

CS201 Lecture 02
Computer Vision:
Image Formation and Basic
Techniques

John Magee
31 August 2012

Computer Vision

How are Computer Graphics and Computer Vision Related?

Recall: Computer graphics in general

Description of scene → Visual representation (Image)

Computer Vision in general:

Image(s) → Some description of the scene



Example -

Input: Image

Output: Face locations

Fujifilm camera demo

Data Structures for Images

- 2D array vs. 1D array
- Interleaved RGB vs. Planar RGB
- Data stored in arrays vs. pointers to pixel class/structure.

Some Easy Techniques

- Color Analysis
- Motion Analysis
- Template matching

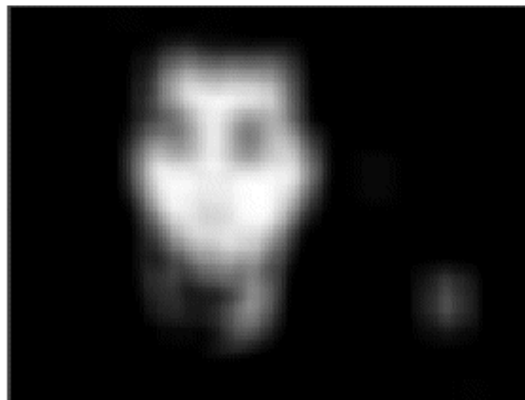
(Some extra detail on the next few slides)

Color Analysis

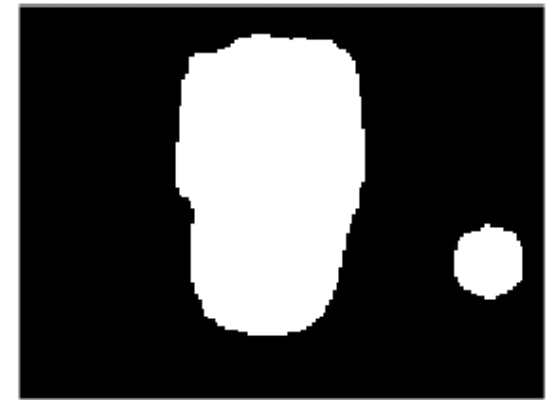
Skin color analyzed by lookup of 2D histogram:



(a) Result of histogram lookup.

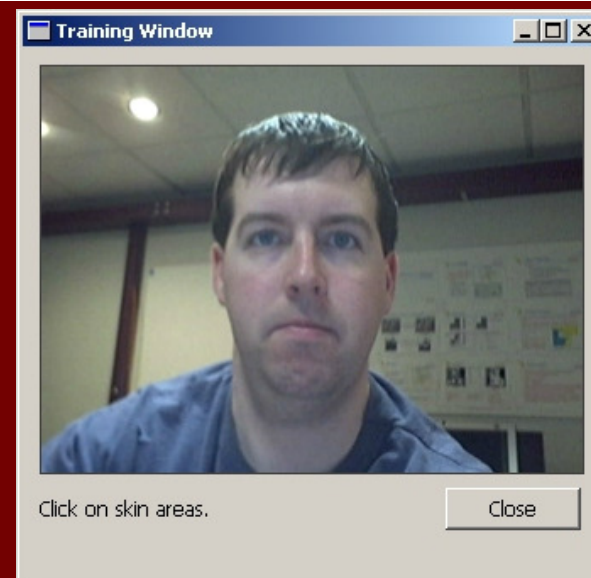


(b) Result after low-pass filter.



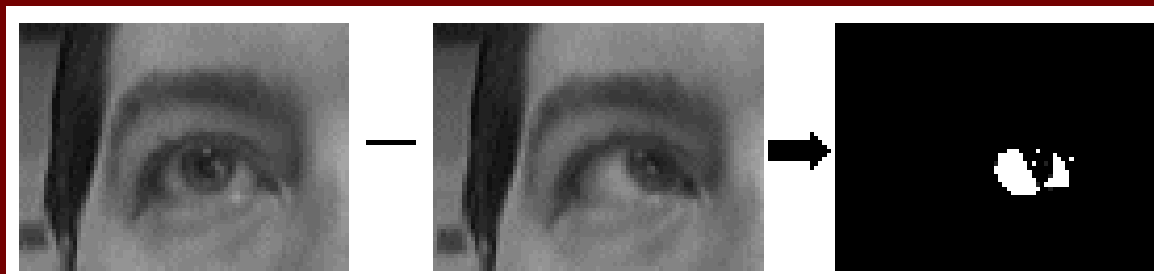
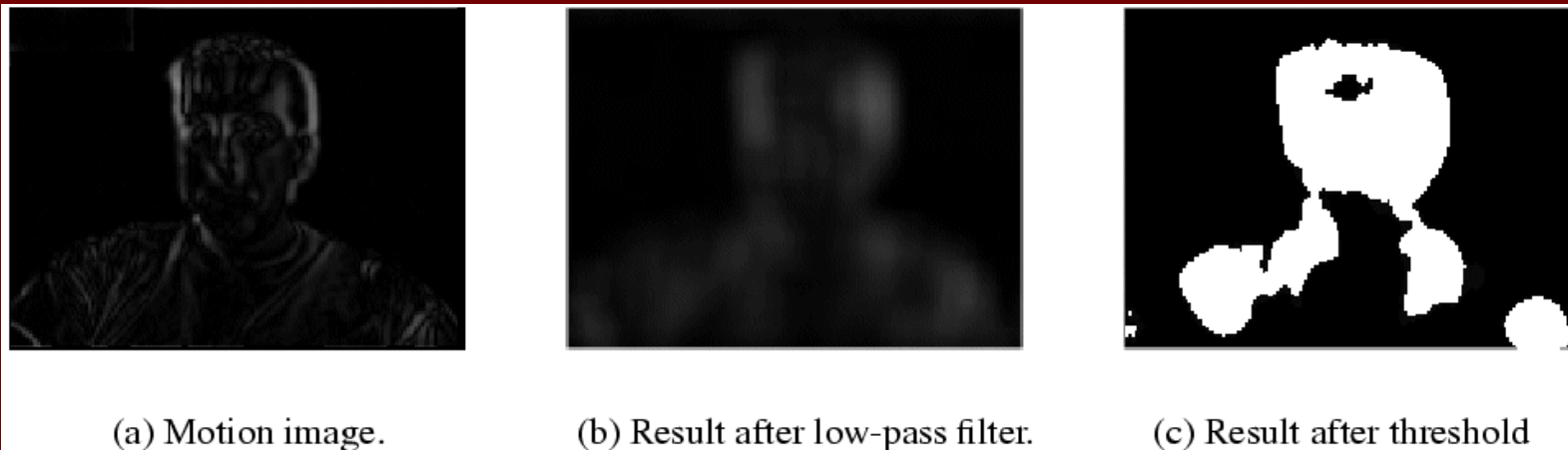
(c) Result after threshold

Histogram can be updated during operation →



Motion Analysis

Motion analysis by frame differencing:



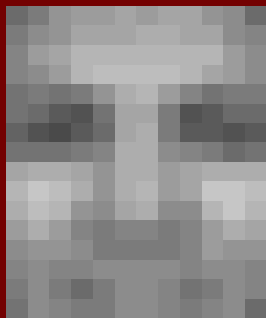
Recall: Video compression uses frame differencing.

Template Matching

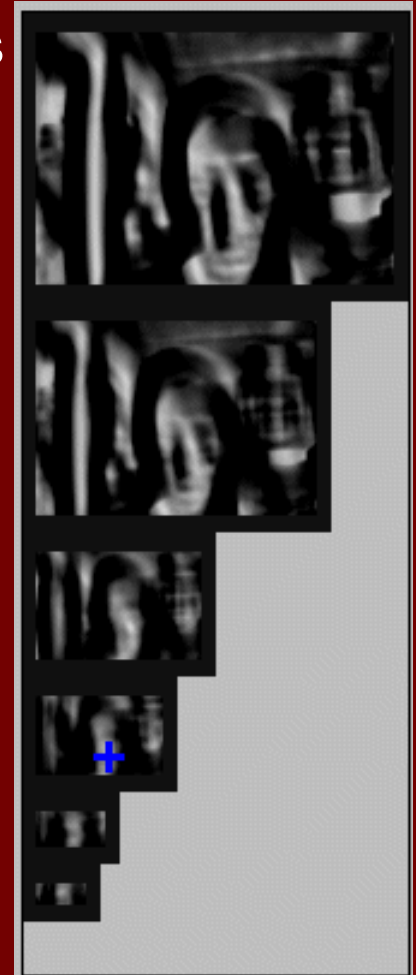
$$SAD(x, y) = \sum_{i=0}^{T_{rows}} \sum_{j=0}^{T_{cols}} \text{Diff}(x + i, y + j, i, j) \quad \text{Sum of Absolute Differences}$$

- Normalized correlation coefficient matching over multi-resolution search space.

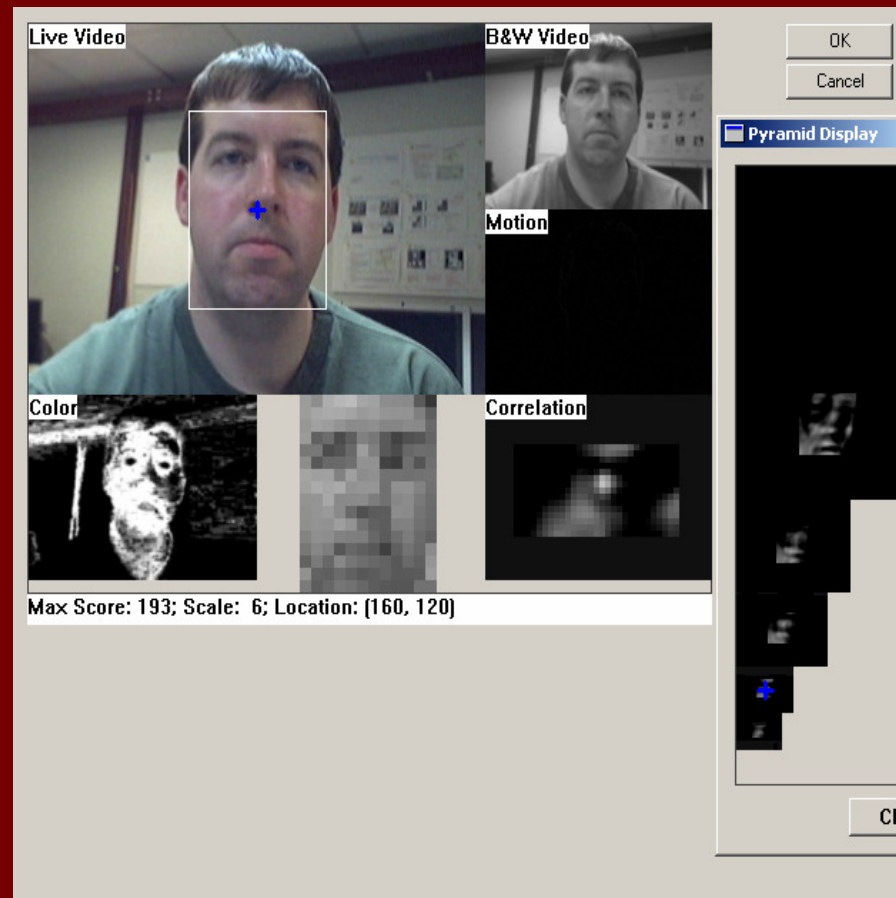
$$\tilde{R}(x, y) = \frac{\sum_{y'=0}^{h-1} \sum_{x'=0}^{w-1} (T(x', y') - \bar{T}) \cdot (I(x + x', y + y') - \overline{I(x, y)})}{\sqrt{\sum_{y'=0}^{h-1} \sum_{x'=0}^{w-1} (T(x', y') - \bar{T})^2 \sum_{y'=0}^{h-1} \sum_{x'=0}^{w-1} (I(x + x', y + y') - \overline{I(x, y)})^2}}$$



← 12 x 16 Template matching over all resolutions →



Face Tracking



Face Detection vs. Face Recognition

Face Detection exploits the **similarities** between human faces.

- Using Probabilistic/Statistical Matching

Face Recognition exploits the **differences** between human faces.

- Using Principle Component Analysis

Gaze Analysis



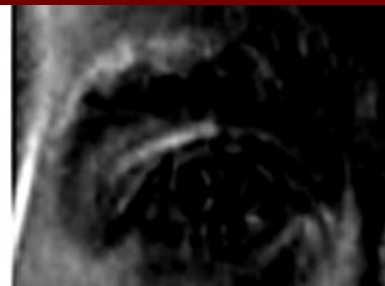
Right Eye



Mirrored
Left Eye



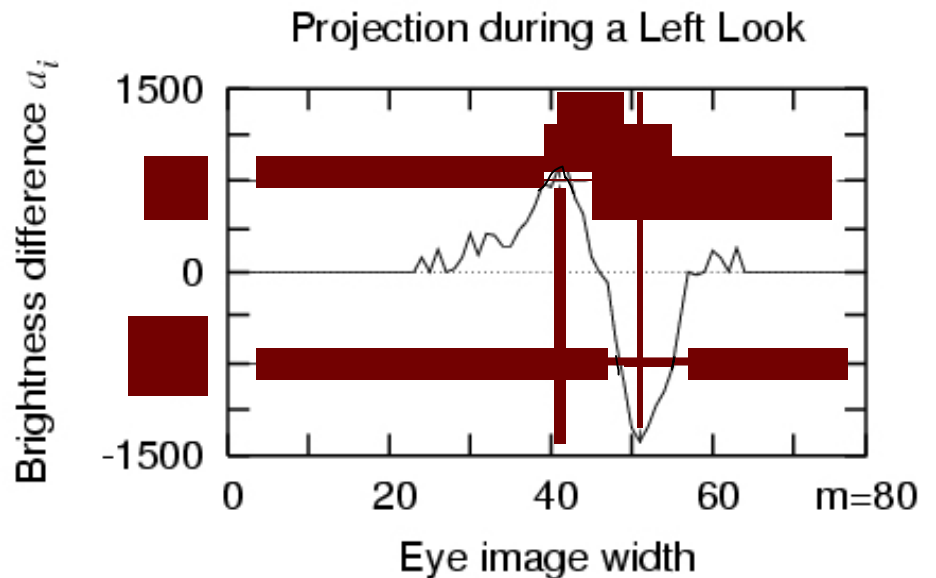
Looking
Left



Looking
Straight

Eye ($m \times n$) image
difference projected
to x-axis:

$$a_i = \sum_{j=1}^n (I_r(i, j) - I_\ell(m - i, j)).$$



Computer Vision

What can go wrong?

- You might not know anything about a scene!
- Lighting could change!
- People could do weird things!

Google Similar Images

<http://www.youtube.com/watch?v=6fD2t4d2Ln4>



Showing only similar images - [Back to results for paris](#)



360 x 480 - 43k - jpg
www.johnnyjet.com
[Similar images](#)



270 x 360 - 40k - jpg
graphics.worldweb.com
[Similar images](#)

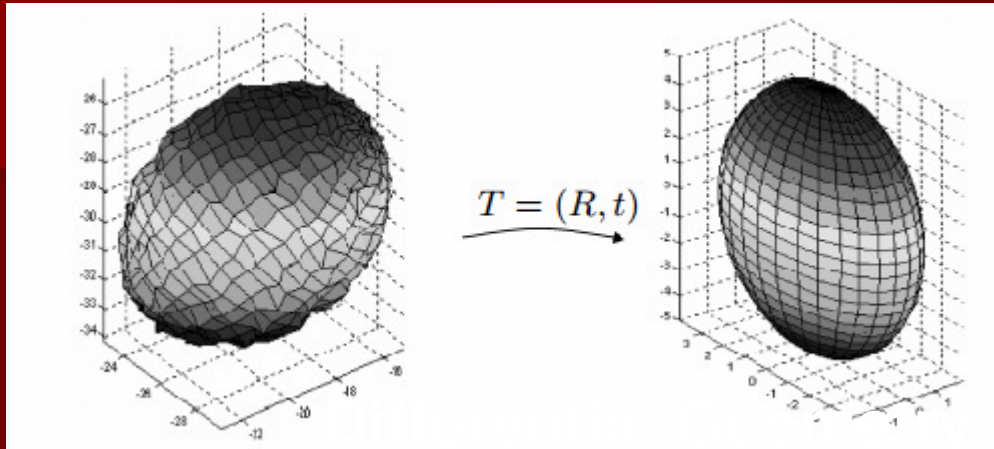


337 x 450 - 36k - jpg
tripadvisor.com
[Similar images](#)

<http://similar-images.googlelabs.com/>

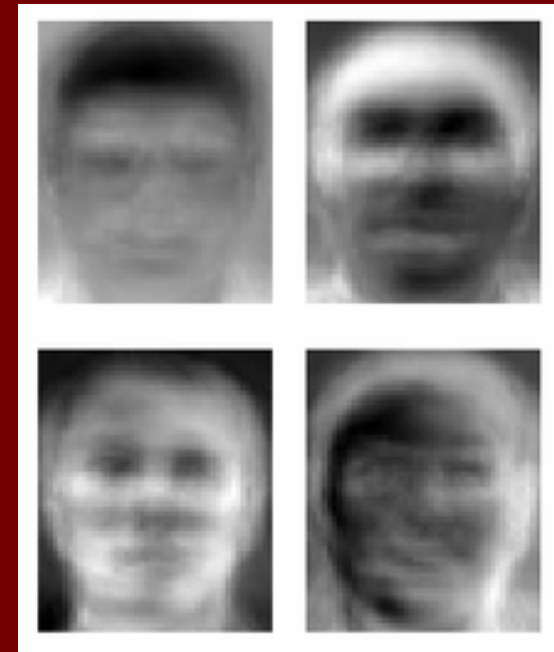
Systems that learn about the world.

Vision: Mathematical Foundations



- Probabilistic and Statistical Models
- Fourier Analysis

Extract high-level but low dimensional information from low-level high dimensional data.



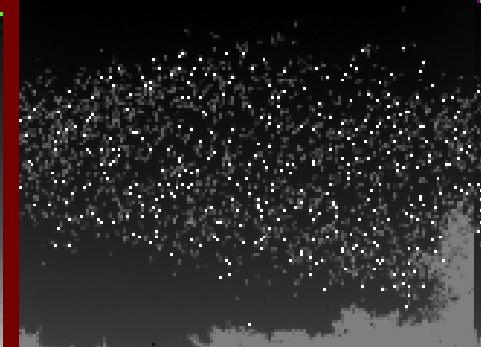
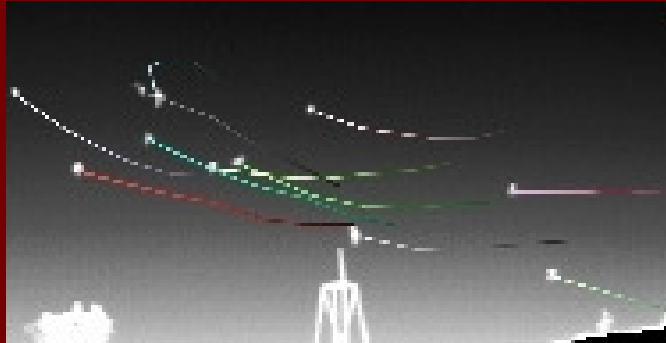
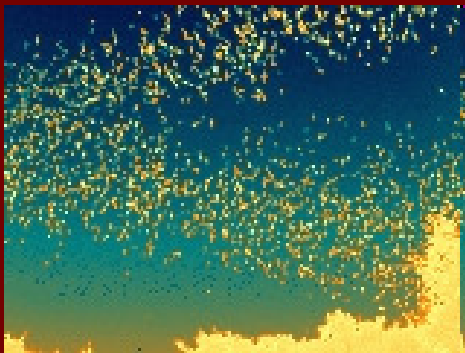
“Eigenfaces” – Principal Component Analysis

Animal Behavior and Census

Bat Tracking:

Collaboration with Biologists

Funded by Office of Naval Research



Demo Video

Cell Tracking / Analysis



House et al. – Boston U

Linguistic Analysis of Sign Language

The screenshot displays a software interface for linguistic analysis of sign language. It features three video windows at the top, each showing a different view of a signer: 'U3Video1 - database' (front view), 'U3Video3 - database' (close-up), and 'U3Video2 - database' (side view). Each video window includes a timestamp of 01/12/2000 18:24:57 and a frame number of 43985. Below the videos is a 'U3 Gloss - database' window. This window has a search bar with 'Datum' and a play button. It shows a participant named '*Norma' and various fields for linguistic analysis. The 'main gloss' field contains the text: **LOVE** fs-MARY |-----> fs-JOHN |----->. The 'english' field contains the text: 'As for loving Mary, John (does).'. The interface also includes buttons for 'Edit Participant', 'Show All Data', 'Show Field', 'Hide Field', and 'Hide Pane'.

Vision Guided Robots



Assistive Robots



Autonomous Vehicles

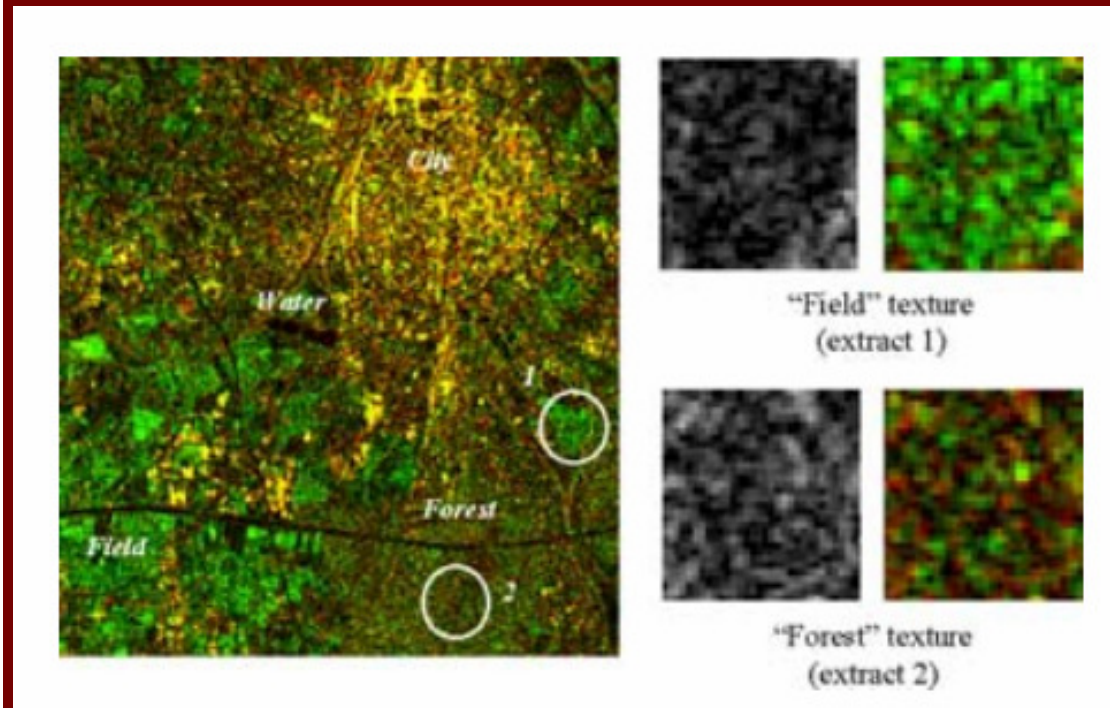
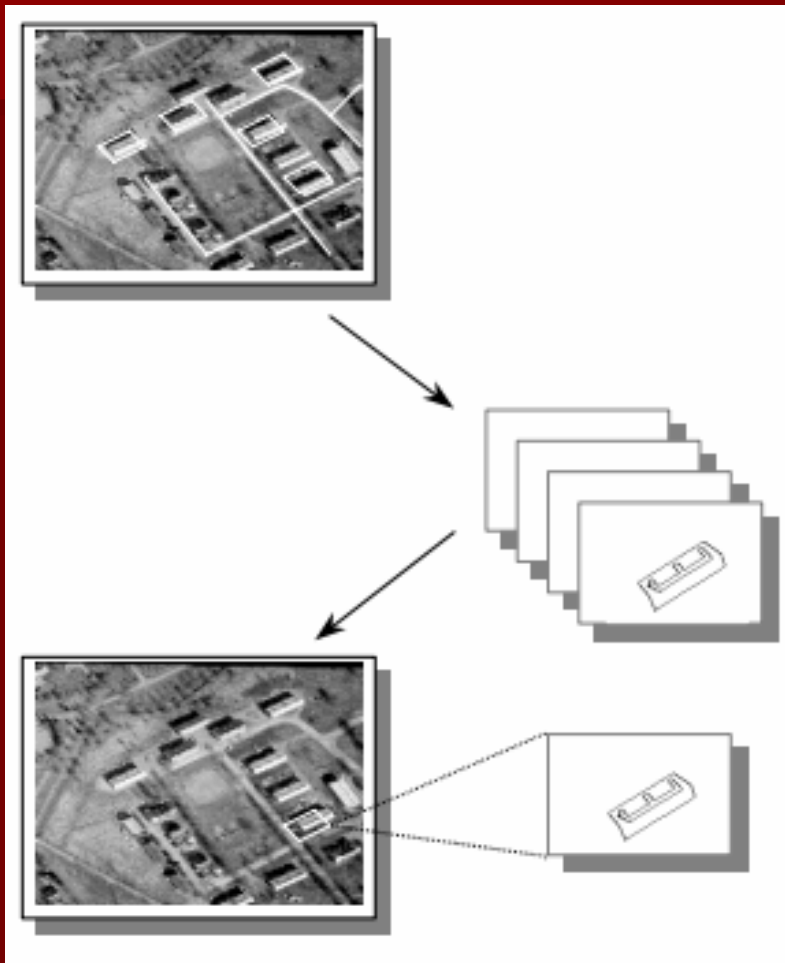


Tele-presence Robots



Manufacturing

Remote Sensing (Geography)



Computational Neuroscience

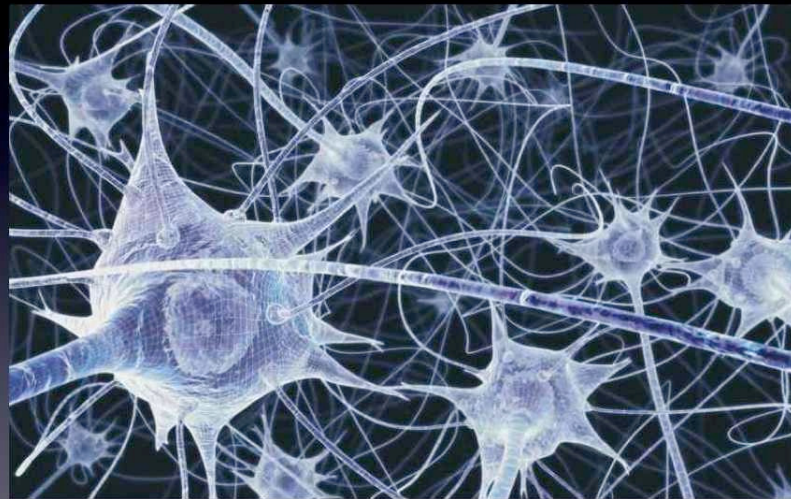
Biologically Inspired Vision:

Machine Learning, Artificial Neural Networks

Brain Modelling

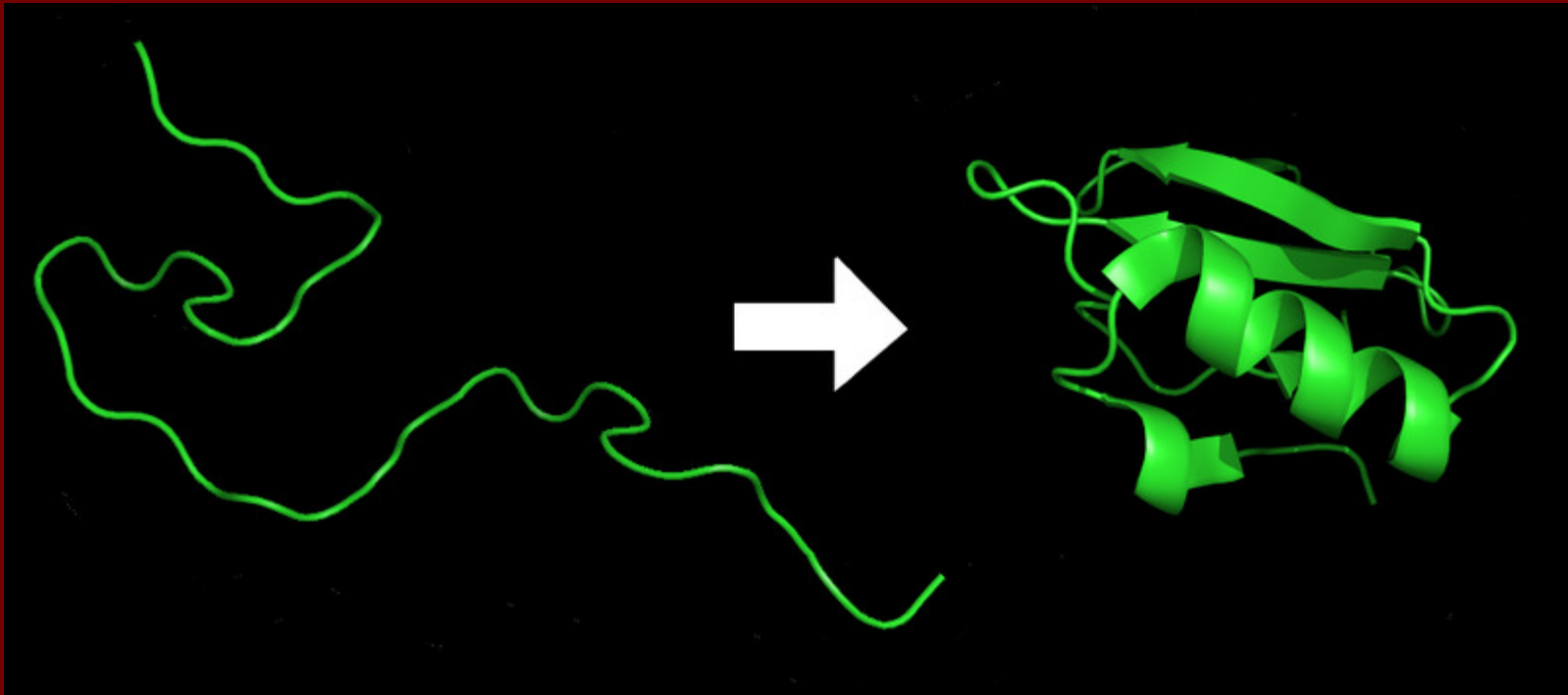
Brain-Computer Interfaces

Neural Networks: Artist Rendition



Protein Folding (Biochemistry)

Many Computer Vision techniques used in computer simulations.



Finance / Machine Learning

Abstract from Bloomberg research talk:

Gary Kazantsev, R&D Machine Learning, 12/05/2013

We will give a brief overview of the machine learning discipline from a practitioner's perspective and discuss the evolution and development of several key Bloomberg projects such as sentiment analysis, market impact prediction, novelty detection, machine translation, social media monitoring and information extraction. We will show that these interdisciplinary problems lie at the intersection of linguistics, finance, computer science and mathematics, requiring methods from signal processing, **machine vision** and other fields. Throughout, we will talk about practicalities of delivering machine learning solutions to problems of finance and highlight issues such as importance of appropriate problem decomposition, feature engineering and interpretability.

Human-Computer Interaction

We're all used to mouse and keyboard...
But you could use a camera to track motion...

Camera Mouse

<http://www.cameramouse.org/> (Free Download)



**A user with severe
paralysis using the
Camera Mouse**

Articles and Videos:

<http://www.bu.edu/today/2009/04/10/seeing-eye-mouse>

<http://www.bu.edu/today/2011/big-meaning-in-the-smallest-movements/>

Reading

- http://en.wikipedia.org/wiki/Template_matching
 - [http://en.wikipedia.org/wiki/Sum of absolute differences](http://en.wikipedia.org/wiki/Sum_of_absolute_differences)
 - <http://en.wikipedia.org/wiki/Cross-correlation>
- [http://en.wikipedia.org/wiki/Netpbm format](http://en.wikipedia.org/wiki/Netpbm_format)
- [http://en.wikipedia.org/wiki/Pinhole camera](http://en.wikipedia.org/wiki/Pinhole_camera)
- [http://en.wikipedia.org/wiki/Perspective projection](http://en.wikipedia.org/wiki/Perspective_projection)
- [http://en.wikipedia.org/wiki/Camera matrix](http://en.wikipedia.org/wiki/Camera_matrix)