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

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

known

known

It is not the probability that a 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 looks like this $)$

missing: Prior

Bayes' theorem

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

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

$\ell(B|A)$ likelihood

$$p(B|A) = \frac{p(A|B)\,p(B)}{p(A)}$$

"prior"

posterior probability

irrelevant most of the tim

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

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

Maximum likelihood & optimization

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

  → Maximum likelihood

  maximizing $\ell(\alpha | x)$

  or

  → Maximum a posteriori (MAP)

  maximizing $p(\alpha|x) \propto \ell(\alpha|x)p(\alpha)$

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

  (ok because log is a monotonic function)

Example: a biased coin

$p(r = H \mid \alpha) = \alpha$

α: parameter and probability of head

$p(r = T \mid \alpha) = 1 - \alpha$

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

probability of obtaining N_H heads and N_T tails

$$= \ell(\alpha \mid N_H, N_T)$$

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

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

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

Example: Gaussian model

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

2. many independent variables (independent measurements)

$$p(x_1, x_2, x_3, \dots, x_N \mid \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\sum_i \frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

$$\left(= \prod_i p(x_i \mid \mu, \sigma^2)\right)$$

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

can check: $\frac{\partial \mathcal{L}}{\partial \mu} = 0 \longrightarrow \mu = \frac{1}{N}\sum_i x_i$

$\frac{\partial \mathcal{L}}{\partial \sigma^2} \longrightarrow \sigma^2 = \frac{1}{N}\sum_i (x_i - \langle x \rangle)^2$

Example: Gaussian model

3. N variables not identically distributed and not independent

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

$(x_1, x_2, \ldots)$

$(\mu_1, \mu_2, \ldots)$

covariance matrix

determinant

"Mahalanobis distance"

If M measurements are made:

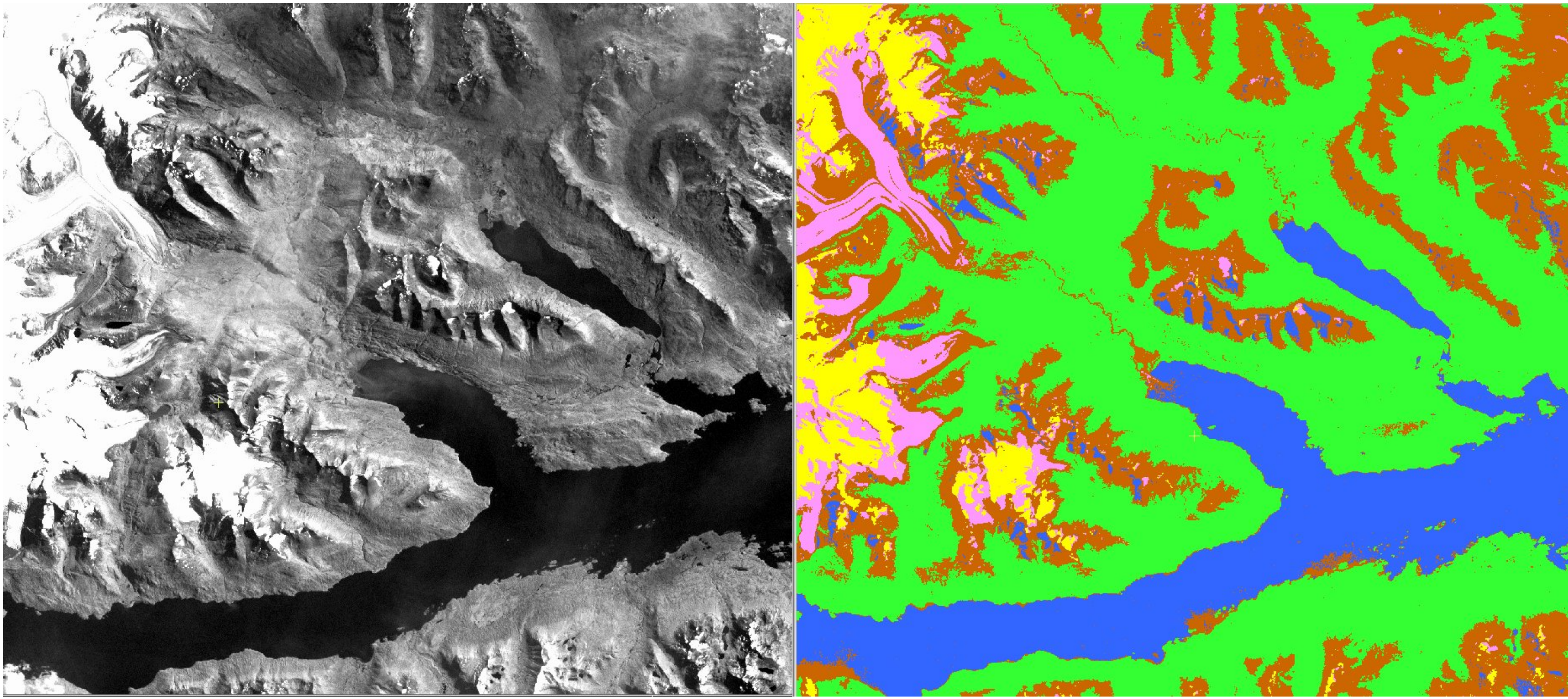
$$p(\vec{x}^{(1)}, \vec{x}^{(2)}, \vec{x}^{(3)}, \ldots \vec{x}^{(M)} \mid \vec{\mu}, C) = \frac{1}{(2\pi)^{\frac{MN}{2}} |C|^{\frac{M}{2}}} \times$$

$$\exp\Big(-\tfrac{1}{2}\sum_i (\vec{x}^{(i)} - \vec{\mu})^T C^{-1} (\vec{x}^{(i)} - \mu)\Big)$$

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

6 others !



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)

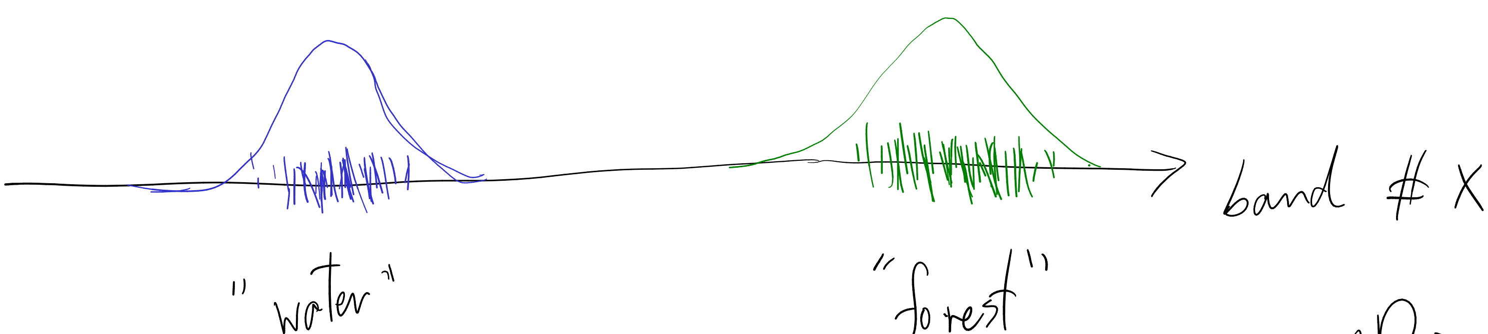
Bands	Wavelength (micrometers)	Resolution (meters)
Band 1 - Coastal aerosol	0.43-0.45	30
Band 2 - Blue	0.45-0.51	30
Band 3 - Green	0.53-0.59	30
Band 4 - Red	0.64-0.67	30
Band 5 - Near Infrared (NIR)	0.85-0.88	30
Band 6 - SWIR 1	1.57-1.65	30
Band 7 - SWIR 2	2.11-2.29	30
Band 8 - Panchromatic	0.50-0.68	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

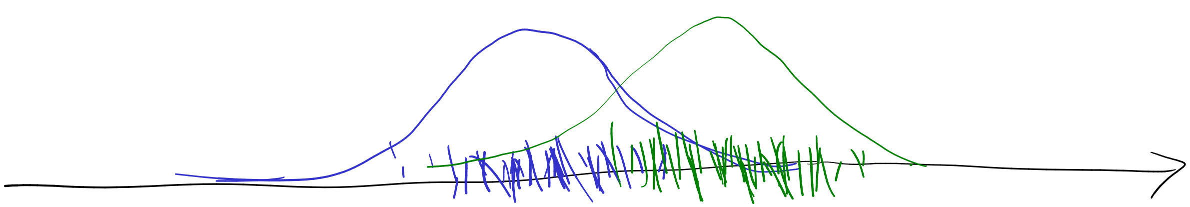
Supervised Maximum Likelihood Classification

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

1D example



... actually



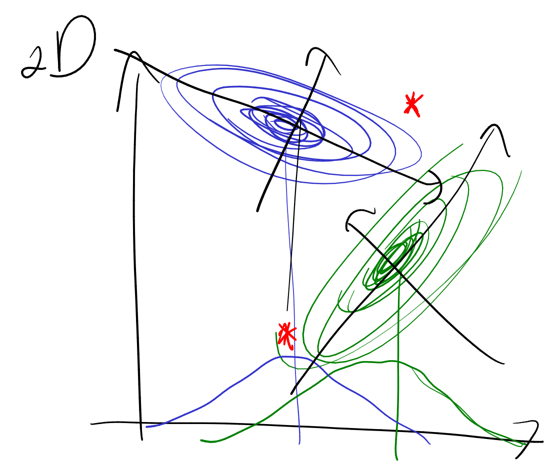


Image classification

Supervised Maximum Likelihood Classification

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

$\vec{\mu}_{water}$ $\vec{\mu}_{field}$ ...

$p(\vec{x} \mid \vec{\mu}, C)$

one observation (one pixel)

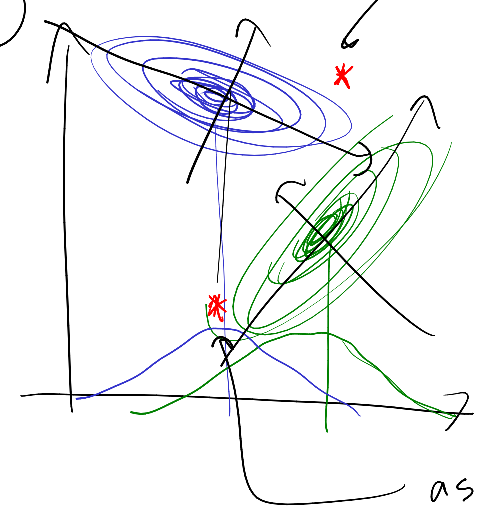
p_{water}

p_{field}

$p_{...}$

2D

assigned to "water" class



assigned to "forest" class

Image deconvolution revisited

In the presence of noise:

PSF (known)

$$g(\vec{r}) = (h * f)(\vec{r}) + n(\vec{r})$$

additive gaussian noise

$p(n_1, n_2, n_3, \ldots)$

Noise correlation: e.g. 2 pixels

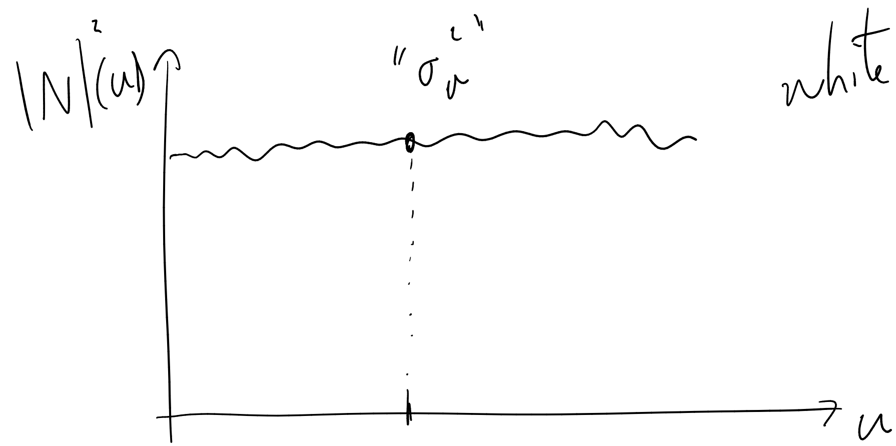
$$p(n_1, n_2) \propto \exp\left(-\frac{1}{2}[n_1, n_2]\, C^{-1} \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}\right)$$

not diagonal if noise is correlated!

Fourier basis: uncorrelated is a good assumption

$$C = \begin{bmatrix} \langle n_1^2 \rangle & \langle n_1 n_2 \rangle \\ \langle n_1 n_2 \rangle & \langle n_2^2 \rangle \end{bmatrix}$$

$$n(\vec{r}) \longrightarrow N(\vec{u})$$



white noise

Image deconvolution revisited

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

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

$$\frac{\partial \mathcal{L}}{\partial F^*(u)} = \frac{1}{|N(u)|^2}\left(F(u)H(u) - G(u)\right)H^*(u) = 0$$

$$F(u)H(u) = G(u)$$

$$F = G/H$$

not good enough!

Image deconvolution revisited

Solution: include prior: impose power spectrum on F (probalistic!)

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

Maximum a posteriori (MAP)

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

acts as a regularizer

$$\mathcal{L}'(F) = \sum_u \frac{1}{|N(u)|^2}\left|F(u)H(u) - G(u)\right|^2 + \sum_u \frac{|F(u)|^2}{S(u)}$$

Wiener filter!

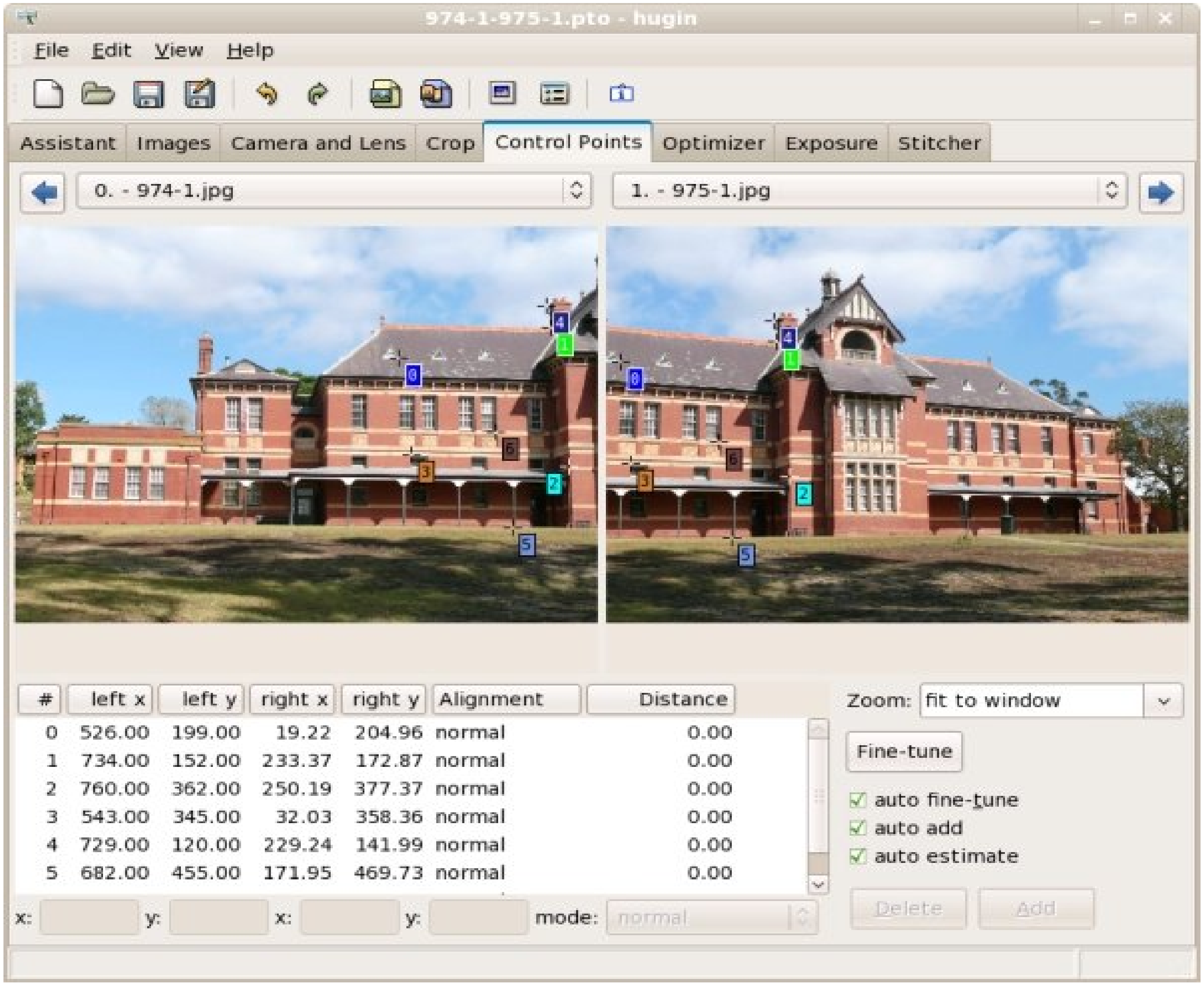
$$\frac{\partial \mathcal{L}'}{\partial F^*} = 0 \;\rightarrow\; \frac{1}{|N|^2}(FH - G)H^* + \frac{F}{S} \;\rightarrow\; F = \frac{H^* G}{\left(|H|^2 + \frac{|N|^2}{S}\right)}$$

Image registration

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.

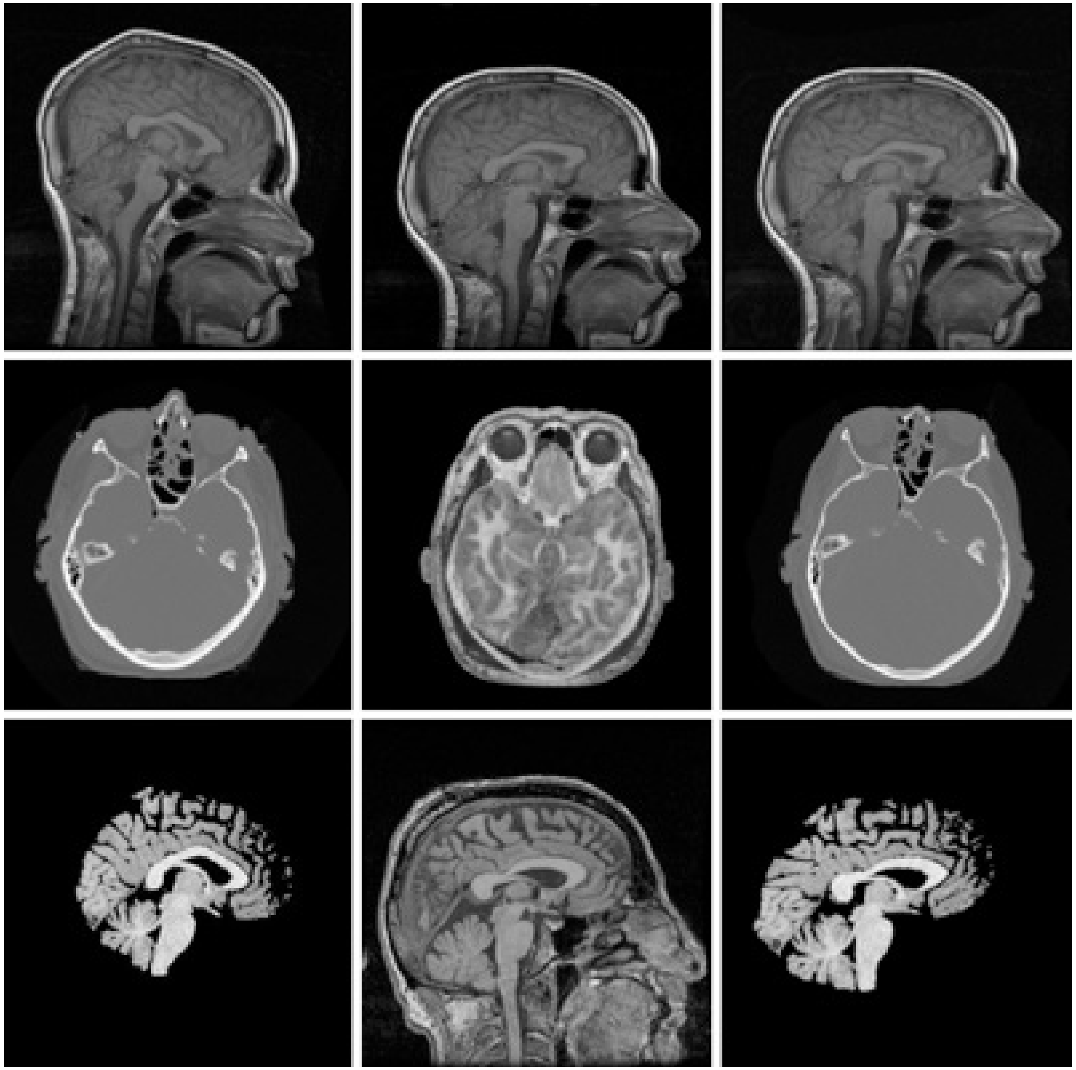
Control points for photo stitching



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

Image registration

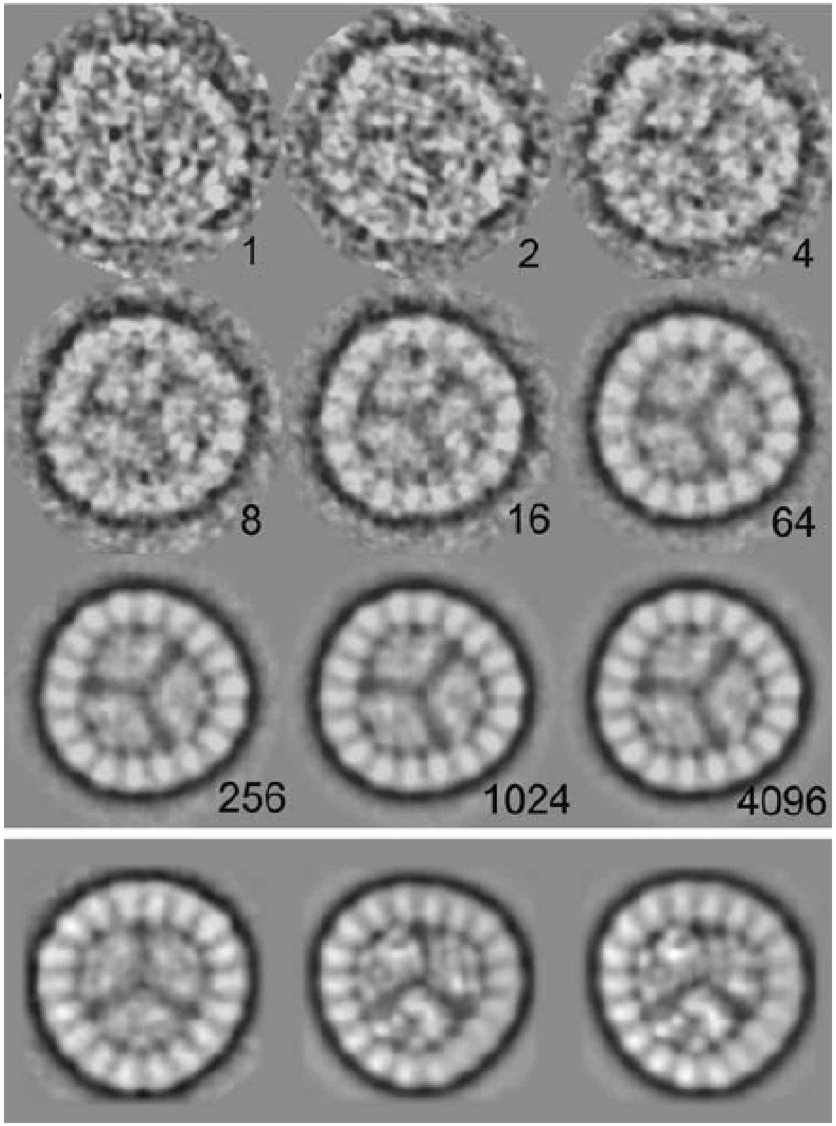
Medical image registration



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

Single particle analysis

1 measurement



"bad" average

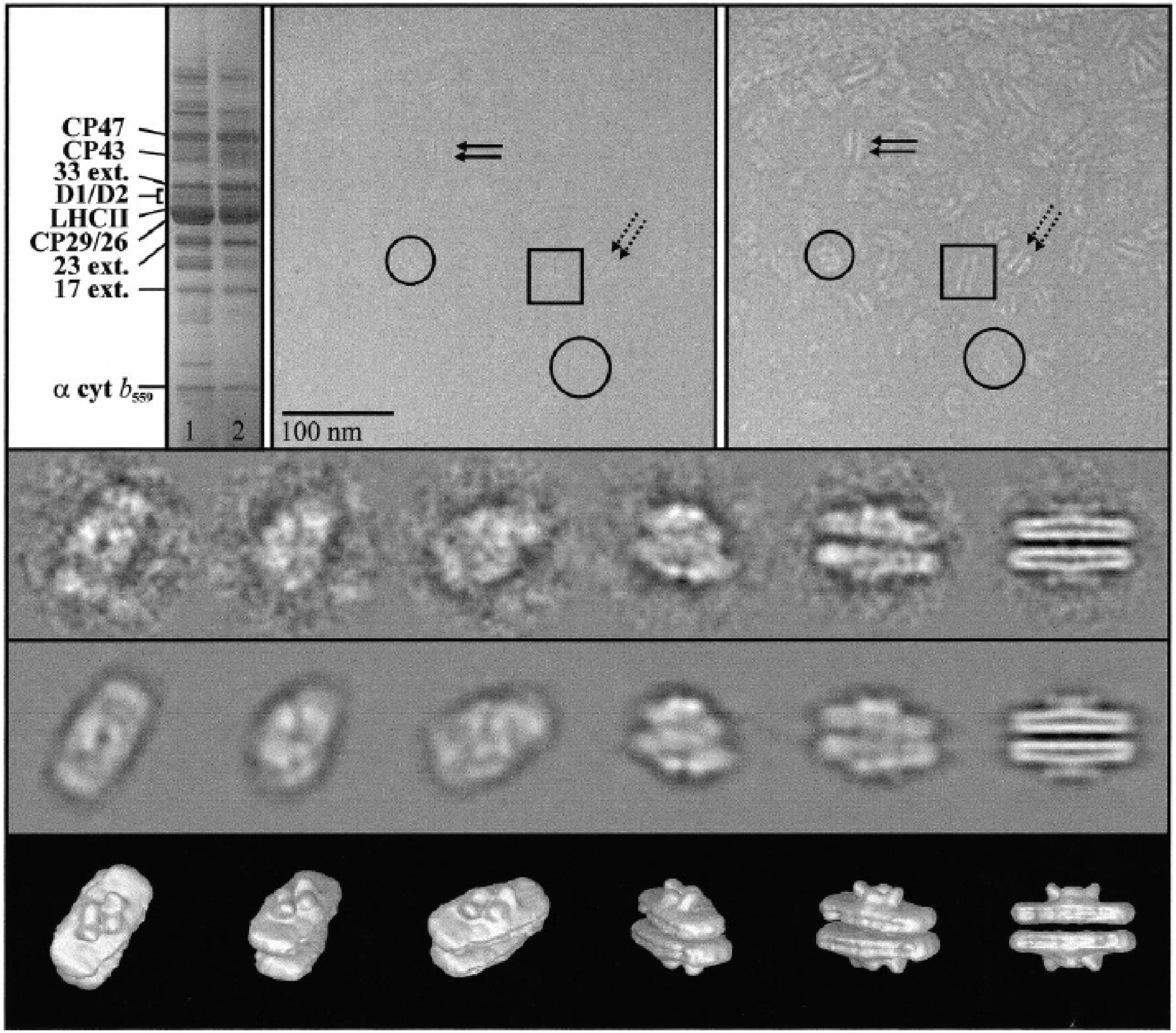
cryoEM

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)

Image registration

Maximum likelihood formulation

$$\mathcal{L} = \sum_i \sum_{\vec{r}} \left(I^{(i)}(\vec{r} - \vec{r}_i) - O(\vec{r}) \right)^2 = \sum_i \mathcal{L}_i$$

measured noisy images

shifts

"object"

unknown

"Expectation maximization" $\longrightarrow$ M step "maximization step"

E step "expectation step"

M step

$$\frac{\partial \mathcal{L}}{\partial O(r)} = 2 \sum_i^N \left(O(r) - I^{(i)}(r - r_i) \right) = 0$$

$$N O(r) = \sum_i I^{(i)}(r - r_i)$$

$$\boxed{O(r) = \frac{\sum_i I^{(i)}(r - r_i)}{N}}$$

Maximum likelihood formulation

E step: what are the most likely $\vec{r}_i$?

$$\mathcal{L}_i = \sum_r \left(I^{(i)^2}(r-r_i) + O^2(r) - 2\, I(r-r_i)\, O(r) \right)$$

constants

cross-correlation

$\Rightarrow$ max cross-correlation!

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, ...