Image Processing for Physicists

Prof. Pierre Thibault
pthibault@units.it

Overview

- Likelihood
- Bayes' theorem
- Application
  - ML Classification
  - Deconvolution
  - Image registration

What is likelihood?

- A likelihood function is a probability distribution expressed as a function of its parameters, and evaluated for a given set of observations.

probability of x given α

$$p(x|\alpha) = \ell(\alpha|x)$$

$\ell(\alpha|x)$ is not the probability that the model is true

Maximum likelihood

Can easily be misunderstood...

Shroud of Turin

p(shroud has this appearance | it really was Jesus) very high ~100%

l(it really was Jesus | it look like this) missing prior

Bayes' theorem

A: I had a good grade in a course

B: I studied a lot

$$p(A \cap B) = p(A|B)\,p(B)$$
$$= p(B|A)\,p(A)$$

"prior"

"posterior"

$$p(B|A) = \frac{p(A|B)\,p(B)}{p(A)} = \frac{p(A|B)\,\overbrace{p(B)}^{\text{"prior"}}}{\int_B p(A|B)\,p(B)}$$

$l(B|A)$

$$p(\alpha|x) = \frac{p(x|\alpha)\,p(\alpha)}{(p(x))} \propto l(\alpha|x)\,p(\alpha)$$

irrelevant

Maximum likelihood & optimization

- Goal: find the parameters that explain best the observed data.

  → Maximum likelihood   maximize $\ell(\alpha|x)$

  or

  → Maximum a posteriori (MAP)

  maximize $\ell(\alpha|x)\underbrace{p(\alpha)}$

  additional knowledge about α

- Very often more convenient to minimize -log().

Example: a biased coin

$p(\text{result is Head} \mid \alpha) = \alpha$ $\qquad$ α: parameter and probability of head

$p(\text{result is Tail} \mid \alpha) = 1-\alpha$

problem: find α given some measurements

N_H : number of Heads in a measurement

N_T : " " Tails

$$p(N_H, N_T \mid \alpha) = \alpha^{N_H} (1-\alpha)^{N_T} = \ell(\alpha \mid N_H, N_T)$$

$$\mathcal{L} = -\ln(\ell) = -N_H \ln\alpha - N_T \ln(1-\alpha)$$

find $\min_{\alpha} \mathcal{L}$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = -\frac{N_H}{\alpha} + \frac{N_T}{1-\alpha} = 0 \qquad (1-\alpha) N_H = \alpha N_T$$

$$\alpha(N_T + N_H) = N_H \qquad \hat{\alpha} = \frac{N_H}{N_H + N_T}$$

Example: Gaussian model

1. A single variable: $p(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\alpha^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

2. Many independent variables with same distribution (many independent measurements drawn from same distribution)

$$p(x_1, x_2, x_3, \ldots, x_N \mid \mu, \sigma^2)$$

$$= \prod_i p(x_i \mid \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\sum_i \frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

$$= \ell(\mu, \sigma^2 \mid x_1, x_2, x_3 \ldots)$$

$$\mathcal{L} = -\ln(\ell) = \frac{N}{2}\ln(2\pi\sigma^2) + \frac{1}{\sigma^2}\sum_i (x_i-\mu)^2$$

least squares

$$\frac{\partial \mathcal{L}}{\partial \mu} = 0 \longrightarrow \mu = \frac{1}{N}\sum_i x_i \qquad \frac{\partial \mathcal{L}}{\partial \sigma^2} = 0 \Rightarrow \sigma^2 = \frac{1}{N}\sum_i (x_i-\mu)^2$$

Example: Gaussian model

3. N variables <u>not</u> identically distributed and <u>not</u> independent

$$p(\vec{x} \mid \vec{\mu}, C) = \frac{1}{(2\pi)^{N/2}\sqrt{|C|}} \exp\left(-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right)$$

$\vec{\mu}$: means

C: covariance matrix

$|C|$: determinant

If M measurements are made:

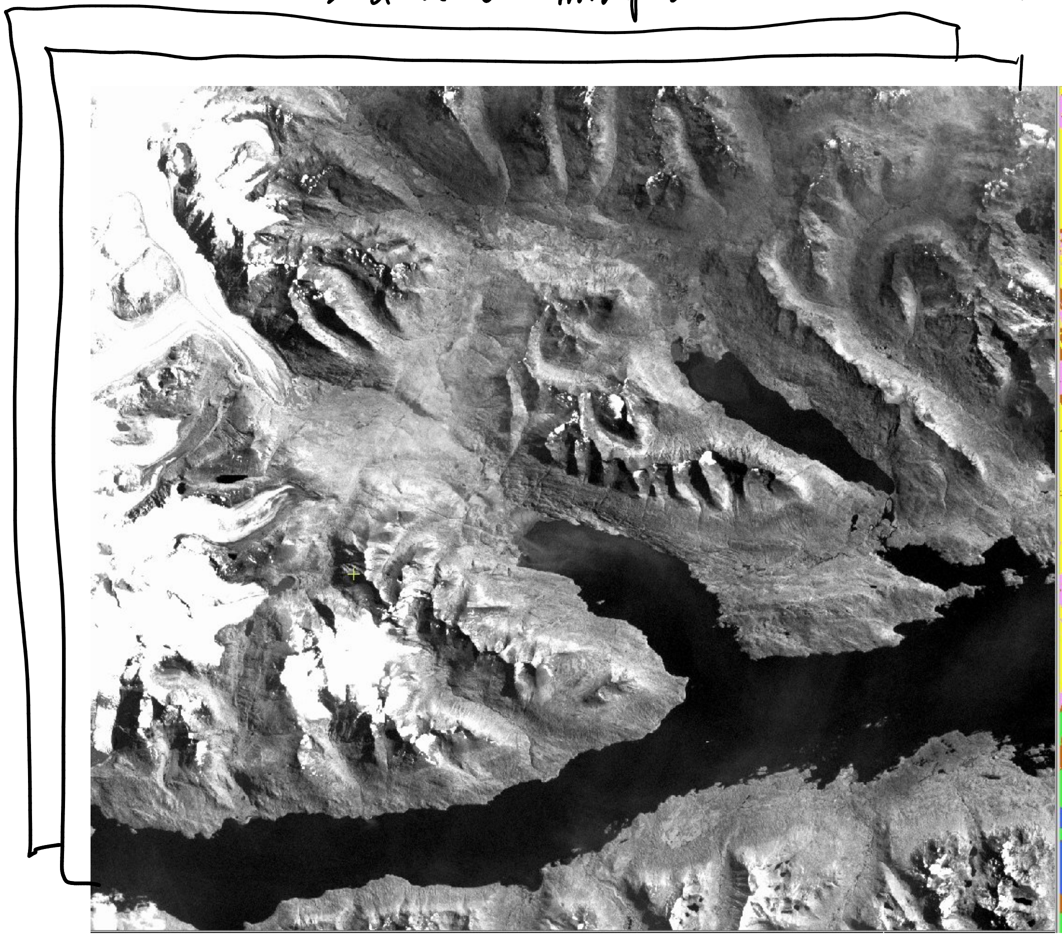
$$p(\vec{x}^{(1)}, \vec{x}^{(2)}, \vec{x}^{(3)}, \dots \mid \vec{\mu}, C) = \frac{1}{(2\pi)^{\frac{MN}{2}} |C|^{M/2}} \exp\left(-\frac{1}{2}\sum_i (\vec{x}^{(i)} - \vec{\mu})^T C^{-1} (\vec{x}^{(i)} - \mu)\right)$$

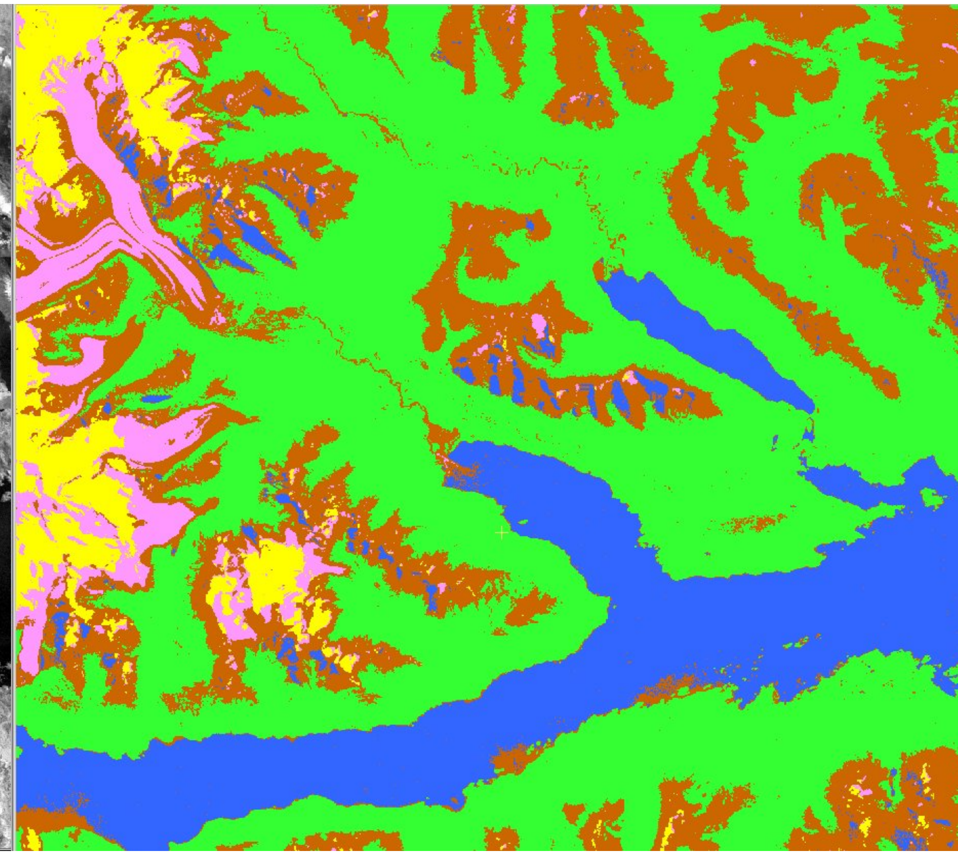
Result:

$$\vec{\mu} = \frac{\sum_i \vec{x}^{(i)}}{M} \qquad C_{lm} = \frac{1}{M}\sum_i (x_l^{(i)} - \mu_l)(x_m^{(i)} - \mu_m)$$

Image classification

stack of images. here sattelite images at various wavelengths





Goal: assign each pixel to a class according to a probability model

Image classification

Landsat 8-9 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS)

Bands	Wavelength (micrometers)	Resolution (meters)
Band 1 - Coastal aerosol	0.43-0.45	30
Band 2 - Blue	0.45-0.51	30
Band 3 - Green	0.53-0.59	30
Band 4 - Red	0.64-0.67	30
Band 5 - Near Infrared (NIR)	0.85-0.88	30
Band 6 - SWIR 1	1.57-1.65	30
Band 7 - SWIR 2	2.11-2.29	30
Band 8 - Panchromatic	0.50-0.68	15
Band 9 - Cirrus	1.36-1.38	30
Band 10 - Thermal Infrared (TIRS) 1	10.6-11.19	100
Band 11 - Thermal Infrared (TIRS) 2	11.50-12.51	100

Image classification

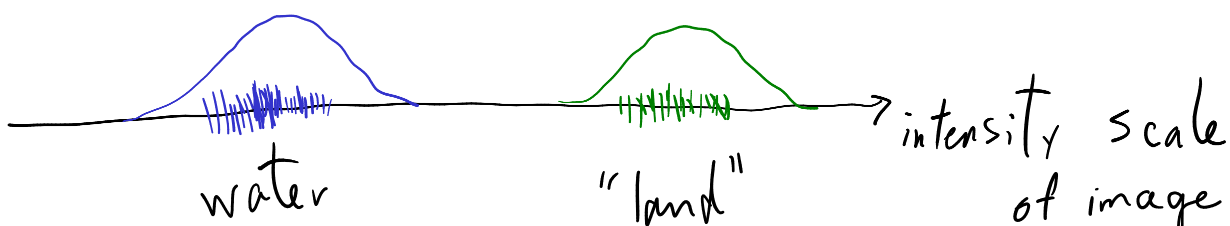
Supervised Maximum Likelihood Classification

1. Training: for each class, evaluate the probability distribution of the measurements.

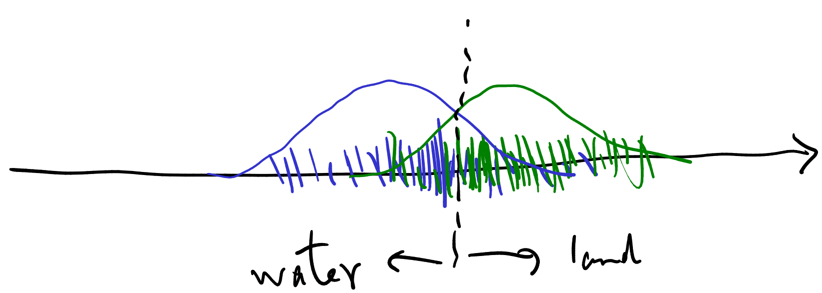
→ extract $\vec{\mu}$, C

1D

one of the images



in reality:



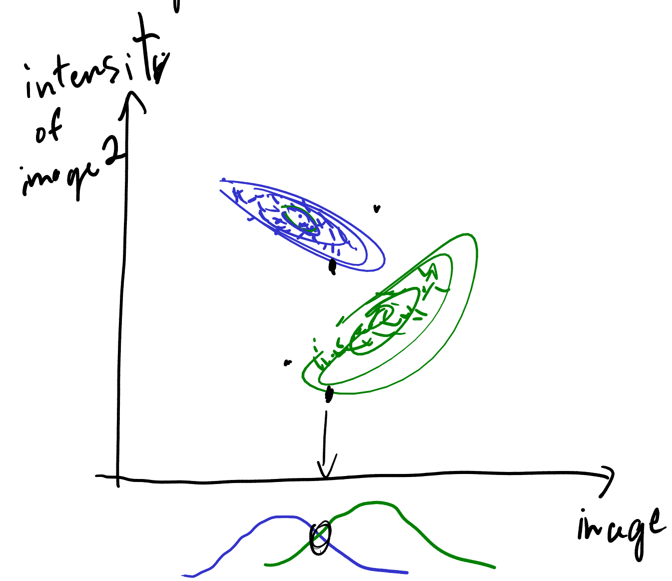


Image classification

Supervised Maximum Likelihood Classification

2. Classification: for each pixel, compute the probability that it belongs to each class. The highest probability wins.

$$p_{water}(\vec{x} \mid \vec{\mu}_{water}, C_{water}) = \ldots \left(\frac{1}{(2\pi)^{N/2}\sqrt{|C|}} \exp\left(\underbrace{-\frac{1}{2}(x-\mu)^T C^{-1} (x-\mu)}_{\text{as computed earlier}} \right) \right)$$

p_{field}

p_{city}

$\vdots$

sometimes called

Mahalanobis distance

$$f = -\ln(p_c) = \frac{1}{2}\ln|C| + \frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)$$

← minimize this

⇕

maximize p

Image deconvolution revisited

Image convolved with -known- PSF in the presence of noise

$$g(\vec{r}) = \underset{\substack{\uparrow \\ \text{PSF}}}{h} * \underset{\substack{\uparrow \\ \text{true image}}}{f}(\vec{r}) + \underset{\substack{\uparrow \\ \text{noise}}}{n}(\vec{r})$$

Fourier space

$$G(\vec{u}) = H(\vec{u}) F(\vec{u}) + N(\vec{u})$$

Often good assumption: $N(\vec{u})$ is uncorrelated $\leftrightarrow$ white noise

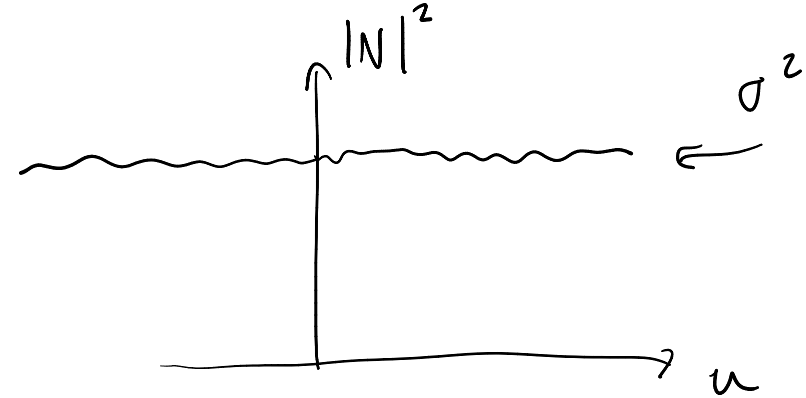


Image deconvolution revisited

Probability of measuring $G(\vec{u})$:

$$\ell(F|G) = p\left(G(\vec{u}) \,|\, F(\vec{u})\right) \propto \exp\left(-\frac{1}{2} \sum_u \frac{1}{|N(u)|^2} \left|F(\vec{u}) H(\vec{u}) - G(\vec{u})\right|^2\right)$$

$$-\ln(\ell) = \mathcal{L}(F) = \sum_u \frac{1}{|N(u)|^2} \left| F(\vec{u}) H(\vec{u}) - G(\vec{u}) \right|^2$$

$$\frac{\partial \mathcal{L}}{\partial F(\vec{u})^*} = 0 \implies F = G/H$$

not good! same instability as discussed before

Solution: include prior: impose power spectrum on F

$$p\left(F(\vec{u})\right) \propto \exp\left(-\frac{1}{2} \sum \frac{|F(u)|^2}{S(u)}\right)$$

Image deconvolution revisited

Maximum a posteriori (MAP)

maximize $l(F|G)\,p(F)$ instead of $l(F|G)$

$$\mathcal{L}'(F) = -\ln\big(l(F|G)\,p(F)\big) = \underbrace{-\ln(l)}_{\mathcal{L}} - \ln(p(F))$$

$$= \sum_u \frac{1}{|N(u)|^2}\,|F(u)H(u) - G(u)|^2 + \sum_u \frac{|F(u)|^2}{S(u)}$$

$$\frac{\partial \mathcal{L}}{\partial F^*} = 0 \longrightarrow \frac{1}{|N(u)|^2}\big(F(u)H(u) - G(u)\big)H^*(u) + \frac{F(u)}{S(u)} = 0$$

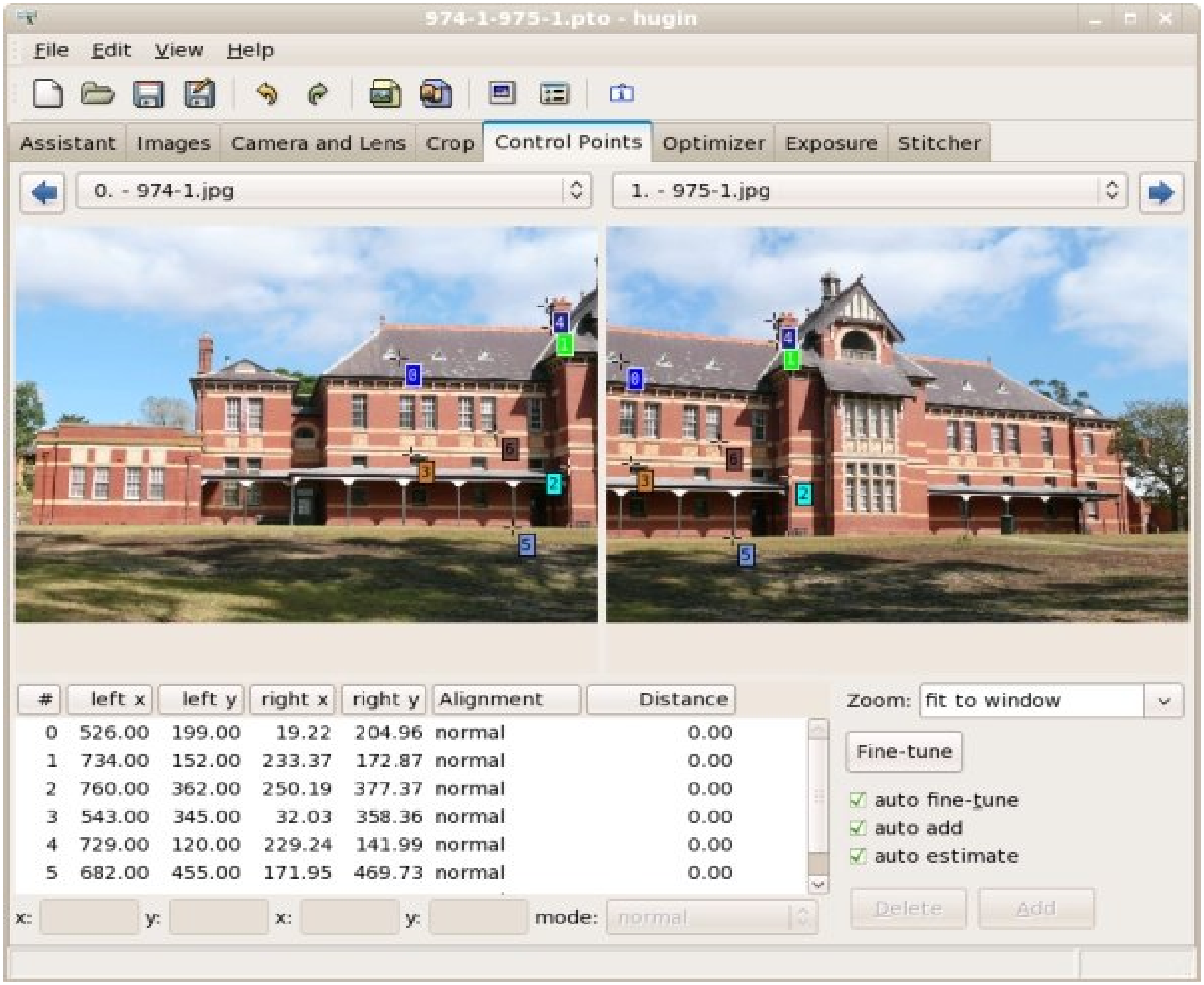
$$F(u) = \frac{H^*(u)\,G(u)}{\left(|H|^2 + \frac{|N|^2}{S}\right)}$$

<u>Wiener filter</u>

What is image registration?

- Geometric transformation of multiple images to make them match
- Transformations can be rigid or non-rigid
  - Rigid: translation, scale, rotation
  - Non-rigid: shear, perspective, ...
- Optimization can be done on the transformed images or on a set of control points.
- In almost all cases, interpolation is required to remap images on a regular grid.

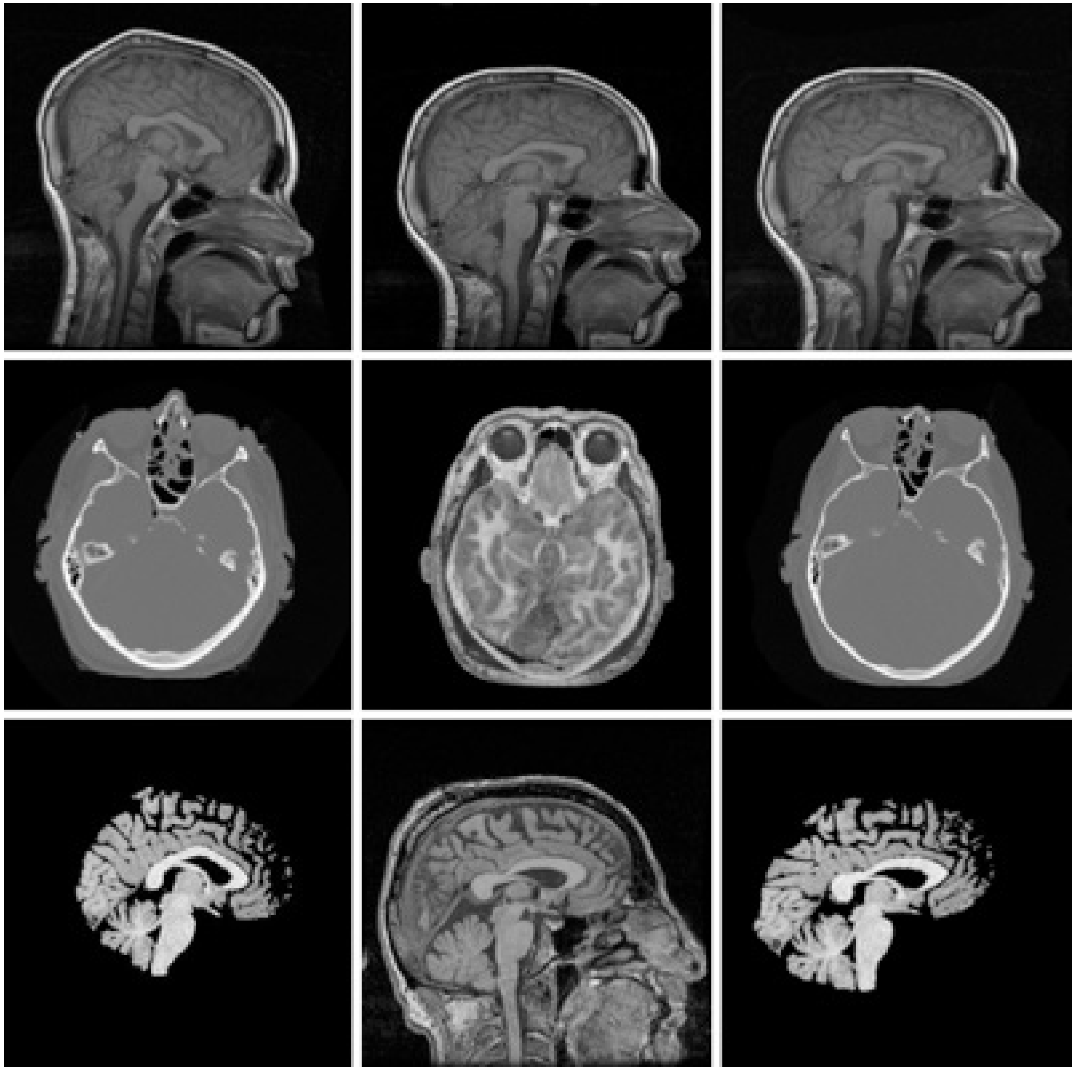
Control points for photo stitching



Source: http://hugin.sourceforge.net/tutorials/two-photos/en.shtml

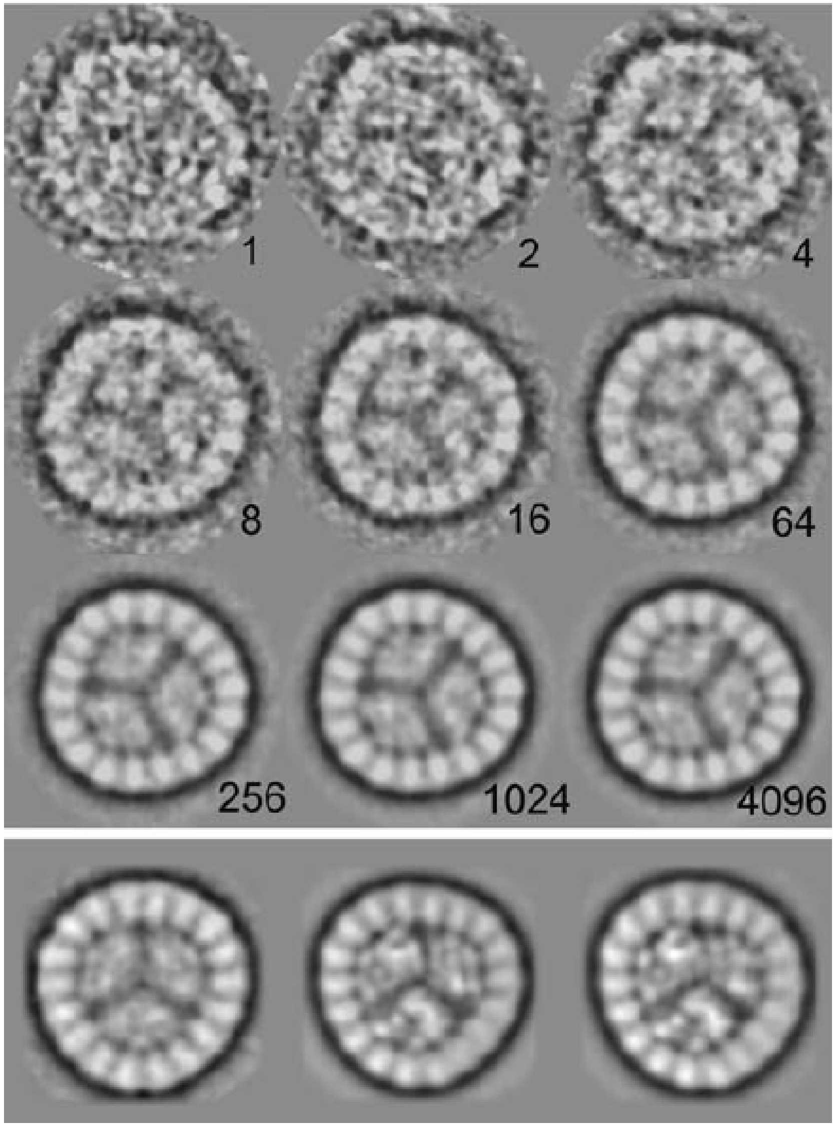
Image registration

Medical image registration



Source: http://www.cs.dartmouth.edu/farid/Hany_Farid/

Single particle analysis

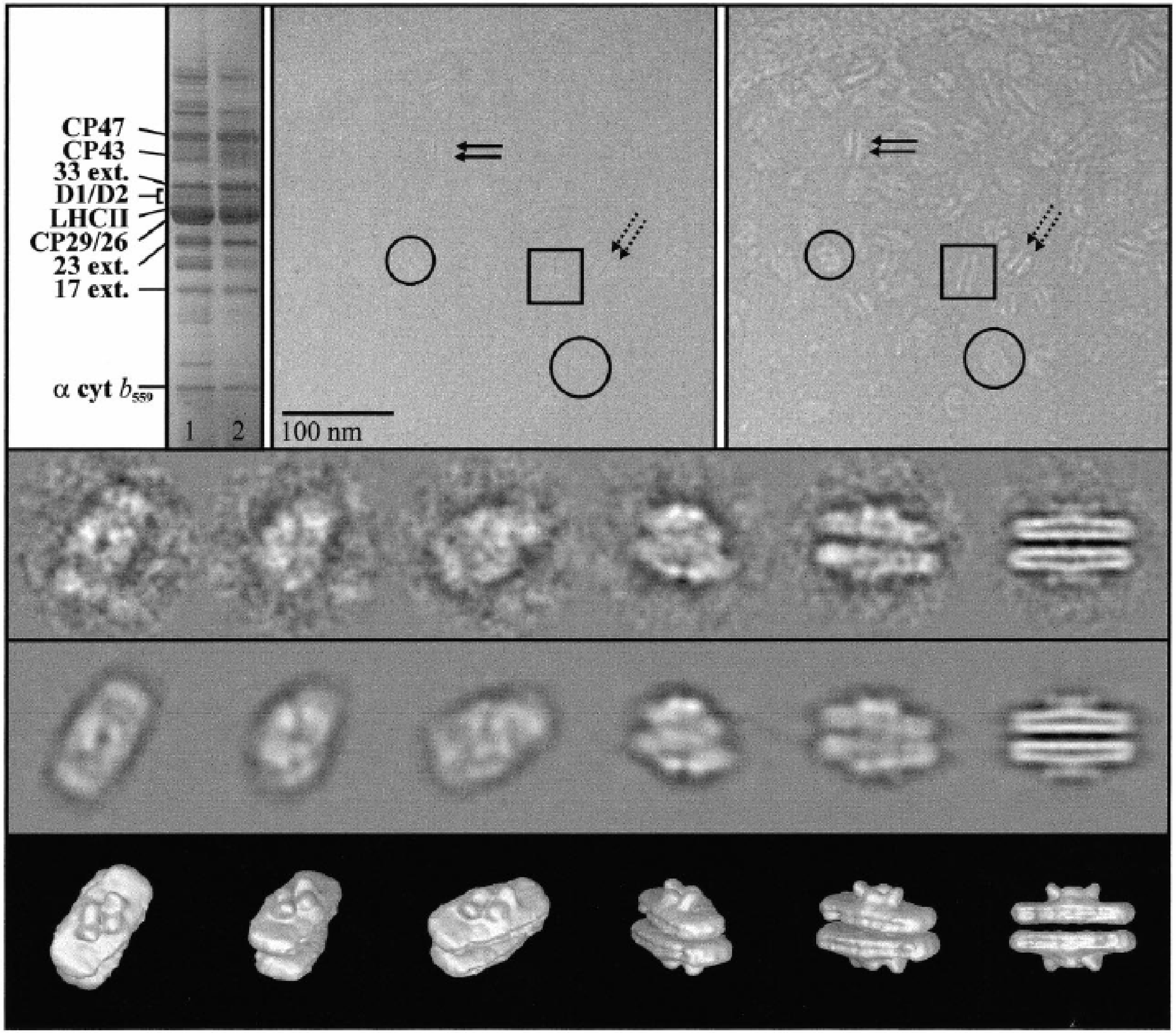


Cryo EM
electron microscopy

Nobel 2017

Source: Boerkema *et al.* Photosynth. Res. **102**, 189-196 (2009)

Single particle analysis



Source: Nield *et al.* Nat. Struct. Bio. **7**, 44-47 (2000)

Summary

- Likelihood maximization: finding parameters that best fit an observation.
    - Powerful, but:
    - Can overfit, can misinterpret
- Maximum A Posteriori (MAP): include prior (probabilistic) knowledge
- Broad range of applications:
    - Classification, registration, enhancements, ...