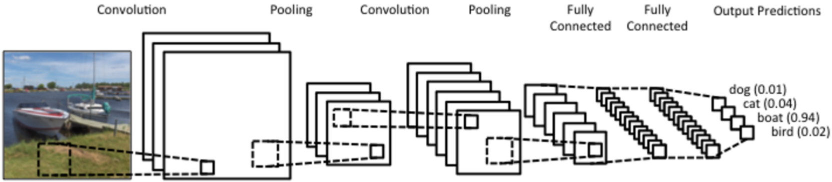
- Neural networks typically consists on layers of neurons
- Possibly images on inputs, some hypotheses on outputs



Deep neural networks learn hierarchical feature representations

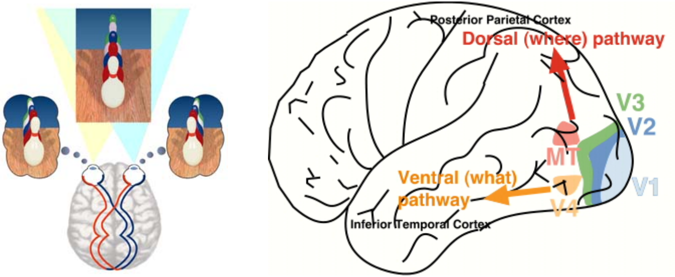
Convolutional Neural Networks (CNN)

- CNNs are more common in computer vision
- Each layer includes three steps:
  1. Convolutions (normal filtering operations)
  2. Non-linear operator (e.g. ReLU: set negative values to zero)
  3. Pooling (e.g. find local maximum and subsample)
- Last layers are fully connected.



Is the problem solved with deep learning then?

Visual cortex with *what* and *where* pathways.

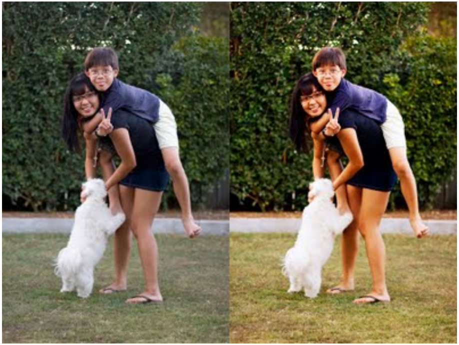


Deep learning can

- benefit from lots of data – but what if you don't have much data?
- answer *what*-questions – but not yet good at *where*-questions.

Computer vision is so much more than image classification.

Image processing $\Longleftrightarrow$ Signal processing



- The image is enhanced for easier interpretation.
- Different levels of processing (often used as pre-processing).

Fourier Transform

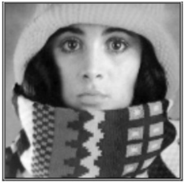


Figure 4a

Original

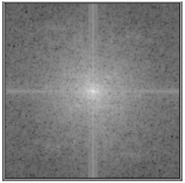


Figure 4b

$\log(|A(\Omega,\Psi)|)$

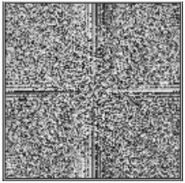


Figure 4c

$\varphi(\Omega,\Psi)$

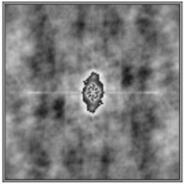


Figure 5a

$\varphi(\Omega,\Psi) = 0$

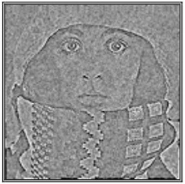
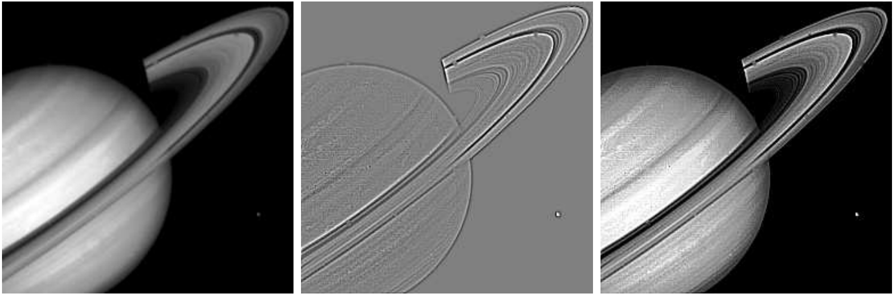


Figure 5b

$|A(\Omega,\Psi)| = \textit{constant}$

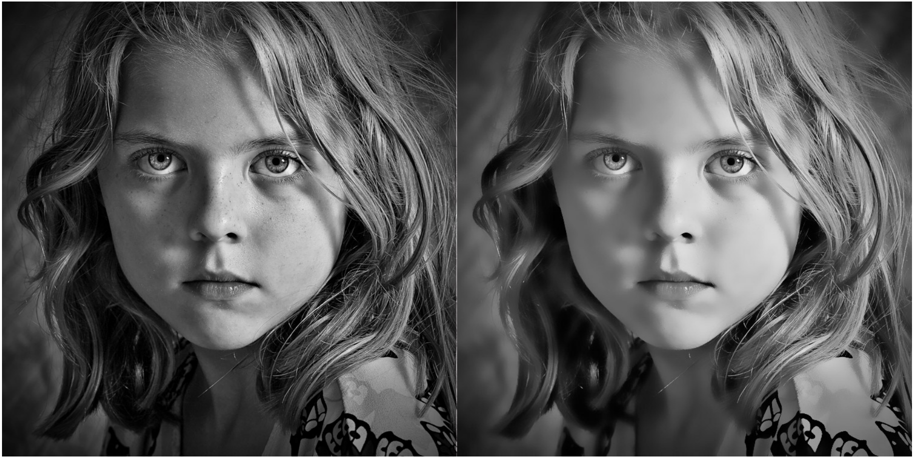
Images can be studied in frequency domain (like audio), but with images phase is more important than magnitude (unlike audio).

Application of the Laplacian operator



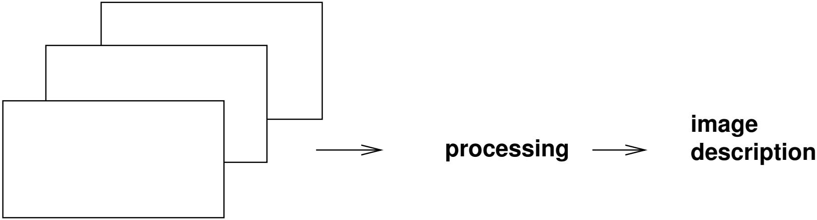
Original image (left), application of Laplacian operator (middle), and subtraction of the Laplacian from the original image (right).

Anisotropic smoothing



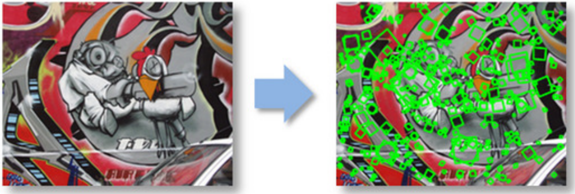
Note: the image is smoother, but individual hairs are not blurred out.

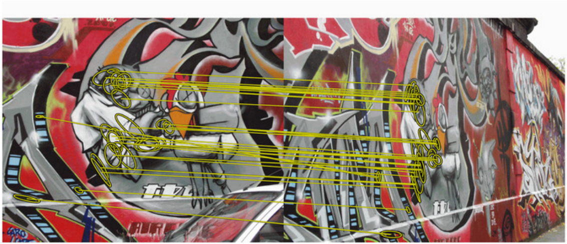
Image analysis



- Purpose: Generate a useful description of the image
- Examples: Character recognition, fingerprint analysis

Feature extraction and matching

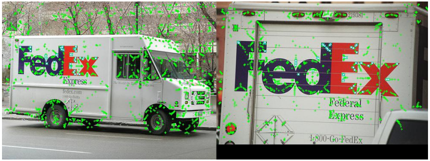




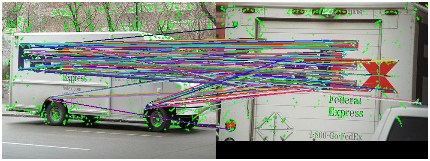
- Features (corners, blobs, lines, etc) can be extracted from images and then matched between images.

Matching planes with homographies

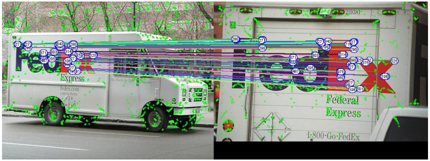
1. Detect point features in both image (e.g. SIFT or SURF)



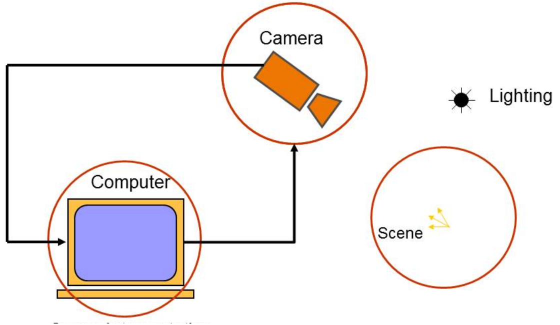
2. Match features between the two images.



3. Use RANSAC to find homography and mismatches (outliers)

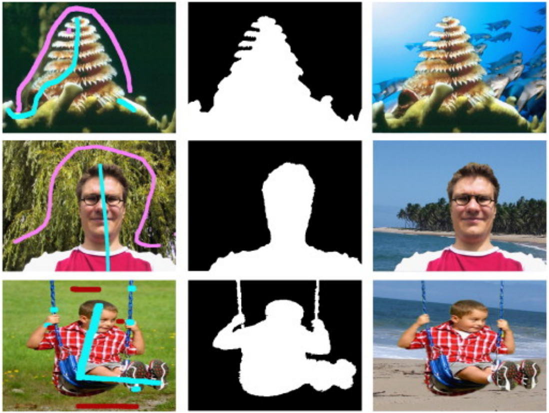


Computer vision



- Purpose: Achieve an understanding of the world, possibly under active control of the image acquisition process.
- Examples: object tracking, activity recognition
- Often people say computer vision, instead of image analysis.

Object segmentation using GrabCut



- Left: A couple of strokes are applied to create colour models of background and foreground.
- Right: Afterwards background can be changed to something else.

Computer vision examples

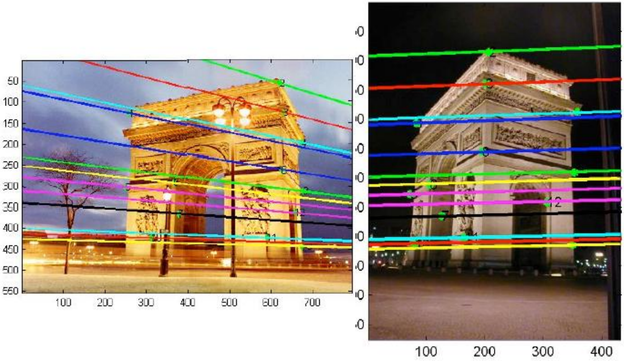
Figure: Scene parsing (Hong Kong)

Active contours example

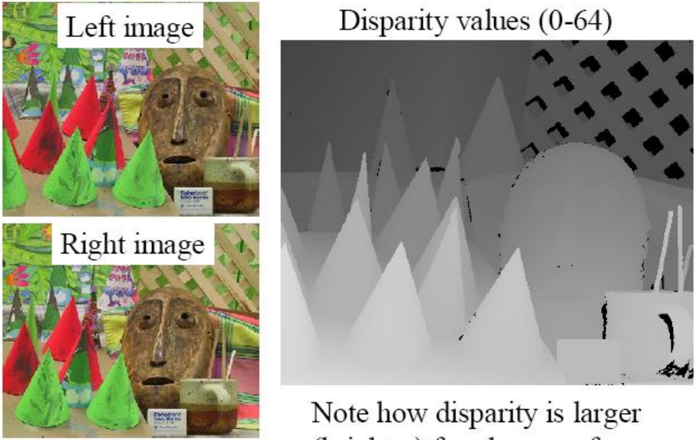
Figure: Liptracking using snakes

Epipolar geometry

- Matches between two images can be found along lines determined by the relative positions between cameras, or vice versa,
- if you have a number of matches between images, you can determine the relative positions between cameras.



Stereo matching

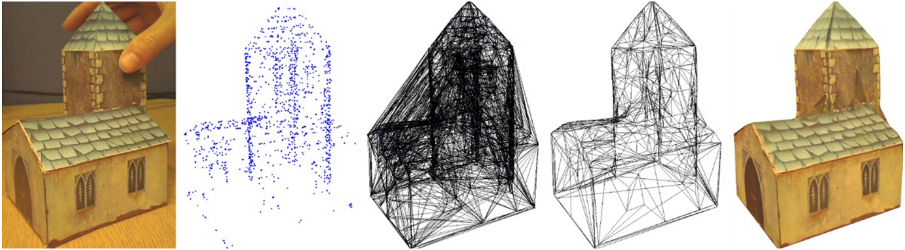


Disparity values (0-64)

Note how disparity is larger (brighter) for closer surfaces.

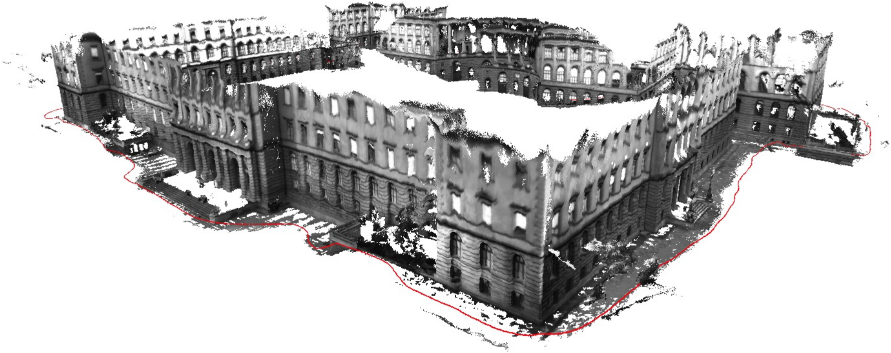
Match left and right images to create a depth image.

Multi-view 3D object reconstruction



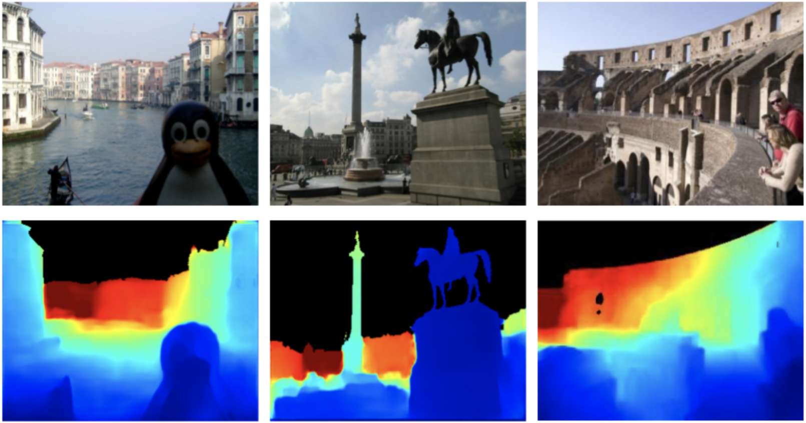
Find image features in multiple views, determine their positions in the 3D world and create surfaces.

Multi-view 3D object reconstruction



Such models can be very large and used to control a robot.

Single-view depth using deep learning



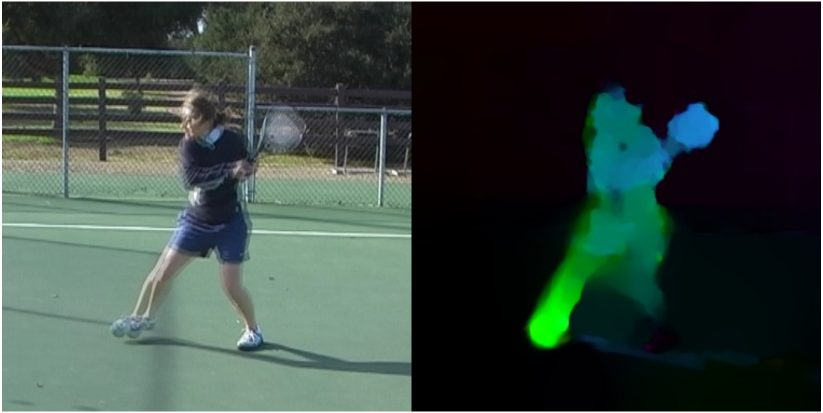
Grand Canal, Venice

Trafalgar Square, London

Colosseum, Rome

Using deep learning one can predict depths from a single view.

Motion analysis



Measure how pixels move over time.

Computer vision examples

Figure: OpenPose: Multi-person tracking (CMU)

Thank you!