International Journal of Computer Science Engineering Research and Development (IJCSERD)

ISSN 2248 – 9363(Print)
ISSN 2248 – 9371(Online),
Volume 3, Number 2,April- May (2013), pp: 01-08

IJCSERD

© PRJ PUBLICATION

IMAGE SIMILARITY BASED ON INTENSITY USING MUTUAL INFORMATION

Darshana Mistry[1], Asim Banerjee[2], Aditya Tatu[3]

[1]Asst. Professor, Computer Engineering, Indus Institute of Technology, Ahmedabad, India
[2,3]Professor, Information Communication Technology, DAIICT, Gandhinagar, India

ABSTRACT

In this paper, image similarity is finding based on intensity. Mutual information measures the amount of information that one image information contains about image information based on intensity. There are three ways to find mutual information using joint entropy, conditional entropy, and relative entropy. Entropy is find uncertainty of image. If images are more similar to other image, mutual information value is high and vice versa. If images are dissimilar or independent then mutual information is zero. When two images are exact same then mutual information value is maximum(same as entropy of an image).

Keywords: Image similarity, Intensity, Mutual Information, Entropy, joint entropy.

I. INTRODUCTION

Image similarity is used in image registration, content based image retrieval, medical application etc. There are two ways to find similarity of images: intensity based method and feature based method. Sum of squared differences, normalized cross correlation, mutual information methods etc. are used in intensity based method. Distance between corresponding points, similarity metric between features values (i.e. curvature based registration) etc. are used in feature based method.

The mutual information (MI) between two variables is a concept with rootsin information theory and essentially measures the amount of information that one variable contains about another[9]. Put another way, it is the reduction inuncertainty of one variable given that we know the other. The mutual information is measured by entropy, joint entropy, relative entropy and/or conditional entropy.

In information theory, entropy is a measure of the uncertainty in a random variable. In this context, the term usually refers to the Shannon entropy, which quantifies the expected value of the information contained in a message[2]. Entropy is typically measured in bits, nats, or bans.

This field concerns the broadcast of a message from a sender to a receiver. The first attempts to arrive at an information measure of a message focused on telegraph and radio communication, sending Morse code or words. However, picture transmission (television) was already considered in the important paper by Hartley [4]. A drawback of Hartley's measure is that it assumes all symbols (and hence all messages of a given length) are equally likely to occur. In line with Hartley's entropy, we can also view Shannon's entropy as a measure of uncertainty. The difference is that Shannon's measure depends not only on the number of possible messages, but also on the chances of each of the messages occurring.

The Shannon entropy can also be computed for an image, in which case we do not focus on the probabilities of letters or words occurring, but on the distribution of the grey values of the image. A probability distribution of grey values can be estimated by counting the number of times each grey value occurs in the image and dividing those numbers by the total number of occurrences. An image consisting of almost a single intensity will have a low entropy value; it contains very little information. A high entropy value will be yielded by an image with more or less equal quantities of many different intensities, which is an image containing a lot of information[2].

II. ENTROPY, JOINT ENTROPY

Entropy is a measure of unpredictability or information content. Let X be a discrete random variable with alphabet χ and probability mass function$p(x) = \Pr\{X = x\},\ x \epsilon \chi$. We denote the probability mass function by p(x) rather than$p_x(x)$ forconvenience. Thus, p(x) and p(y) refer to two different random variables, and are in fact different probability mass functions, $p_x(x)$ and $p_y(y)$ respectively.

The entropy H(X) of a discrete random variable X is defined by

$$H(X) = -\sum_{x \epsilon \chi} p(x) \log p(x)$$

The log is to the base 2 and entropy is expressed in bits. For example, the entropy of a fair coin toss is 1 bit.

We will use the convention that 0 log 0 = 0, which is easily justified by continuity since $x\log x \to 0$ as $x \to 0$. Thus adding terms ofzero probability does not change the entropy. If the base of the logarithm is b, we will denote the entropy as $H_b(X)$. If the base of the logarithm is e, then the entropy is measured in nuts. Unless otherwise specified, we will take all logarithms to base 2, and hence all the entropies will be measured in bits.

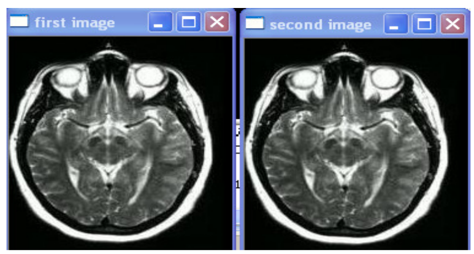
Note that entropy is a functional of the distribution of X. It does not depend on the actual values taken by the random variable X, but only on the probabilities.

The term of$\log\left(\frac{1}{p(x)}\right)$, signifies that the amount of information gained from an event with probability $p(x)$is inversely related to the probability that the event takes place. The information per event is weighted by the probability of occurrence. The resulting entropy term is the average amount of information to be gained from a certain set of events.

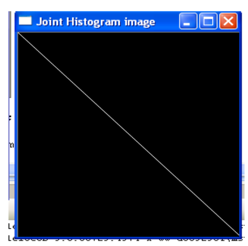
Joint entropy is a measure of the uncertainty associated with a set of variables. The joint entropy H(X, Y) of a pair of discrete random variables (X, Y) with a joint distribution p(x, y) is defined as

$$H(X, Y) = -\sum_{x \epsilon \chi} \sum_{y \,\epsilon\, Y} p(x, y) \log\ p(x, y)$$

Joint entropy is finding in gray level image using joint histogram. Hill et al. [7] proposed an adaption of Woods' measure. They constructed a feature space(or joint histogram/ scattered plot), which is a two dimensional plot showing the combinations of grey values in each of the two images for all corresponding points. The feature space is constructed by counting the number of times a combination of grey values occurs. For each pair of corresponding points (x, y), with x a point in the input image and y is a point in the reference image, the entry (input_image(x), Reference_image(y)) in the feature space on the right is increased. The feature space (or joint histogram) changes as the alignment of the images changes. When the images are correctly registered, corresponding anatomical structures overlap and the joint histogram will show certain clusters for the grey values of those structures(see fig. 1 and 2). Table 1 shows entropy of images and joint entropy of images

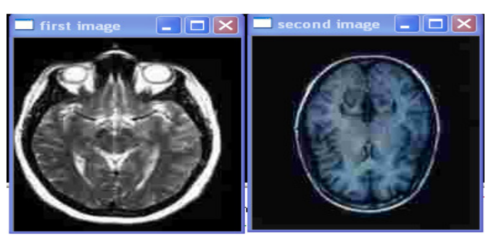


(a) (b)

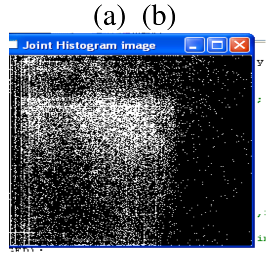


(c)

Fig 1. (a) Input image (b) Reference image, (c) joint histogram of (a) and (b) images



(a) (b)



(c)

Fig. 2. (a) Input image (b) Reference image, (c) joint histogram of (a) and (b) images

TABLE 1: ENTROPY OF IMAGES

Image	Input Image(Entropy)- H(A)	Reference Image(Entropy)- H(B)	Joint Entropy H(A,B)
Fig. 1	6.43535	6.43535	5.391131
Fig. 2	6.35373	4.987587	10.204270

III. CONDITIONAL ENTROPY AND RELATIVE ENTROPY

If *(X,Y)~p(x, y),* then the conditional entropy *H(Y/X)* is defined as

$$H(Y|X) = \sum_{x \in \chi} p(x) H(Y|X = x)$$

$$= -\sum_{x \in \chi} p(x) \sum_{y \epsilon \gamma} p(y|x) \log p(y|x) = -\sum_{x \in \chi} \sum_{y \epsilon \gamma} p(x, y) \log\, p(y|x)$$

The naturalness of the definition of joint entropy and conditional entropy is exhibited by the fact that the entropy of a pair of random variables is the entropy of one plus the conditional entropy of the other.

The relative entropy (Kullback Leibler Distance) is a measure of the distance between two distributions. Is statistics, it arises as an expected logarithm of the likelihood ratio. The relative entropy *D(p||q)* is a measure of the inefficiency of assuming that the distribution is q when the true distance is p. If we knew the true distribution of the random variable, then we could construct a code with average description length H(p). If, instead, we used the code for a distribution q, we would need H(p) + D(p|| q) bits on the average to describe the random variable.

The relative entropy between two probability mass function p(x) and q(x) is defined as

$$D(p||q) = \sum_{x \in \chi} p(x) \log \frac{p(x)}{q(x)}$$

Relative entropy is always non-negative and is zero if and only if p = q.

IV. MUTUAL INFORMATION

Mutual information measures the amount of information that one variable contains about another reference variable. There are three ways to represent of mutual information. All three forms are identical; each can be rewritten into the other two. Each form of definition, however, explains the relation to registration in a different way.

The first form of definition we discuss is the one that best explains the term "mutual information". For two images A and B, mutual information I can be defined as [2]

$$I(A, B) = H(B) - H(B|A) = H(A) - H(A|B)$$

where H(B) is the Shannon entropy of image B, computed on the probability distribution of the grey values. H(B |A)denotes the conditional entropy, which is based on the conditional probabilities p(b | a), the chance of grey value b in image B given that the corresponding voxel in A has grey value a. mutual information is the amount by which the uncertainty about B decreases when A is given: the amount of information A contains about B. Because A and B can be interchanged, I(A,B) is also the amount of information B contains about A.

The second form of definition is most closely related to joint entropy. It is

$$I(A,B) = H(A) + H(B) - H(A,B)$$

This form contains the term –H(A,B), which means that maximizing mutual information is related to minimizing joint entropy. The advantage of mutual information over joint entropy per se, is that it includes the entropies of the separate images. Mutual information and joint entropy are computed for the overlapping parts of the images and the measures are therefore sensitive to the size and the contents of overlap. A problem that can occur when using joint entropy on its own is that low values (normally associated with a high degree of alignment) can be found for complete miss registrations. For example, when transforming one image to such an extent that only an area of background overlaps for the two images, the joint histogram will be very sharp. There is only one peak, that of background. Mutual information is better equipped to avoid such problems, because it includes the marginal entropies H(A) and H(B).These will have low values when the overlapping part of the images contains only background and high values when it contains anatomical structure. The marginal entropies will thus balance the measure somewhat by penalizing for transformations that decrease the amount of information in the separate images. Consequently, mutual information is less sensitive to overlap than joint entropy, although not completely immune.

The final form of definition we discuss is related to the Kullback-Leibler distance(Relative Entropy). The mutual information of image A and B is defined as

$$I(A,B) = \sum_{a,b} p(a,b) \log \frac{p(a,b)}{p(a)p(b)}$$

The interpretation of this form is that it measures the distance between the joint distribution of the images' grey values p(a, b) and the joint distribution in case of independence of the images, p(a)p(b). It is a measure of dependence between two images. The assumption is that there is maximal dependence between the grey values of the images when they are correctly aligned. Miss registration will result in a decrease in the measure.

V. IMPLEMENTATION

OpenCV 2.3(Open source for Computer vision) and Microsoft Visual studio 2010 are used to implement entropy based finding similarity between of two images. Brain image data set 512 x 512 is used for implementation.

Fig. 3 to Fig. 8 denotes joint histogram of an images and Table 2 denotes entropy, joint entropy and mutual information of an images.

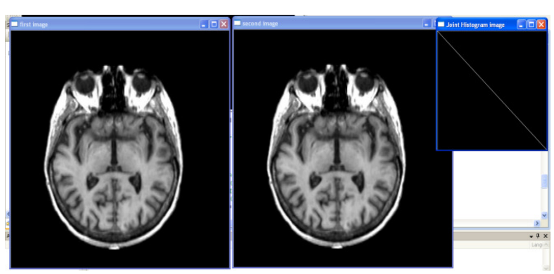


Fig 3. (a) Input image(CT image) (b) Reerence image(CT image), (c) joint histogram of (a) and (b) images

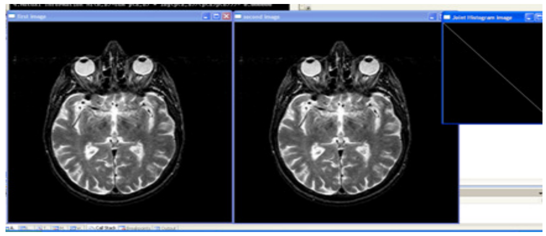


Fig 4. (a) Input image(MRI image (b) Reference image(MRI image), (c) joint histogram of (a) and (b) images

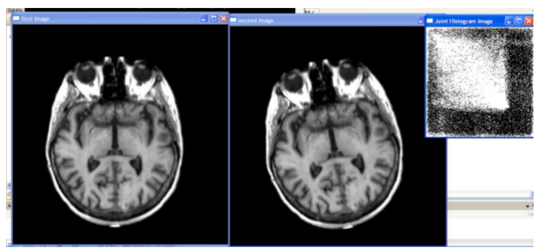


Fig 5. (a) Input image (CT image) (b) Reference image(CT image rotate with 10 degree), (c) joint histogram of (a) and (b) images

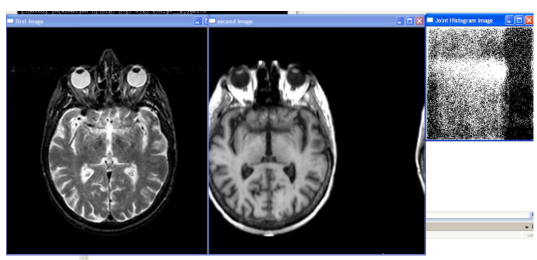


Fig 6. (a) Input image(MRI image) (b) Reference image(CT image-transport image), (c) joint histogram of (a) and (b) images

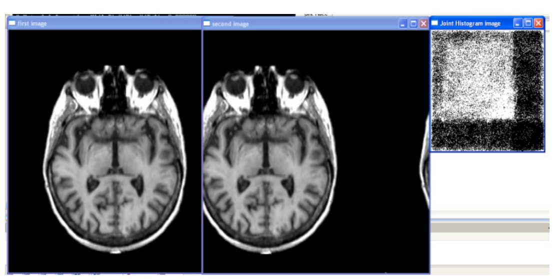


Fig 7 (a) Input image(CT image) (b) Reference image (CT image-transport image), (c) joint histogram of (a) and (b) images

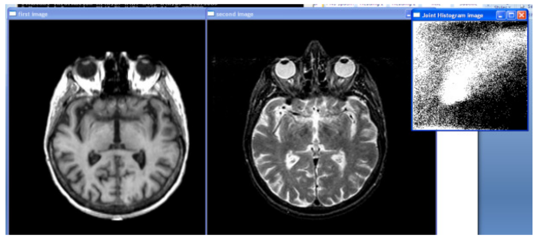


Fig 8 (a) Input image(CT image) (b) Reference image(MRI image), (c) joint histogram of (a) and (b) images

TABLE 2: ENTROPY, MUTUAL INFORMATION OF IMAGES

Image	Input Image (Entropy)-H(A)	Reference Image (Entropy)-H(B)	Joint Entropy H(A,B)	Mutual Information I(A:B)
Fig. 3	4.058281	4.058281	4.058281	4.058281
Fig. 4	4.089286	4.089286	4.089286	4.089286
Fig. 5	4.058283	4.138616	7.137251	1.051646
Fig. 6	4.08928	4.065977	7.853091	0.302172
Fig. 7	4.05828	4.065977	7.837682	0.286576
Fig. 8	4.058283	4.089286	7.053765	1.093803

If two images are same then joint histogram(scatter plot) of two images are diagonal line from top left corner to bottom right corner which denoted that two images are exact similar. If are not totally similar, there are distortions in scatter plot (see fig. 3 to 8).
Here, mutual information is finding based on joint entropy. If two images are exact same then mutual information is maximum (entropy value of an image). If two image are totally independent(do not same any point) then mutual information is minimum value(zero). If mutual information value is high then image similarity is high. If mutual information value is less, image similarity is less.

CONCLUSION

Entropy of an image is measured through distribution of the grey values of the image. A probability distribution of grey values can be estimated by counting the number of times each grey value occurs in the image and dividing those numbers by the total number of occurrences. So entropy find uncertainty of an image. Joint histogram denotes how images are more/less similar. Mutual information measures the amount of information that one image. If mutual information value is maximum then tow images are totally similar. If it is zero then two images are totally dissimilar. If mutual information value is high the image similarity is high and vice versa.

ACKNOWLEDGMENT

I am thankful to Dr Asim Banerjee who inspired and helped me to do my dissertation work on image similarity which is used in image registration for continuous guidelines. I am also very thankful to Dr Aditya Tatu who help me to solve fundamental knowledge in information theory.

REFERENCE

[1] R. M. Gray, "Entropy and Information Theory", Springr-Verlag, New York, 2009.
[2] J. P.W.Pluim, J.B. A. Maintz, and M. A. Viergever, "Mutual Information based registration of medical images: a survey", IEEE Transactions on Medical Images, Vol.-XX, Issues-Y, 2003.
[3] T. Carter, "An introduction to information theory and entropy", Complex System Summer School, June 2011.
[4] R. V. L. Hartley, "Transmission of information," Bell System Technical Journal, vol. 7, pp. 535–563, 1928.
[5] T. M. Cover, J. A. Thomas, "Elements of Information Theory", John Wiley & Sons, pp 12-49, 1991.
[6] D. L. G. Hill, D. J. Hawkes, N. A. Harrison, and C. F. Ruff, "A strategy for automated multimodality image registration incorporating anatomical knowledge and imager characteristics," in Information Processing in Medical Imaging, H. H. Barrett and A. F. Gmitro, Eds. 1993, vol. 687 of Lecture Notes in Computer Science, pp. 182–196, Springer-Verlag, Berlin.
[7] D.Mistry, A. Banerjee, A. Tatu, "Image Similarity based on Joint Entropy (Joint Histogram)", will be published in May 2013.
[8] T.M. Cover, J. A. Thomas, "Entropy, Relative Entropy and Mutual Information", Chapter 2, Elements of Information Theory, John Wiley & Sons.
[9] D.B. Russsakoff, C. Tomasi, T. Rohlfing, C. R. Maurer, "Image Similarity Using Mutual Information ofRegions", Springer-Verlag Berlin- Heidelberg, pp 596-607, 2004.