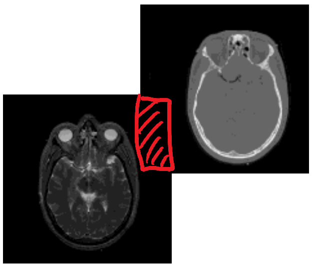
Mutual Information

L31

Goal: When does joint entropy fail, and what other information-theoretic objective functions work better?

There is a problem with joint entropy:
Consider the case when only the background of the images overlap.



The histogram is only computed for the overlapping portion. As this region gets smaller, there are fewer samples, and it's easier to achieve a single peak in the histogram.

ie. Once it's only background, then the joint histogram will be all zero except for one bin containing the background intensities for both images.

$\Rightarrow$ tight, compact histogram

$\Rightarrow$ low joint entropy

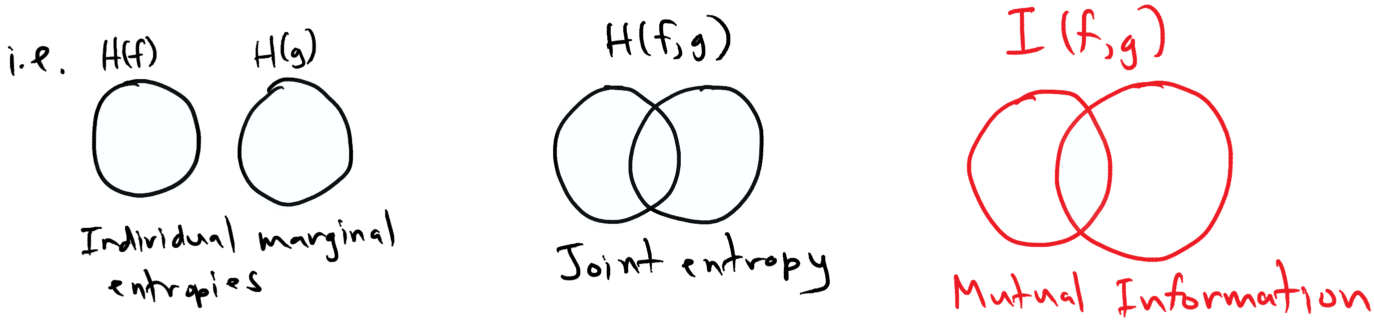
(increadible Matlab demo)

The cost function somehow has to also encourage nontrivial image content in the overlap region.

Mutual Information

Recall that H(f,g) is the amount of information that is in either f or g or both. If you subtract H(f,g) from the sum of H(f) and H(g), you are left with the total amount of information that both f and g share. This is called the mutual information of f and g.

i.e.



Mutual Information is defined as:

$$I(f,g) = H(f) + H(g) - H(f,g)$$

Recall that

$$H(f) = -\sum_i p_i^{(f)} \ln p_i^{(f)} \qquad H(g) = -\sum_j p_j^{(g)} \ln p_j^{(g)}$$

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

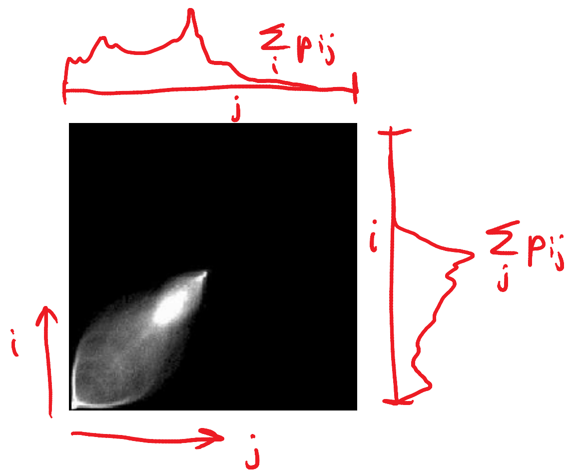
Thus,

$$I(f,g) = -\sum_i p_i^{(f)} \ln p_i^{(f)} - \sum_j p_j^{(g)} \ln p_j^{(g)} + \sum_i \sum_j p_{ij} \ln p_{ij}$$

Notice that

$$p_i^{(f)} = \sum_j p_{ij}$$

$$p_j^{(g)} = \sum_i p_{ij}$$



We call $p_i^{(f)}$ and $p_j^{(g)}$ the marginal probabilities.

Continuing the derivation...

$$I(f,g) = -\sum p_i^{(f)} \ln p_i^{(f)} - \sum p_j^{(g)} \ln p_j^{(g)} + \sum \sum p_{ij} \ln p_{ij}$$

Continuing the derivation...

$$I(f,g) = -\sum_i p_i^{(f)} \ln p_i^{(f)} - \sum_j p_j^{(g)} \ln p_j^{(g)} + \sum_i \sum_j p_{ij} \ln p_{ij}$$

$$= -\sum_i \left[\sum_j p_{ij} \ln\left(\sum_j p_{ij}\right)\right] - \sum_j \left[\sum_i p_{ij} \ln\left(\sum_i p_{ij}\right)\right] + \sum_i \sum_j p_{ij} \ln p_{ij}$$

$$= \sum_i \sum_j \left[p_{ij}\left(\ln p_{ij} - \ln\left(\sum_j p_{ij}\right) - \ln\left(\sum_i p_{ij}\right)\right)\right]$$

$$= \sum_i \sum_j p_{ij} \ln\left(\frac{p_{ij}}{p_i^{(f)} p_j^{(g)}}\right)$$

This is another way of writing mutual information.

Another Problem

As it turns out, MI decreases with more background. More background creates a higher peak in the histogram at (0,0).

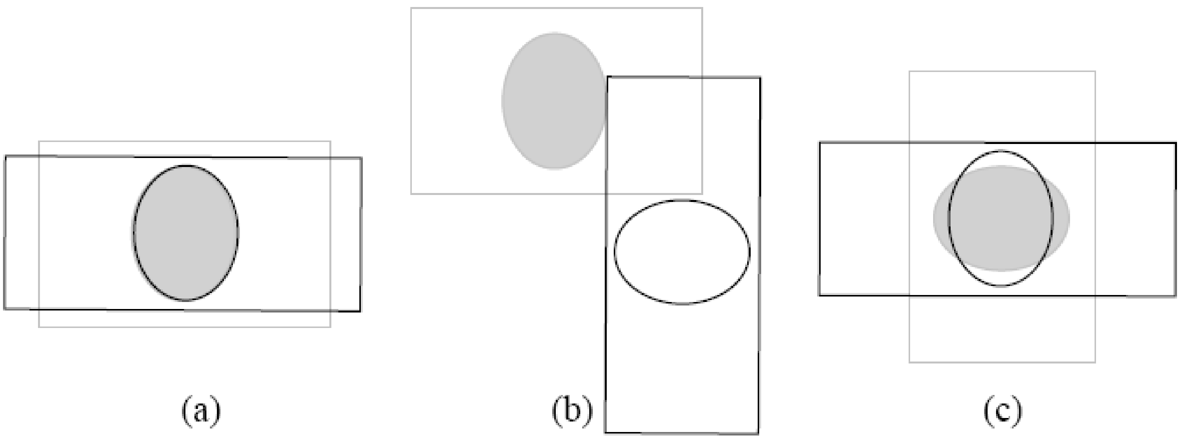


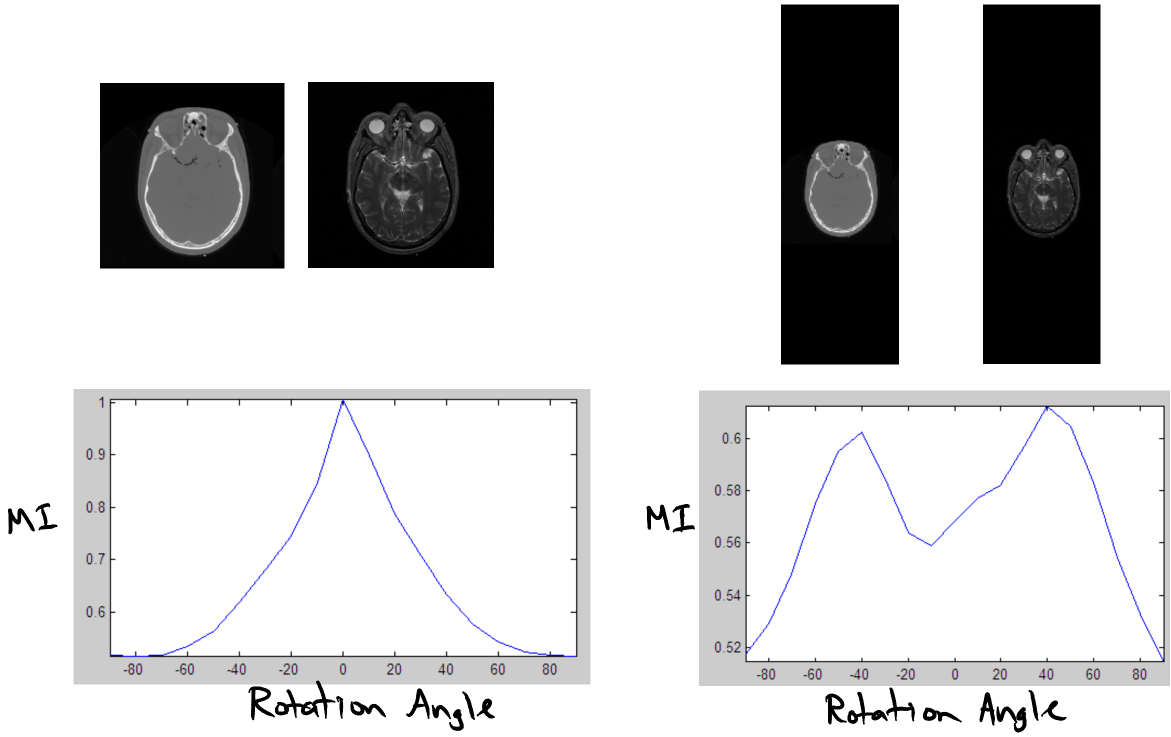
Figure 2.8: Alignment (a) is the correct alignment. However, the entropy of the joint histogram is smaller for (b) because it is calculated only on the overlapped portion. Alignment (c) shows how maximizing mutual information can be inappropriate, since including less background in the overlap increases mutual information. (Adapted with permission from Studholme *et al.* [86])

(from Jeff Orchard's PhD thesis)

Contrast (a) and (c):

(a) has more background, so H(f) and H(g) are lower.
(c) has less background, so H(f) and H(g) are higher, but the decrease in -H(f,g) is no enough to counteract H(f) and H(g).

(truly exceptional Matlab demo)



Normalized Mutual Information

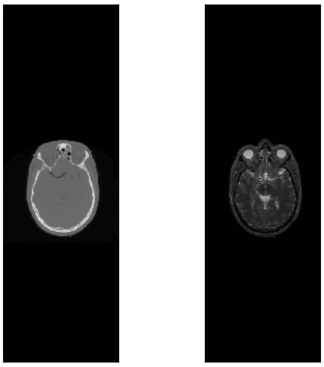
To combat the problem of the amount of background influencing the cost, we have Normalized MI (NMI). There are various forms, including,

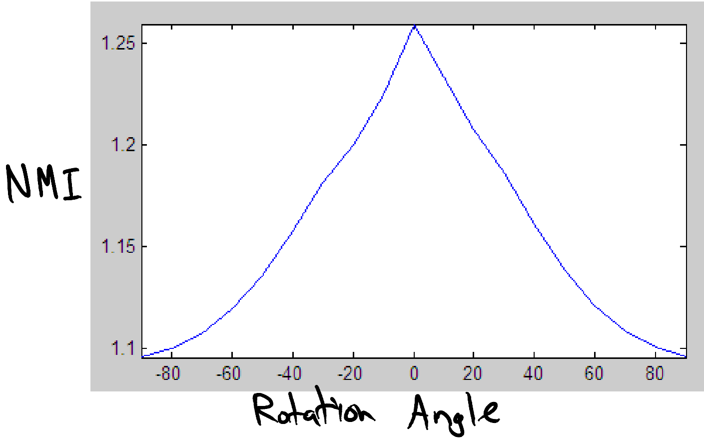
$$NMI(f,g) = \frac{H(f) + H(g)}{H(f,g)}$$

Or, equivalently,

$$NMI(f,g) = \frac{H(f) + H(g) - H(f,g)}{H(f,g)} = \frac{I(f,g)}{H(f,g)}$$

(totally awesome Matlab demo)





MI and NMI are routinely used to register images from different modalities. They constitute the current state-of-the-art.