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 observation x:

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

$\ell(\alpha|x)$ is <u>not</u> the probability that the model is correct

<u>not</u> a probability distribution for α

Maximum likelihood

Can easily be misunderstood...

Shroud of Turin

missing: a <u>prior</u>

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

ℓ (it really was Jesus | the shroud has this appearance) : high also because = p

Bayes' theorem

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

"prior"

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

"posterior"

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

not relevant: fixed given an observation

Maximum likelihood & optimization

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

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

  or

  → Maximum a posteriori (MAP)

  maximize $\ell(\alpha|x)\, p(\alpha)$

  additional knowledge about α

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

Example: a biased coin

$p(H|\alpha) = \alpha$

$p(T|\alpha) = 1-\alpha$

α : parameter to extract
(= probability of head)

Observation: N_H = number of heads

N_T = number of tails

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

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

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

Example: Gaussian model

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

2) Many independent variables with same distribution:

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

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

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

if σ^2 is known, we recover least square principle

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

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

Example: Gaussian model

3) M variables not identically distributed and not independent

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

covariance matrix

determinant

If N measurements are made:

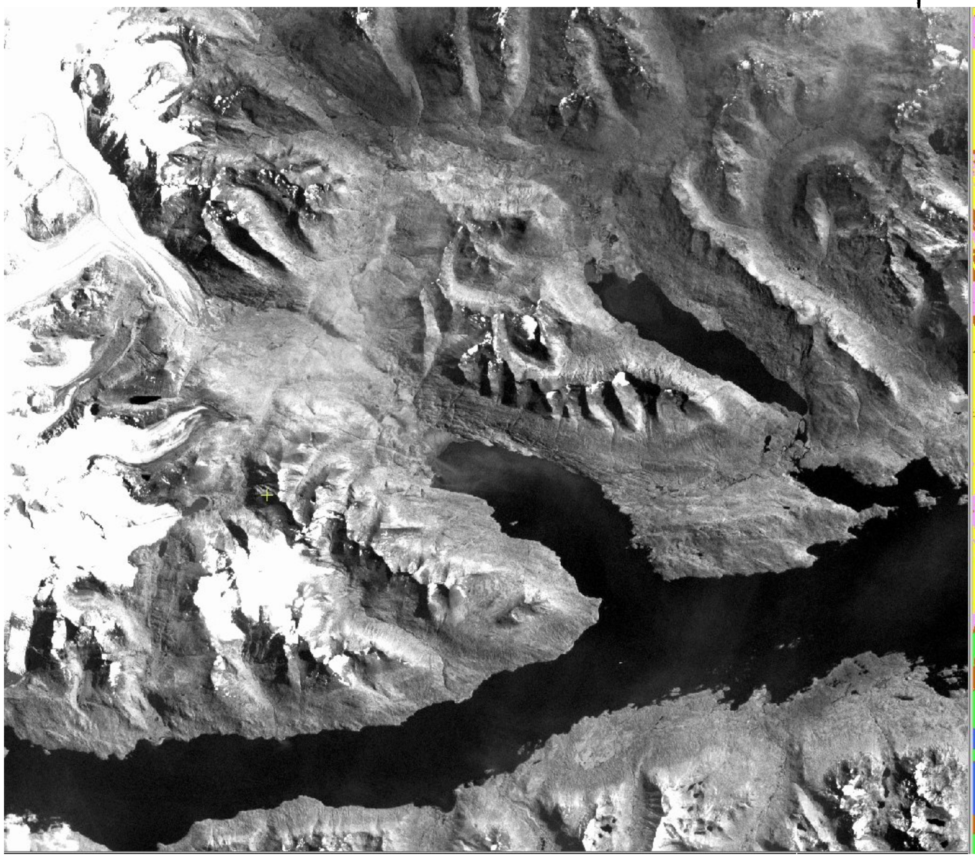
$$p(\vec{x}^{(0)}, \vec{x}^{(1)}, \vec{x}^{(2)}, \ldots \vec{x}^{(N)} \mid \vec{\mu}, C) = \frac{1}{(2\pi)^{\frac{MN}{2}} |C|^{N/2}} \exp\left(-\frac{1}{2}\sum_i (x^{(i)}-\mu)^T C^{-1}(x^{(i)}-\mu)\right)$$

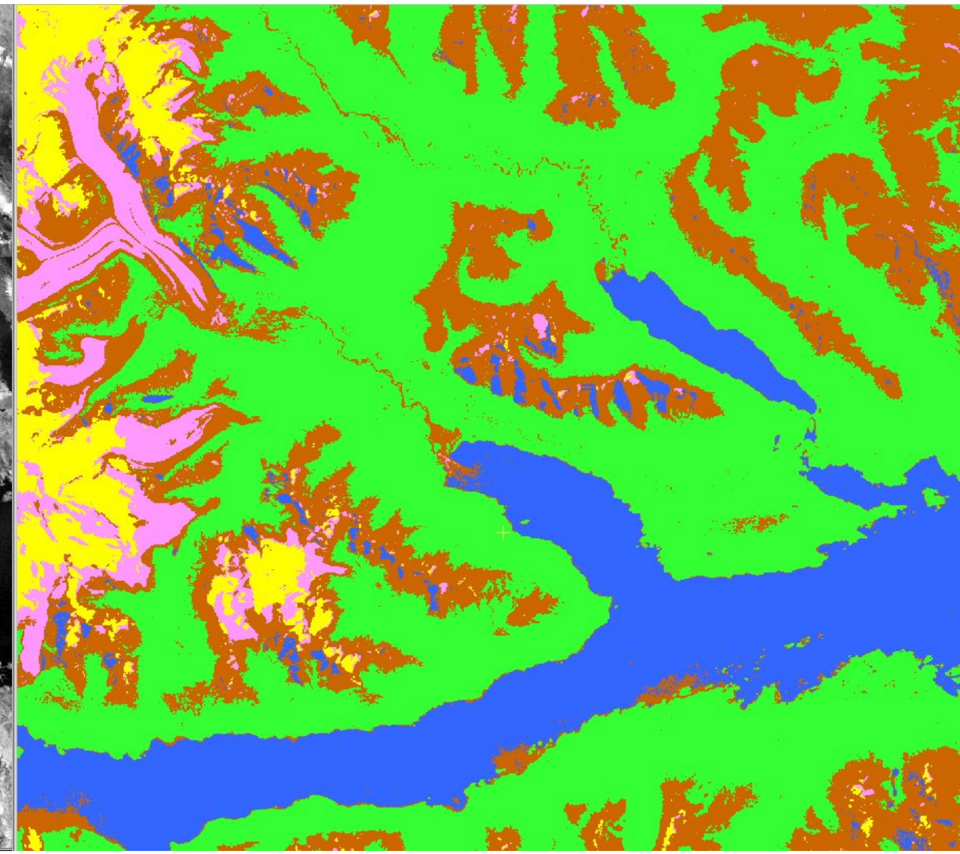
$$l = p \quad , \quad \mathcal{L} = -\ln l \qquad \frac{\partial \mathcal{L}}{\partial \vec{\mu}} = 0 \quad \frac{\partial \mathcal{L}}{\partial C} = 0$$

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

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	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

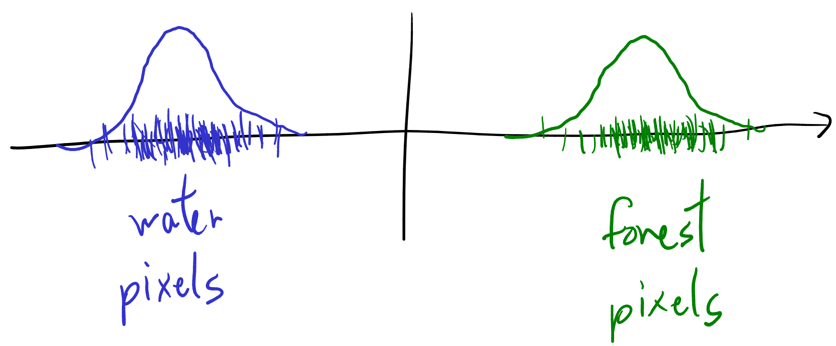
Image classification

Supervised Maximum Likelihood Classification

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

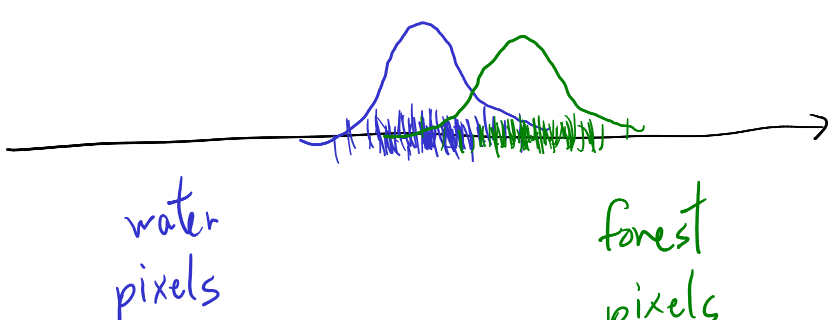
1st step: evaluate parameters for these distributions from known samples

1D



intensity of a single image

in reality:



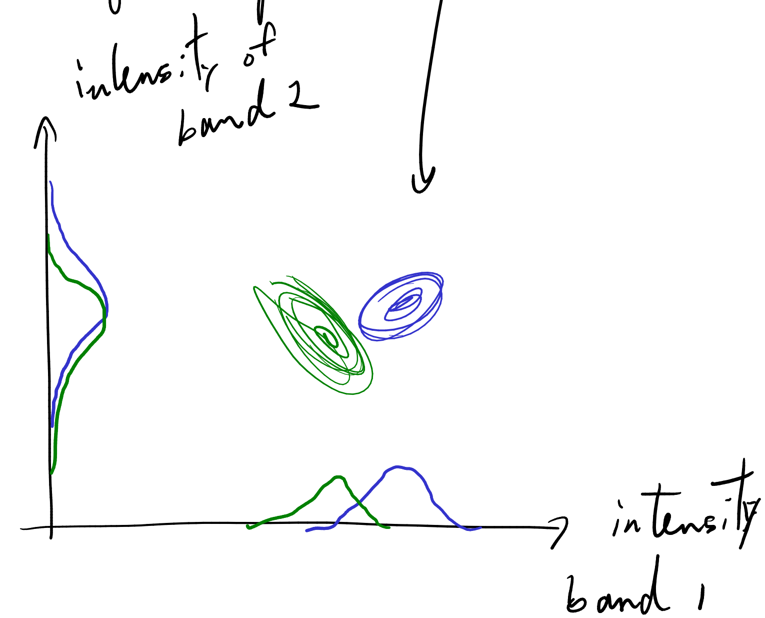


Image classification

Supervised Maximum Likelihood Classification

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

likelihood $\quad p(\text{pixel} \mid \text{class}) = \ell(\text{class} \mid \text{pixel})$

e.g. water class:

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

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

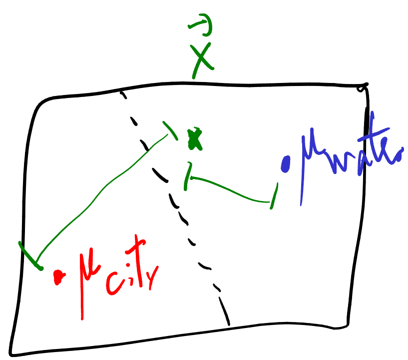


Image deconvolution revisited

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

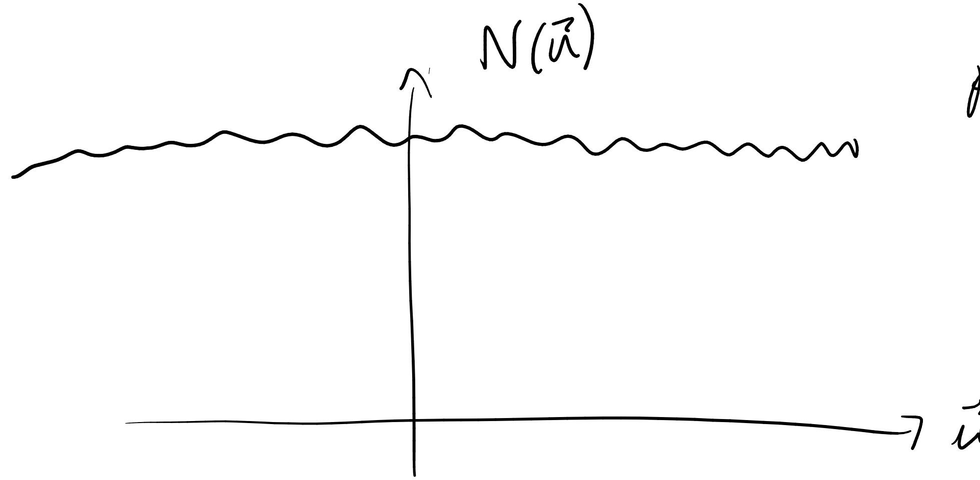
$$\underbrace{g(\vec{r})}_{\text{measurement}} = (\underbrace{h}_{\text{PSF}} * \underbrace{f}_{\text{true image}})(\vec{r}) + \underbrace{n(\vec{r})}_{\text{noise}}$$

zero-mean

Fourier space

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

Often, a reasonable assumption is that $N(\vec{u})$ is flat (white noise)



probability distribution for $N(\vec{u})$: gaussian

$\sigma^2 = |N(\vec{u})|^2$

Image deconvolution revisited

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

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

Maximum likelihood? .... $F(\vec{u}) = G(\vec{u})/H(\vec{u})$

not good, unstable, amplifies noise...

Solution: include prior knowledge about F.

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

S: power spectrum

Image deconvolution revisited

Maximum a posteriori (MAP): maximize

$$p(F(\vec{u}) \mid G(\vec{u}))\, p(F(\vec{u}))$$

$$\Updownarrow$$

minimize $-\ln$: $\mathcal{L} = \sum_u \left[\frac{1}{|N(u)|^2} \left| F(\vec{u})\, G(\vec{u}) - H(\vec{u}) \right|^2 + \frac{|F(\vec{u})|^2}{S(\vec{u})} \right]$

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

Wirtinger calculus: range of $\mathcal{L}$ is $\mathbb{R}$

$\Rightarrow$ $F(\vec{u})$ and $F^*(\vec{u})$ can treated as independent variables

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

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

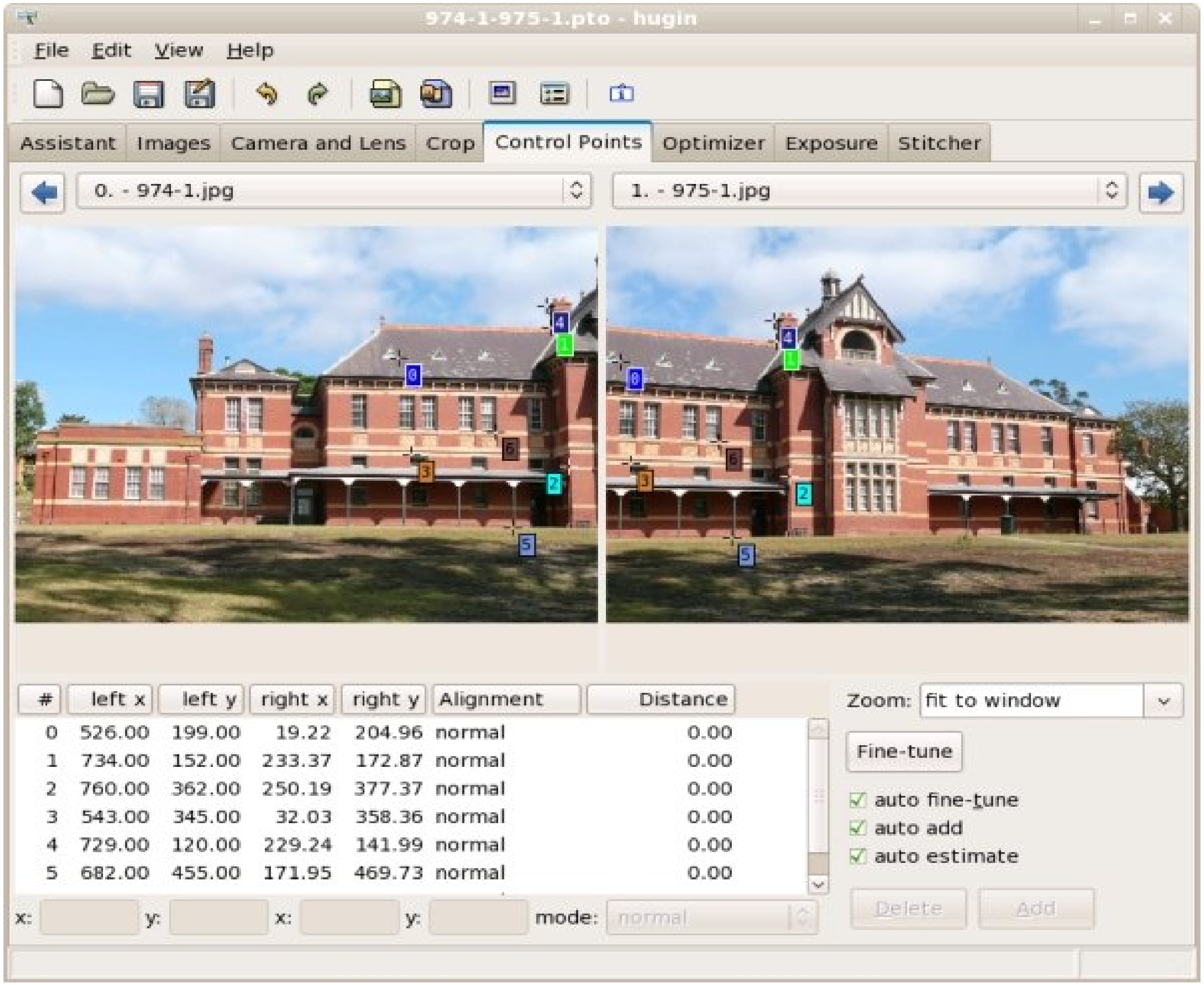
Wiener filter

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.

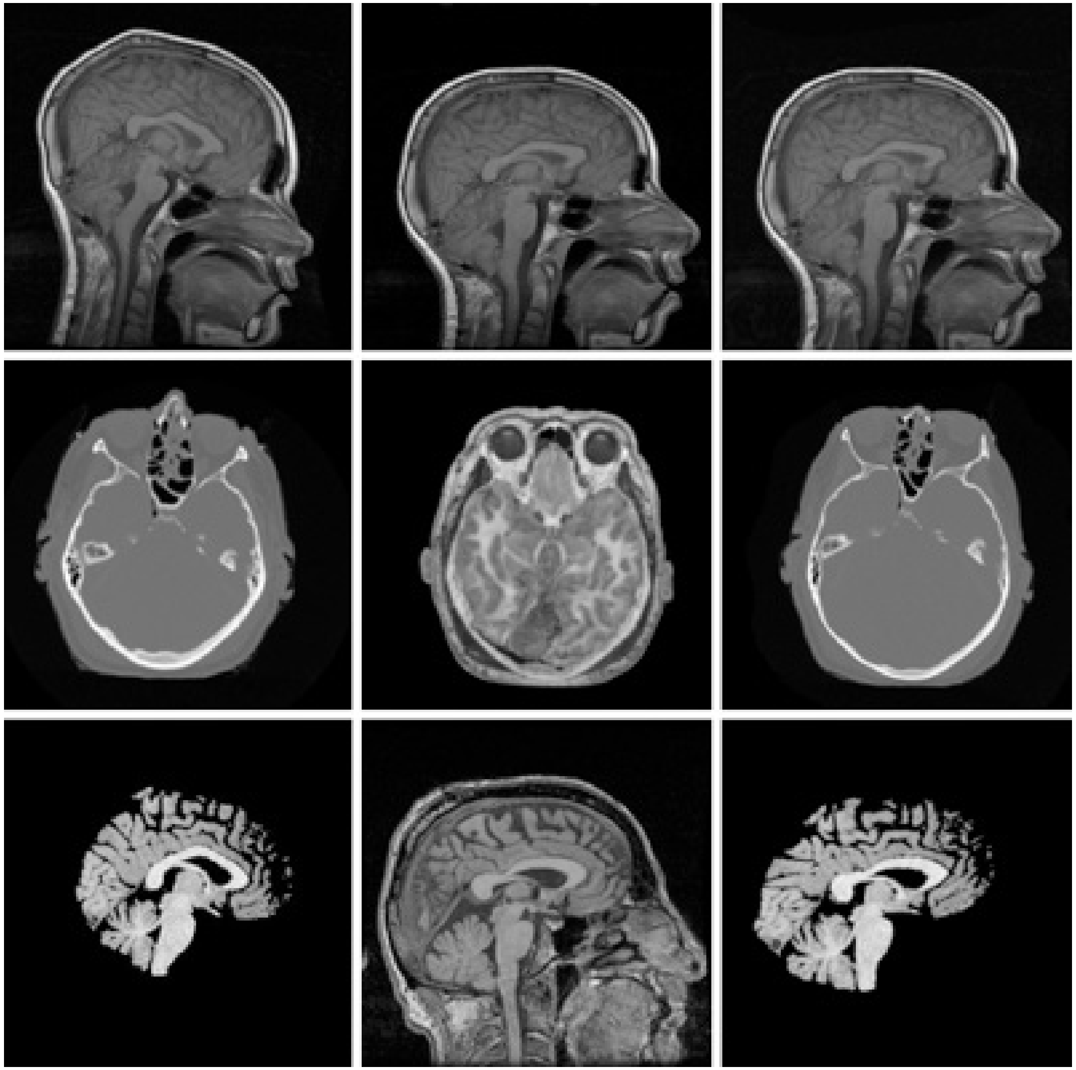
Control points for photo stitching



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

Image registration

Medical image registration

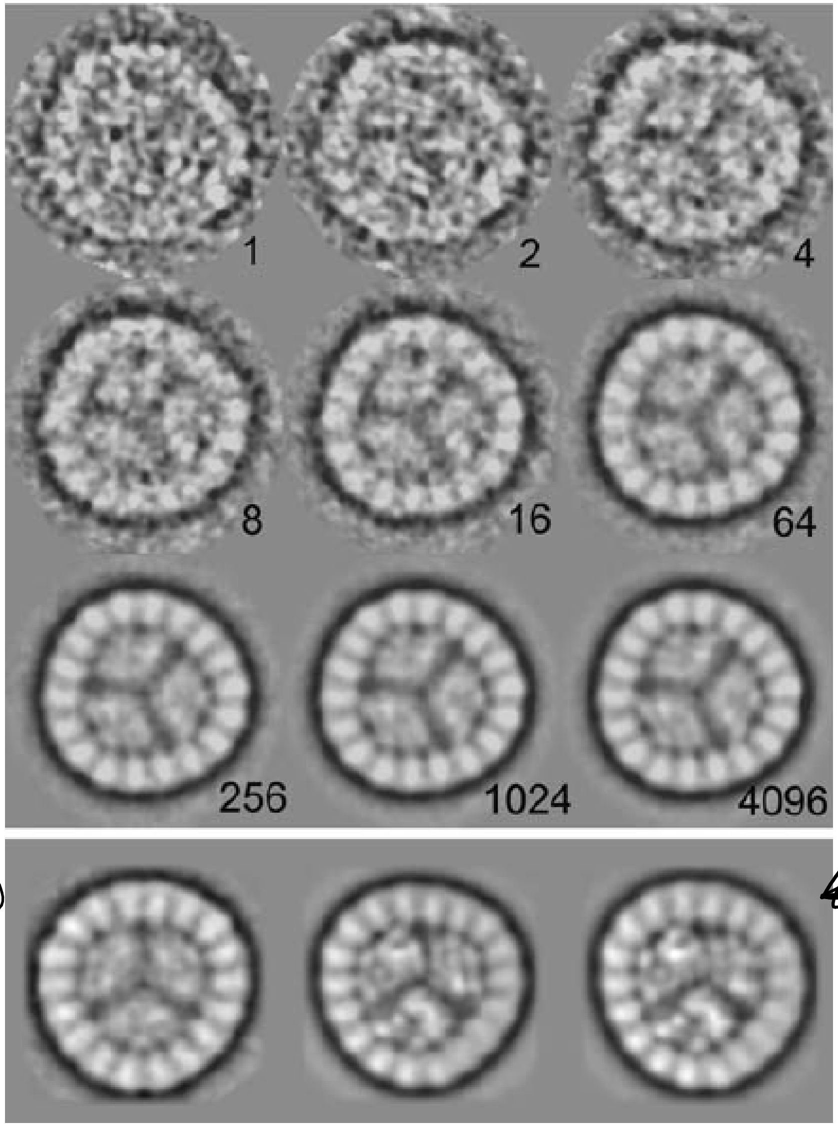


Source: http://www.cs.dartmouth.edu/farid/Hany_Farid/

Single particle analysis

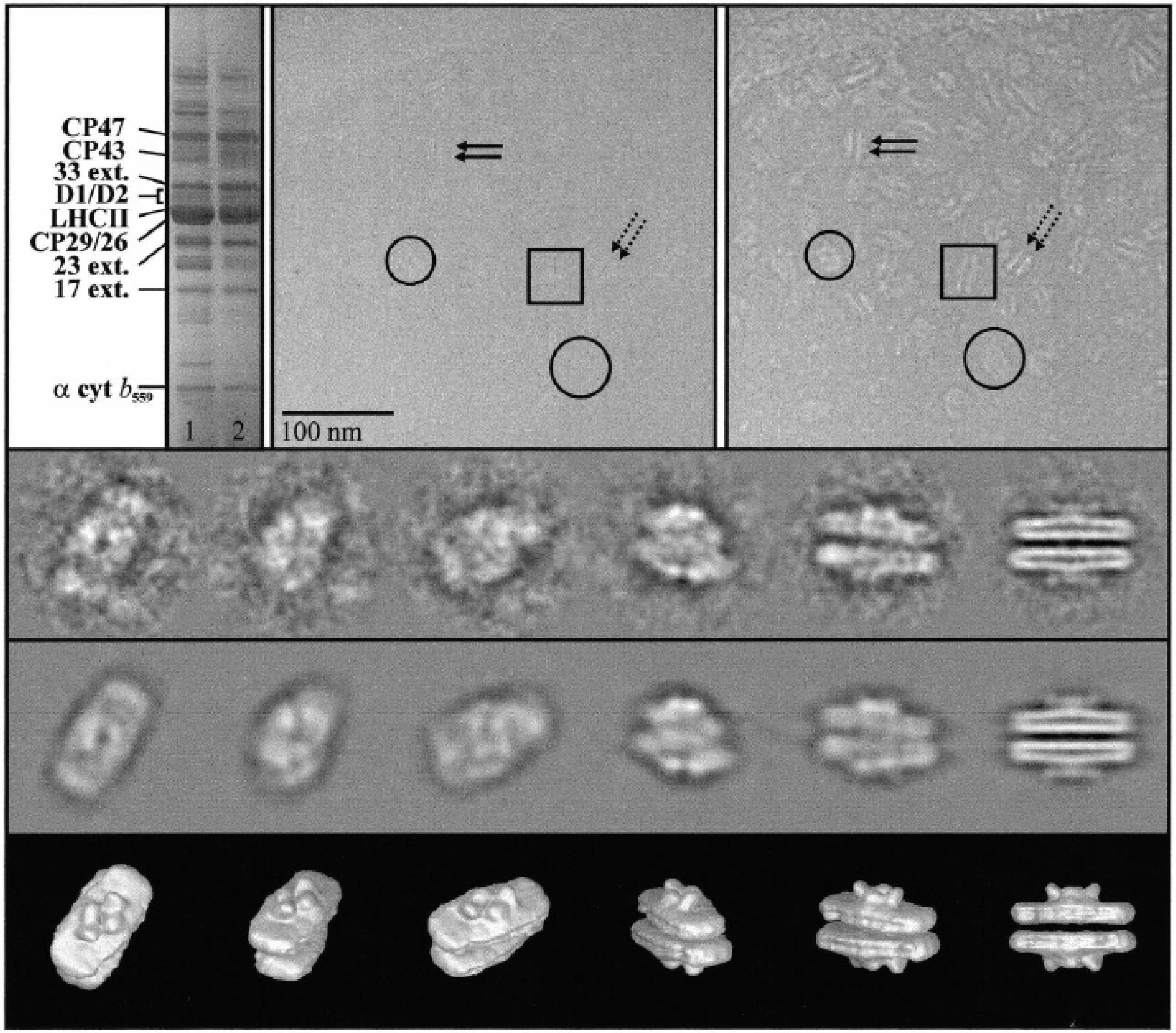
one noisy image

TEM images



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

Single particle analysis



Nobel 2017

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

Image registration

Maximum likelihood formulation

$I_i(\vec{r})$: multiple noisy images

$I_0(\vec{r})$: real image

$$p\left(I_i(\vec{r}) \mid I_0(\vec{r}), \vec{r}_i\right) \propto \exp\overbrace{\left(-\frac{1}{2}\sum_{\vec{r}} \frac{1}{\sigma^2}\left(I_i(\vec{r}+\vec{r}_i) - I_0(\vec{r})\right)^2\right)}^{\text{true because additive gaussian noise}}$$

$$p\left(I_1(\vec{r}), I_2(\vec{r}) \ldots \mid I_0(\vec{r}), \vec{r}_1, \vec{r}_2, \ldots\right) \propto \exp\left(-\frac{1}{2}\sum_{i} \frac{1}{\sigma^2} \sum_{\vec{r}} \left(\qquad \right)^2\right)$$

strategy : minimize in alternance w.r.t. I_0 and $\{\vec{r}_1, \vec{r}_2, \vec{r}_3 \ldots\}$

more complicated and robust approach (to noise) is called Expectation-maximization algorithm

Maximum likelihood formulation

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, ...