Image Processing for Physicists

Prof. Pierre Thibault

pthibault@units.it

Maximum likelihood principle

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: $p(x|\alpha)$

likelihood of α given observations x:

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

$\ell(\alpha|x)$ is <u>not</u> the probability that the model is true (or a given α)

Maximum likelihood

Can easily be misunderstood...

Shroud of Turin

missing a prior

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

ℓ(it really was Jesus | shroud looks like this) : high because = p
but : not a probability!

Bayes' theorem

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

"prior"

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

"posterior"

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

not relevant: fixed number given by the observation x

Maximum likelihood & optimization

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

  $\rightarrow$ Maximum likelihood maximize $l(\alpha|x)$

  or

  $\rightarrow$ Maximum a posteriori (MAP)

  maximize $l(\alpha|x)\,p(\alpha)$ additional knowledge about α

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

Example: a biased coin

$p(\text{result is Head} \mid \alpha) = \alpha$

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

α: parameter to extract from observations

Oberations: N_H = number of heads

N_T = number of tails

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

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

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

Example: Gaussian model

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

2) Many independent variables with the same distribution:

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

$$= \ell(\mu, \sigma^2 | x_1, x_2, x_3, \dots, x_N)$$

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

if σ^2 is known: this is equivalent to least squares!

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

$\hat{\mu}$, $\hat{\sigma}_2$: maximum likelihood estimators for μ and σ^2

Example: Gaussian model

3) M variables not identically distributed and not independent

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

means   covariance matrix   determinant

If N measurements are made:

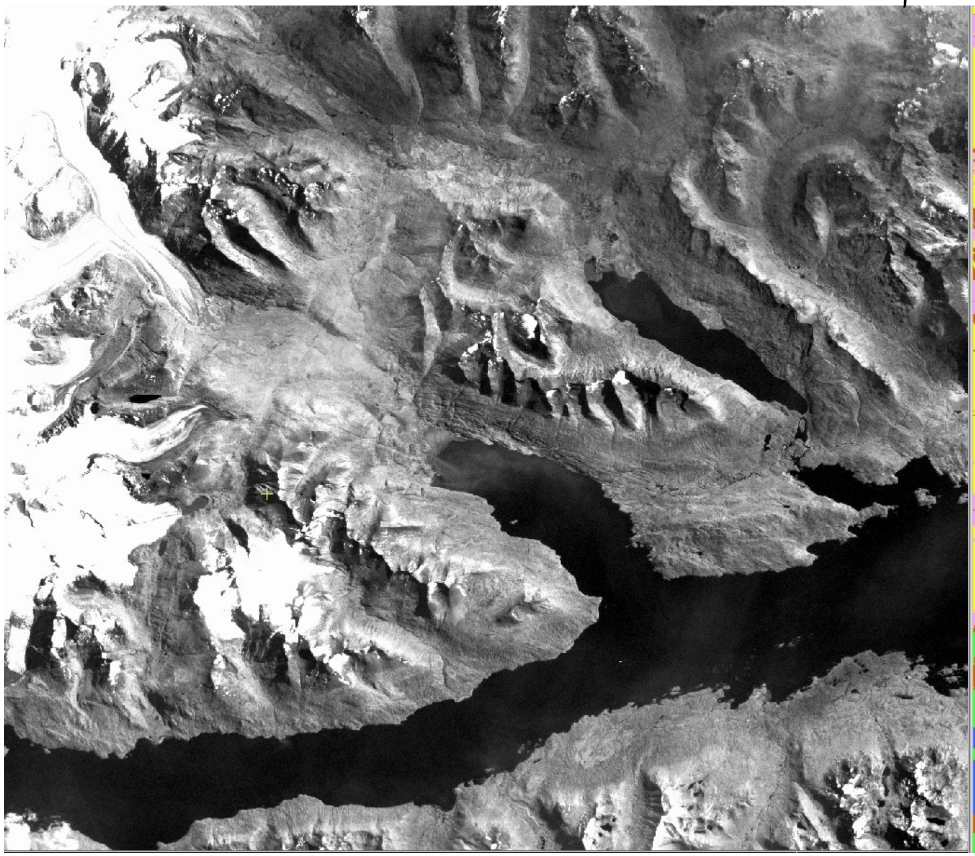
$$p(\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)} \mid \vec{\mu}, C) = \frac{1}{(2\pi)^{\frac{MN}{2}} |C|^{\frac{N}{2}}} \exp\left(-\frac{1}{2}\sum_i (x^{(i)}-\mu)^T C^{-1} (x^{(i)}-\mu)\right)$$

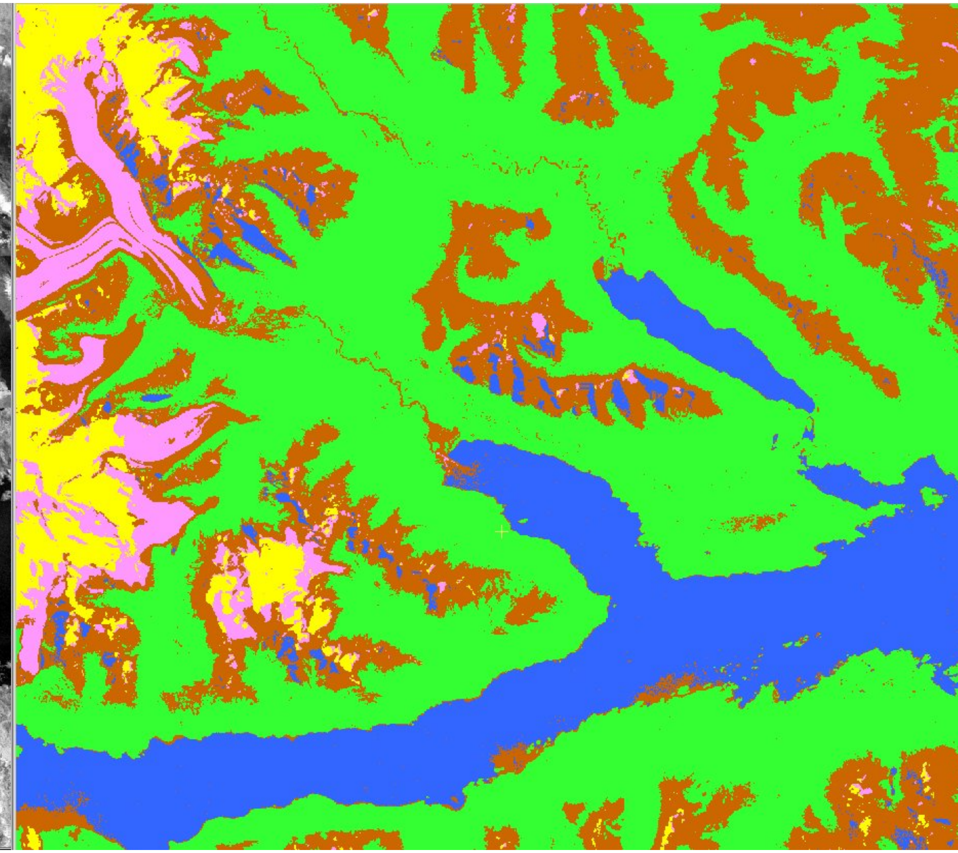
$$\ell = p, \quad \mathcal{L} = -\ln \ell, \quad \frac{\partial \mathcal{L}}{\partial \vec{\mu}} = 0 \quad \frac{\partial \mathcal{L}}{\partial C} = 0$$

$$\Rightarrow \vec{\mu} = \frac{1}{N}\sum_i \vec{x}^{(i)} \qquad C_{\ell m} = \frac{1}{N} \sum_i (x_\ell^{(i)} - \mu_\ell)(x_m^{(i)} - \mu_m)$$

Image classification

stack of images





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 B	30
Band 3 - Green	0.53-0.59 G	30
Band 4 - Red	0.64-0.67 R	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

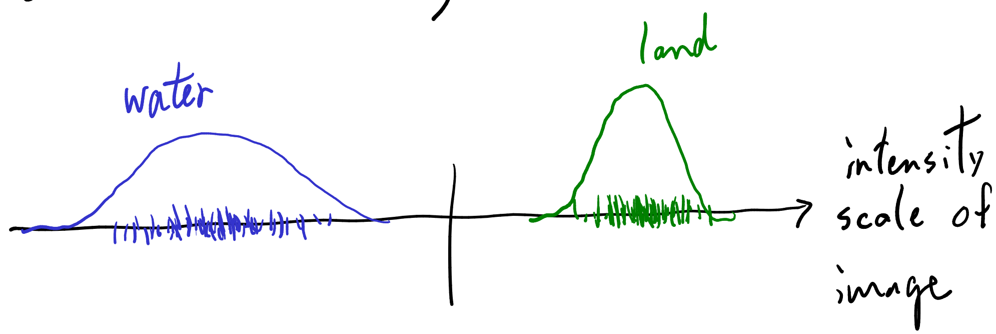
Image classification

– a kind of segmentation!

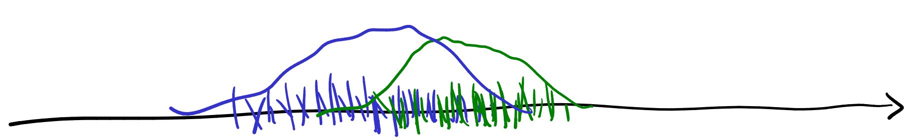
Supervised Maximum Likelihood Classification

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

1D (e.g. one of the bands)



in reality



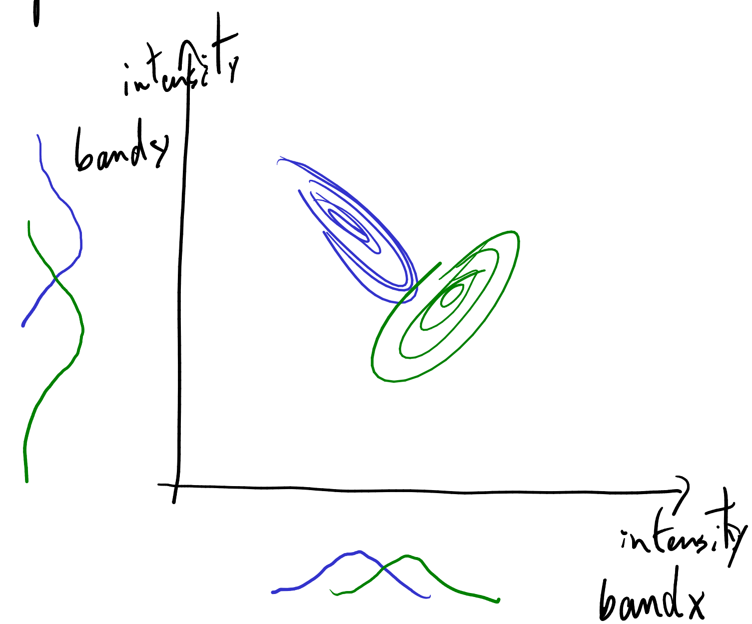


Image classification

Supervised Maximum Likelihood Classification

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

$\parallel$ likelihood! $p(\text{pixel} \mid \text{class})$

$\parallel$

$\ell(\text{class} \mid \text{pixel})$

e.g.

$$p_{water}(\vec{x} \mid \vec{\mu}_{water}, C_{water}) = (\ldots) \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu}_{water})^T C_{water}^{-1} (\vec{x}-\vec{\mu}_{water})\right)$$

$$L = -\ln(p_{water}) = \frac{1}{2}\ln|C_w| + \underbrace{\frac{1}{2}(\vec{x}-\vec{\mu}_w)^T C_w^{-1} (\vec{x}-\vec{\mu}_w)}_{\text{"Mahalanobis distance"}}$$

find minimum of this expression depending on class (water? field?, forest?...)

Image deconvolution revisited

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

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

PSF   true image   noise

Fourier space:

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

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

noise: normal distribution

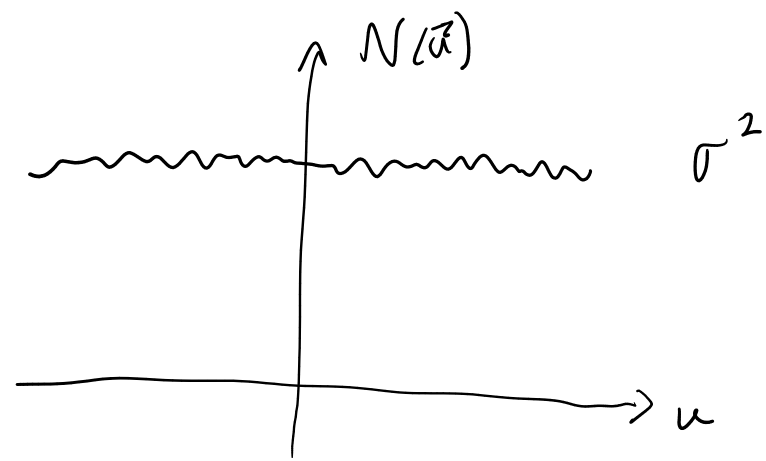


Image deconvolution revisited

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

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

$$\|$$

$$l(F(\vec{u})\,|\,G(\vec{u}))$$

Maximum likelihood? $-\ln(l) = \frac{1}{2}\sum_u \frac{1}{|N(\vec{u})|^2}\,|F(\vec{u})H(\vec{u}) - G(\vec{u})|^2$

minimum w.r.t. $F(\vec{u})$? $\underbrace{F = G/H}$

not good! noise amplification
division by small numbers, ...

Solution: include prior knowledge. Here: "impose" power spectrum on F

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

Image deconvolution revisited

Maximum a posteriori

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

(Bayes)

$$\mathcal{L}'(F(\vec{u})) = \mathcal{L} - \ln(p(F))$$

$$= \frac{1}{2}\sum_{u} \frac{1}{|N(u)|^2} |F(\vec{u})H(\vec{u}) - G(\vec{u})|^2 + \overbrace{\frac{1}{2}\sum_{u} \frac{|F(u)|^2}{S(u)}}^{\text{this can be seen as a regularizer}}$$

minimize w.r.t. $F(\vec{u})$

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

function $\mathbb{C}^N \to \mathbb{R}$:

as if z and z^* are independent

"Wirtinger derivatives"

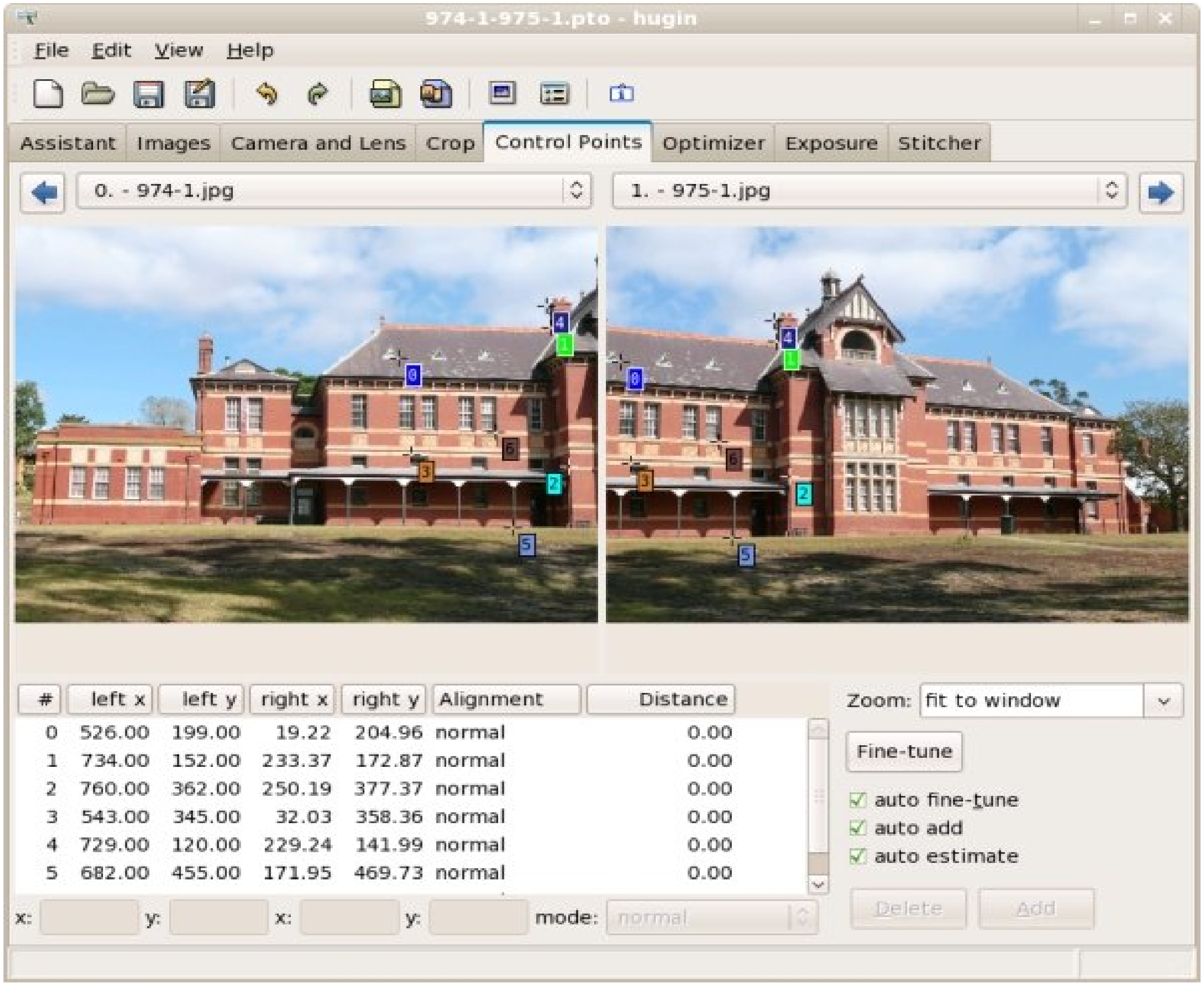
$$F(\vec{u})\frac{|H(u)|^2}{|N(u)|^2} - \frac{G(\vec{u})H^*(\vec{u})}{|N(\vec{u})|^2} + \frac{F(\vec{u})}{S(u)} = 0$$

$$F(\vec{u}) = \frac{H^*(\vec{u})}{\left(|H(\vec{u})|^2 + |N|^2/S\right)} G(\vec{u}) \quad \text{Wiener filter}$$

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.

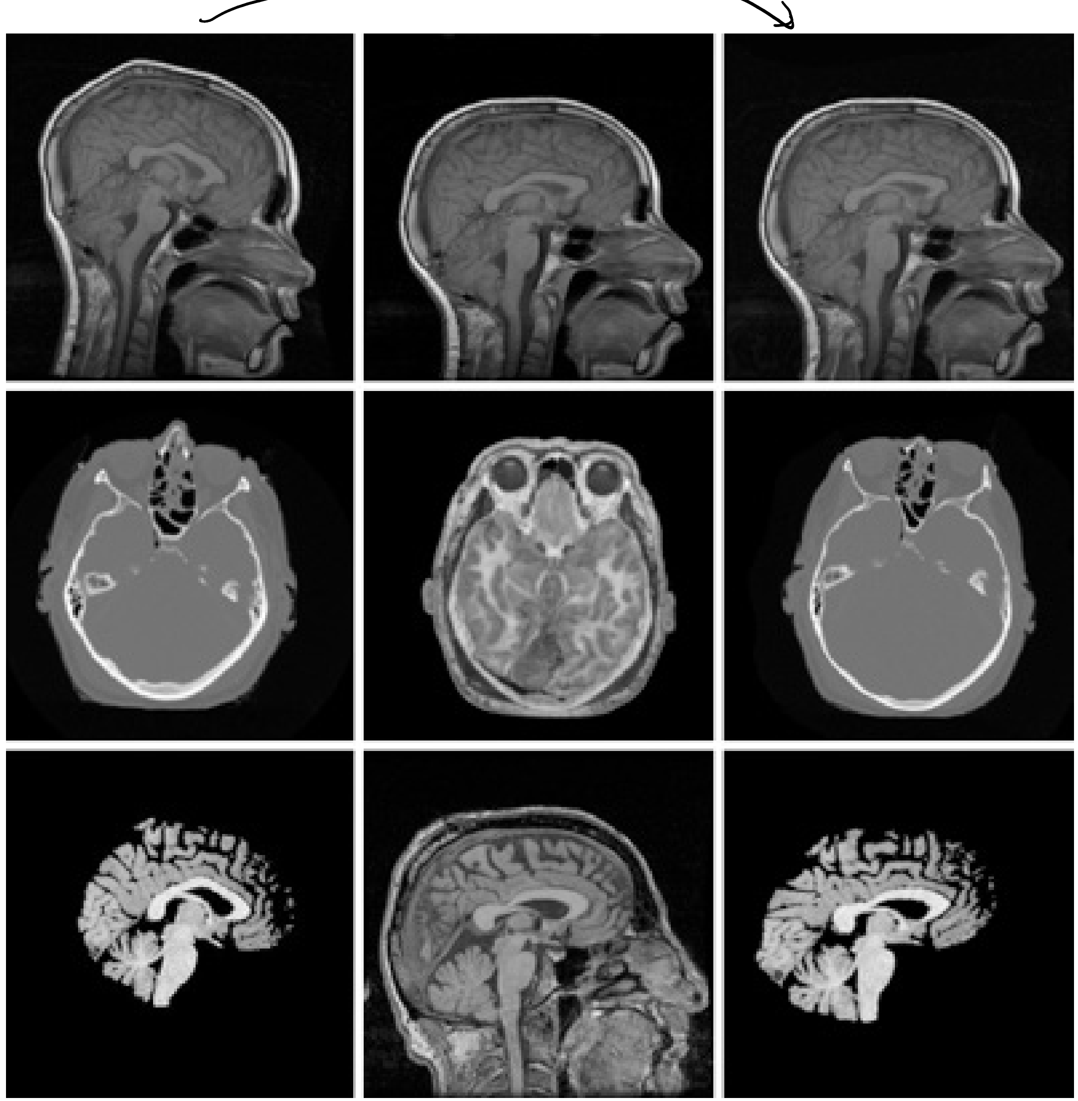
Control points for photo stitching



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

Image registration

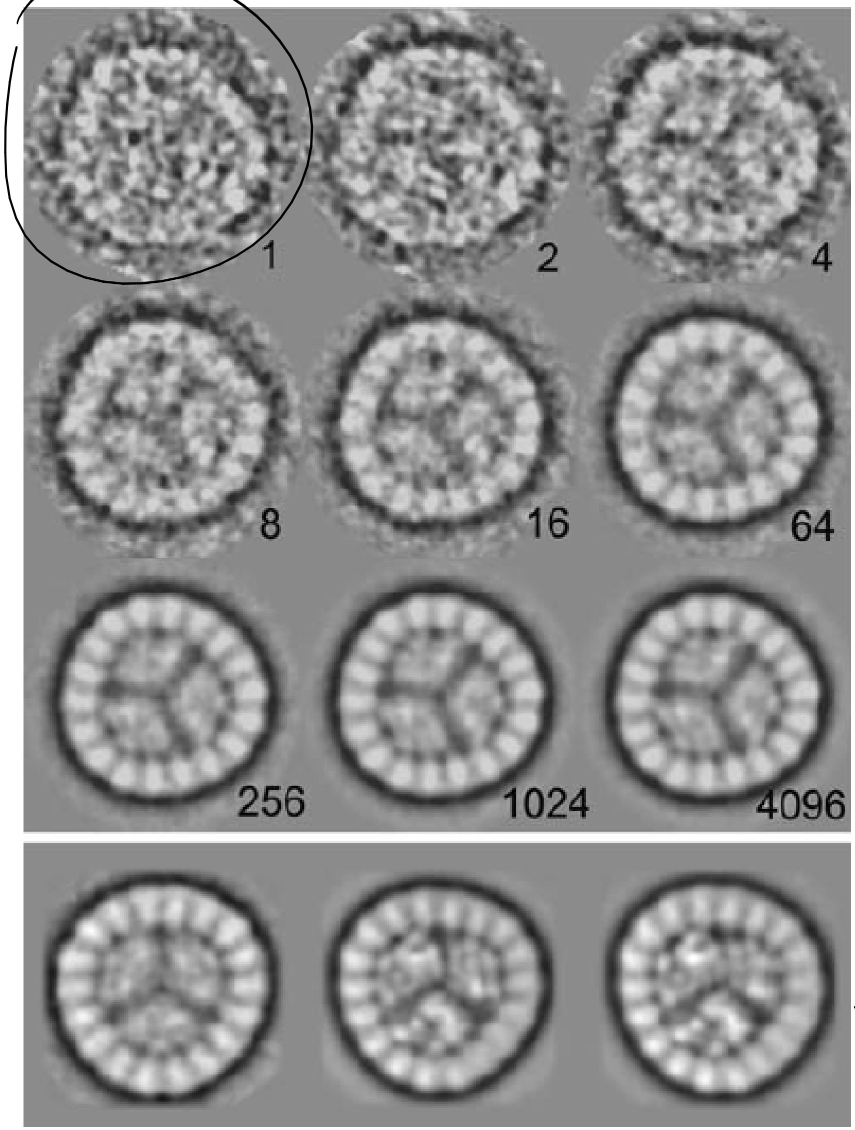
Medical image registration



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

Single particle analysis

a single TEM image

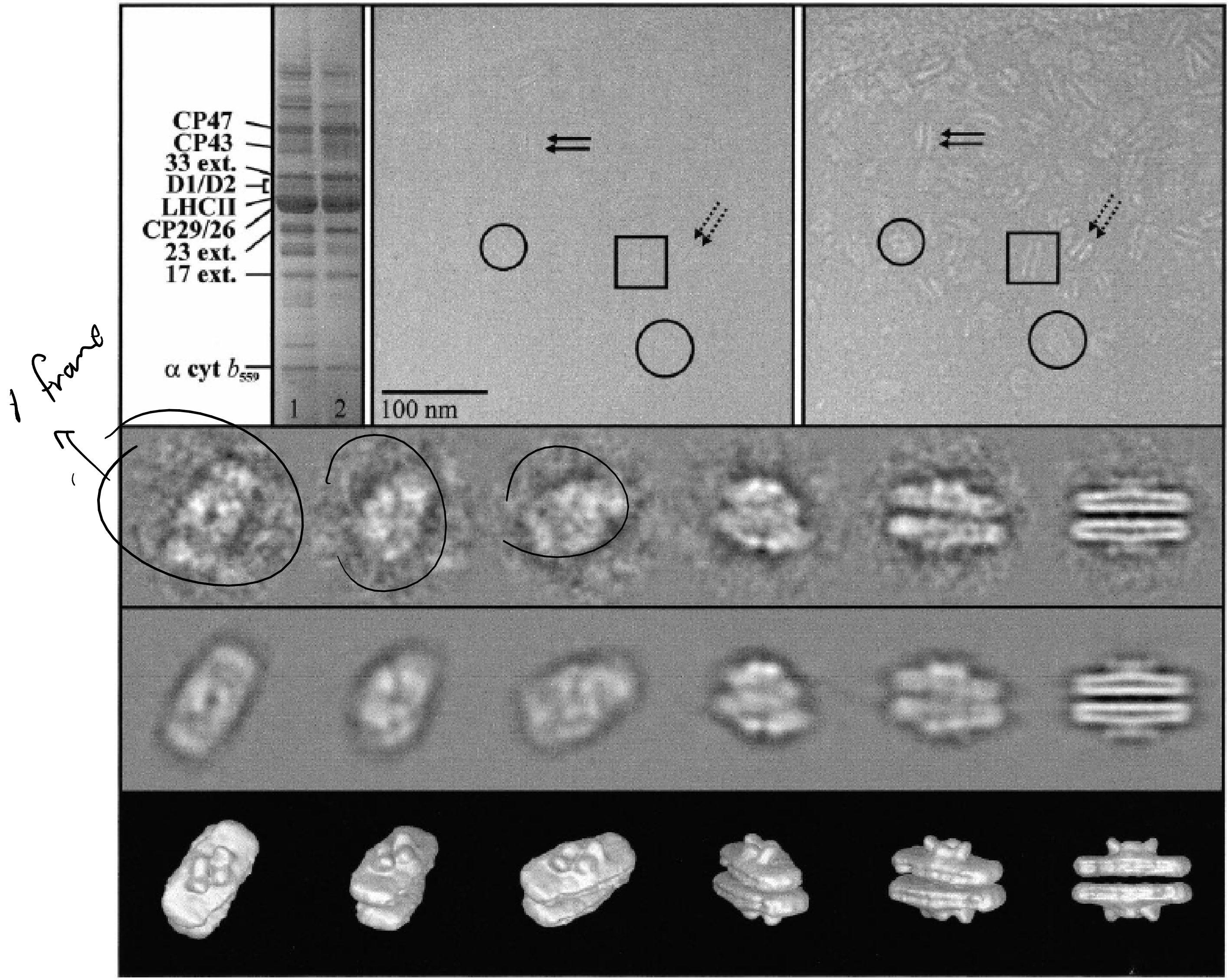


cryo EM
electron microscopy

(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)

Image registration

image stacking
↳ registration combined
with averaging

Summary

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