Mutual Information for Image Registration (Notes G. Gerig)

Excellent descriptions of applications can be found in the following papers [1, 2]. The theory is explained in the textbook [3].

References

[1] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, *Multi-modality image registration by maximization of mutual information*, IEEE TMI, 16(2):187-198, 1997

[2] Wells et al. *Multi-Modal Volume Registration by Maximization of Mutual Information*, Medical Image Analysis, 1(1):35-51 (1996)

[3] Papoulis, A.. Probability, Random Variables, and Stochastic Processes. McGraw-Hill, Inc., third edition, 1991

General Definition

Entropy and joint entropy of a random variable x:

$$
\begin{aligned}
h(x) &= -\int p(x)\ln p(x)dx \\
h(x,y) &= -\int\int p(x,y)\ln p(x,y)\; dxdy
\end{aligned}
$$

Entropy can be interpreted as a measure of uncertainty, variability, and complexity.

Mutual information:

$$
I(x,y) = h(x) + h(y) - h(x,y)
$$

Application: Registration of two volume datasets

We follow the notation of [2]:

$$
I(u(x), v(T(x)) = h(u(x)) + h(v(T(x))) - h(u(x), v(T(x)))
$$

where the volume of the reference volume is denoted as $u(x)$ and the volume of the test volume as $v(x)$. T is a transformation from the coordinate frame of the reference volume to the test volume and x the coordinates of the voxel.

The mutual information can be explained as follows (citing [2]): The mutual information defined in the equation above has three components. The first term on the right is the entropy in the reference volume. The second term is the entropy of the part of the test volume into which the reference volume projects. It encourages transformations that project u into complex parts of v. The third term, the (negative) joint entropy of u and v, contributes when u and v are *functionally related*. The negative joint entropy encourages transformations where u explains v well. Together the last two terms identify transformations that find complexity and explain it well. This is the essence of mutual information.

Binary Case

Entropy

The pixels of a *binary* segmentation results have two states of nature, i.e. 0 or 1. The probabilities of a pixel $\mathbf{x}$ to be 0 or 1 and the binary entropy can be expressed as follows:

$$\begin{aligned}
P_X(0) &= p \\
P_X(1) &= 1-p \\
H(X) &= -(p\ \log_2(p) + (1-p)\ \log_2(1-p)) \ = \ h(p)
\end{aligned}$$

The binary entropy is 0 for $p=0$ and $p=1$ and has a maximum for p=0.5. The entropy H(X) explains how many bits per pixel are needed to describe the image, if p is estimated over the whole image.

Mutual Entropy

One could calculate an entropy measure between a pair of binary images to tell how much information is given by one image to describe a second one, for example to compare the similarity of two segmentation results.

$$\begin{aligned}
H(X|y=0) &= H(\frac{P(x=1|y=0)}{P(y=0)}) \\
H(X|y=1) &= H(\frac{P(x=1|y=1)}{P(y=1)}) \\
H(X,Y) &= P(y=0) * H(X|y=0) \ + \ P(y=1) * H(X|y=1) \\
\\
I(X,Y) &= H(X) + H(Y) - H(X,Y)
\end{aligned}$$

The mutual entropy $I(X,Y)$ expresses the amount of information given by image Y on image X. A value of 1.0 would reflect perfect agreement and 0.0 complete disagreement between the two images X and Y.

Guido Gerig, April 15, 2010