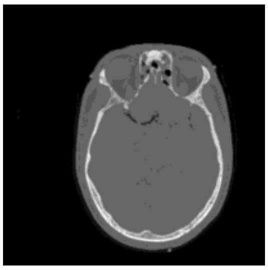
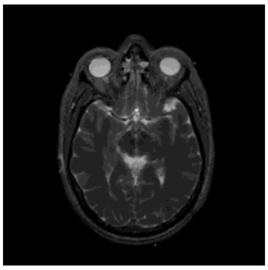
Joint Entropy Registration

$\boxed{L\ 30}$

Goal: An introduction to information-theoretic methods for image registration.

Consider the situation of registering a CT scan (f) to an MRI (g) of the same anatomy.

$f =$ 

$g =$ 

This is an example of multimodal registration.
In particular, there is no known functional relationship between the intensities of the two images.

e.g. $\nexists\ a, b$ s.t. $a + b f(x) = g(x)$ (linear)

$\nexists\ a, b, c$ s.t. $a + b f(x) + c f^2(x) = g(x)$ (quadratic)

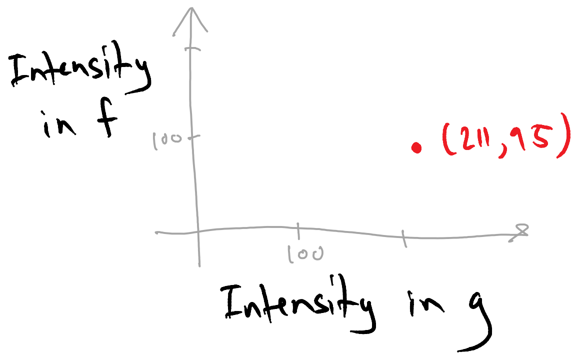
$\vdots$ etc.

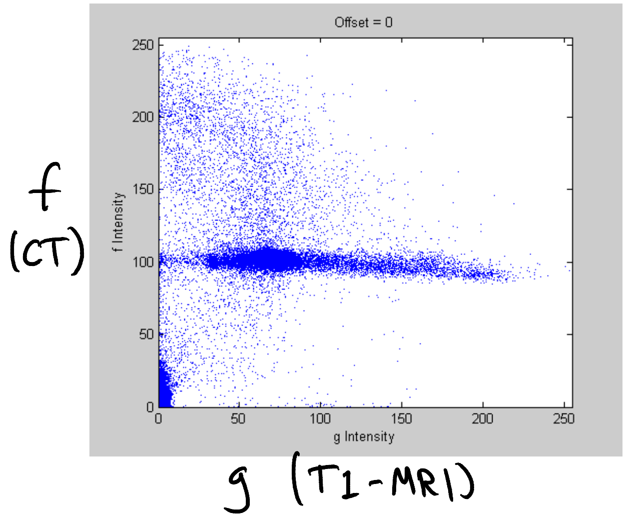
To see this more clearly, let`s look at images that are already registered. Each pixel has an intensity from each image, so each pixel can be represented as a point in 2-space.

e.g. $(\text{row}, \text{col}) = (5, 18)$

$g(5,18) = 211$ $\qquad f(5,18) = 95$

$\therefore$ we plot a point at $(211, 95)$





(Brilliant Matlab demo)

The dispersion (compactness) of the Joint Intensity Scatter Plot (JISP) can be used as a measure of registration. But how do we get a number that reflects the dispersion of the JISP?

Entropy

The entropy of a discrete random variable X is

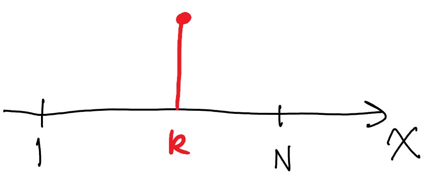
$$H(X) = -\sum_i p(x_i) \ln p(x_i)$$

where $p(x)$ is the prob. of observing x.

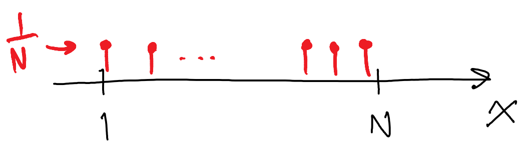
Entropy is a measure of disorder or randomness. In particular, the more spread out a signal's histogram, the higher the entropy.

eg. Consider two signals, one with a constant value, and the other random. Here are their pdf's:

Const. val. X_1



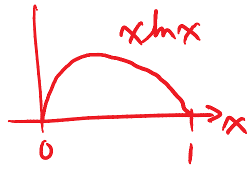
Random X_2



$$p(x) = \begin{cases} 1 & \text{if } x = k \\ 0 & \text{otherwise} \end{cases} \qquad p(x) = \frac{1}{N}$$

For X_1, $\boxed{H(X_1)} = -1 \ln 1 - \sum_i 0 \ln 0 = \boxed{0}$

Note: $\lim_{x \to 0} x \ln x = 0$ (use L'Hôpital's theorem)



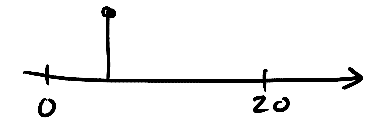
For X_2, $\boxed{H(X_2)} = -\sum_i \frac{1}{N} \ln \frac{1}{N} = -\ln \frac{1}{N} = \boxed{\ln N}$

Note: For continuous-domain variables, the formula for entropy is an integral.

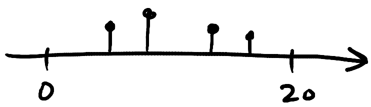
Interpretation

- A constant signal has no information, and no randomness. The next sample can easily be predicted. Hence, a low entropy histogram contains little information, but has only a few tight and compact peaks.

eg.



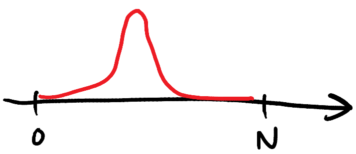
→ 5, 5, 5, 5, ___



→ 5, 16, 8, 8, 18, 8, 16, 5, 16, ___

- A signal with lots of different values, each occurring with similar probability, is more random or disorderly, and is not easily predictable (in general). Hence, high entropy is associated with lots of inforation, and the histogram tends to be more spread out.

eg.



vs.

In terms of registration, we can look at the joint histogram, a binned version of the JISP. We partition the joint intensity space into a collection of bins, and estimate the joint pdf by counting the sample frequency for each bin.

eg.

$$f = \begin{bmatrix} 1 & 3 & 8 \\ 4 & 2 & 9 \\ 5 & 2 & 7 \end{bmatrix} \qquad g = \begin{bmatrix} 4 & 2 & 2 \\ 1 & 6 & 5 \\ 3 & 1 & 7 \end{bmatrix}$$

$$\text{hist} = \begin{array}{c|c|c} f \backslash g & 1\text{-}4 & 5\text{-}9 \\ \hline 1\text{-}4 & |||| & | \\ \hline 5\text{-}9 & || & || \end{array} \Rightarrow \text{joint pdf} \approx \begin{bmatrix} \frac{4}{9} & \frac{1}{9} \\ \frac{2}{9} & \frac{2}{9} \end{bmatrix}$$

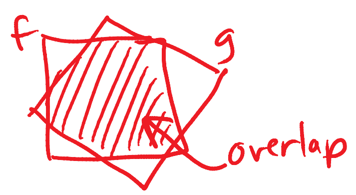
Then we evaluate the entropy of the joint histogram, and use that as a measure of registration. Thus, we can use the entropy of the joint histogram as a registration objective function. We the images are registered, the cost function will (hopefully) attain a **minimum** value.

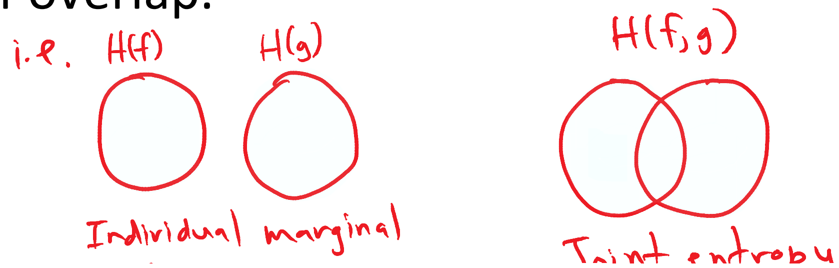
(Inspiring Matlab demo)

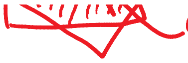
We refer to the entropy of the joint histogram as **"joint entropy"**, and denote it

$$H(f,g) = -\sum_{ij} p_{ij} \ln p_{ij}$$

It effectively measures the amount of information in f or g within the region of overlap.







overlap

Individual marginal entropies

Joint entropy