COMPARATIVE ANALYSIS OF INFORMATION MEASURES FOR DIGITAL IMAGE PROCESSING

J.A.R. Blais and M. Boulianne
Department of Surveying Engineering
The University of Calgary
Calgary, Alberta, T2N 1N4

Commission III

Automation in digital image processing and related applications requires information measures for decision and quality control purposes. Several information measures and distance functions have recently been suggested for estimating quantitative information contents and error bounds in various application areas. Using information theoretic and practical considerations such as robustness, sensitivity to noise and computational efficiency, some of the proposed measures are compared and analyzed, taking into account their applicability in digital image processing. Simulations with pseudo-random noise and experimental projects involving adaptive filtering and template matching applications are reported on along with recommendations for further experimentation and analysis.

1. INTRODUCTION

Digital information requires appropriate methodology for quantization, processing and analysis. Depending on the nature and representation of the digital information, the requirements can be very different and the measuring techniques adaptive and reliable. Research and development in this field are very much in demand in view of the digital information replacing much of the conventional analog information in computerized and automated mapping as well as in other applications. Information measures and distance functions are required whenever digital information needs to be measured or assessed in comparison to alternative representations.

The transition from photographs to digital image data has had profound implications for researchers and users of the information. On the positive side, the change implies practically limitless possibilities for data processing, communication and analysis. With the current computer and related technology, the analytical tools at one's disposal are extremely varied in terms of complexity, reliability and efficiency.

From another perspective, an average aerial photograph easily yields well over one million pixels at the usual photogrammetric resolution. One obvious implication is the enormous amounts of data that have to be dealt with in any data processing operation. The problem is compounded when considering tasks such as point and entity identification, image correlation, stereocorrelation and image refinement procedures using only the digital data available. In other words, without access to the conventional photographic images, the usual tasks are much more challenging.

Appropriate quantitative measures of information contents and distance separations are required for automating most image processing operations. Optimal algorithms often result from the application of the principle of maximum information or entropy which is based on information theory. The current technical literature on image processing and analysis contains numerous references to information theoretic results which provide the mathematical foundations [e.g., Shiozacki, 1986; Geman and Geman, 1984]. Decision problems have to be formulated using objective or discrimination functions in terms of information measures and distance functions as other approaches usually fail to have the necessary generality and reliability.

Following a general introduction to information measures and distance functions, their principal characteristics will be briefly discussed in view of the intended applications. Test results from simulated and real data sets will also be discussed with some practical considerations. Concluding remarks will then be included with recommendations for further research and experimentation on some outstanding problems in this field.

2. INFORMATION MEASURES

The concept of information in mathematics goes back to the definition of probability by Jacob Bernoulli [1713] in the "Ars Conjectandi". In mathematical terms, information, or uncertainty, is an appropriate real-valued function of the probability of occurrence of some events [e.g., Guiasu, 1977; Martin and England, 1981; Blais, 1987 and 1988]. Depending upon the application requirements, different functions may be used but the fundamental characteristics of information and the associated measures are respected. The following discussion will be limited to the Shannon-Wiener entropy, the corresponding relative entropy, the average quadratic and cubic entropies which appear to have advantageous properties for image processing applications. The introduction of weights and fuzzy components will also be briefly mentioned in view of the possible applications.

The Shannon-Wiener entropy for a digital image with grey-level frequencies f_1, f_2, ..., f_n is usually defined as

$$H[p] = H[p_1, p_2, \ldots, p_n] = - k \sum_{i=1}^{n} p_i \log p_i$$

where the probabilities $p_i = f_i/N$ with $N = \Sigma f_i$ and some appropriate positive constant k. Such entropy H[p] is a measure of the information contained in the digital image considered as a realization of a stochastic field.

Alternatively, H[p] can be considered as a measure of the uncertainty in the digital image. The maximum possible entropy H[p] is easily verified to correspond to a uniform distribution $p_i = 1/n$ for all the n grey levels. Any additional constraint imposed on the realization of the stochastic field will tend to lower the entropy of the image.

The corresponding relative entropy associated with the same digital image is defined as

$$H[p;q] = H[p_1,p_2, \ldots,p_n; q_1,q_2, \ldots,q_n]$$

$$= -k \sum_{i=1}^{n} p_i \log p_i/q_i$$

where again, the probabilities $p_i = f_i/N$ for the grey-level frequencies f_1, f_2, ..., f_n with $N = \Sigma f_i$ and some positive constant k in relation to some respective background probabilities q_1, q_2, ..., q_n. It can be shown that with a constant or uniform background, this relative entropy is identical to the Shannon-Wiener entropy. The negative of H[p; q] has been called by various names such as information gain, directed divergence, discrimination information and the Kullback-Leibler information measure.

Average quadratic and cubic entropies associated with the same digital image can also be defined as

$$H_2[p] = H_2[p_1,p_2, \ldots,p_n] = 1 - \sum_{i=1}^{n} p_i^2$$

and

$$H_3[p] = H_3[p_1,p_2, \ldots,p_n] = 1 - \sum_{i=1}^{n} p_i^3$$

respectively. It is worth noting that the average quadratic entropy results from the substitution

$$- \log p_i \rightarrow 1 - p_i$$

in the Shannon-Wiener entropy formula. Similarly, the average cubic entropy results from substituting

$$- \log p_i \rightarrow (1-p_i) - \frac{1}{2} (1-p_i)^2$$

as can easily be verified [Chen, 1977].

In various application areas, the different frequency contributions have to be weighted to take into account any nonuniform sensitivity or resolution for the individual spectral bands. In pattern recognition applications, weights can be associated with the different characteristics to reflect the uncertainty levels associated with the different aspects of the problem [e.g., Guiasu, 1977]. The Shannon-Wiener entropy definition has been generalized to the following:

$$H_W[p] = - k \sum_{i=1}^{n} w_i \, p_i \log p_i$$

for the previously described digital image. The other information measures can be generalized in an analogous manner.

For some applications such as in cluster analysis, the cluster or entity boundaries are not always clearly defined by zones with negligible thickness. In other words, uncertainty is associated with the boundary definition of objects which are to be analyzed for information content or separation distance. Using the Shannon function

$$S(x) = - x \log x - (1-x) \log (1-x)$$

for x in [0, 1] which attains its maximum at x = 1/2, one can define the fuzzy entropy, or measure of uncertainty in the frequency class definitions, as

$$H_S[p] = k' \sum_{i=1}^{n} p_i \, S(\phi(x_i))$$

for the previous digital image situation with $\phi(x_i)$ denoting the characteristic function of the i-th interval x_i in the dynamic range, and some appropriate constant k' [e.g., DeLuca and Termini, 1972]. The corresponding total entropy is simply the sum of the regular entropy and the fuzzy entropy as these are clearly independent.

3. DISTANCE FUNCTIONS

A distance function is a non-negative real-valued function of two variables such that the following metric conditions are satisfied:

(a) $d(x,x) = 0$ for all x,
(b) $d(x,y) > 0$ unless $x = y$,
(c) $d(x,y) + d(y,z) \geq d(x,z)$,

where the last condition is often referred to as the triangle inequality. An example of a trivial distance is the Kronecker distance with $d(x,y) = 1$ except for $d(x,x) = 0$ for all x and y. Examples of more useful distance functions are the Euclidean and non-Euclidean distances in real physical space.

An entropy measure can be converted into a distance function d(X,Y) between the entities X and Y as follows:

$$d(X,Y) = 2H(X,Y) - H(X) - H(Y)$$

where $H(X,Y)$ is the joint entropy of X and Y or the entropy in terms of the joint probabilities of X and Y. In other words, the previous expression for $d(X,Y)$ can be rewritten as

$$d(X,Y) = H(X,Y) - W(X,Y)$$

where the so-called Watanabe measure

$$W(X,Y) = H(X) + H(Y) - H(X,Y)$$

quantifies the interdependence between X and Y. It can easily be verified that $W(X,Y) \geqq 0$, with equality in the case of independence. The latter has been used as a cohesion measure for classification schemes in different application areas [Watanabe, 1969].

The converse proposition for arbitrary distance functions is obviously not true as most distance functions are quite unrelated to information measures. In terms of norms, a distance function implies a norm for vector quantities in the usual manner. Furthermore, entropy norms are known to be intermediate between the variation and the uniform norms [Dabrowski, 1984; Korenblum, 1985].

4. THEORETICAL CONSIDERATIONS

The error probability is normally used to characterize an estimation or decision problem. In other words, error probability means the minimum probability of a false decision for a given digital image. As the error probability is only a particular value of the risk function, the entropy remains most valuable as a measure of the uncertainty of the information. A number of results on the relation between the conditional entropy and the corresponding error probability can be found in Kovalevsky [1980; pp. 78 - 90].

The previous discussion of information measures and distance functions made no references to any probability density functions or distributions. This approach can therefore be qualified as robust in theory and also in practice as will become evident in the sequel. However, information-theoretic considerations can be used to decide on most appropriate distributions for the available observational information [e.g., Blais, 1987]. For instance, no information implies a uniform distribution, first-moment information implies a Gibbs distribution, second-moment information implies a Gaussian distribution, and so on.

One very important consideration for the information measure or entropy is its sensitivity to additive noise. According to a theorem of Aczel and Daroczy [1963 and 1975] for average entropies which are "small for small probabilities", only the Shannon-Wiener entropy and the so-called Renyi entropies of order α with $\alpha > 0$, i.e.,

$$H^{\alpha}[p_1,p_2, \ldots,p_n] = (1 - \alpha)^{-1} \log \sum_{i=1}^{n} p_i^{\alpha}$$

are additive. Recalling the definitions of average quadratic and cubic entropies, it is easy to verify that

$$H^2[p_1,p_2, \ldots,p_n] = - \log \{1 - H_2[p_1,p_2, \ldots,p_n]\}$$

and

$$H^3[p_1,p_2, \ldots,p_n] = - 2^{-1} \log \{1 - H_3[p_1,p_2, \ldots,p_n]\}.$$

When adding noise to a digital image, linear additivity for the Renyi entropies of order 2 and 3 implies exponential growth for the average quadratic and cubic entropies. The implication is that the average quadratic and cubic entropies are much more sensitive to additive noise than the Shannon-Wiener entropy.

5. PRACTICAL CONSIDERATIONS

For image processing and related applications, it is also imperative to have information measures which are coordinate system independent. In other words, any rotation and translation of a digital image should not affect the information content of the image. The relative entropy measure is easily verified to be coordinate-system independent but the other previously discussed information measures are not. Furthermore, the relative entropy measure is known to be unaffected by any orthogonal transformation of the image which is clearly an important property for image processing applications [Andrews, 1970].

For image correlation and matching applications, the problem often reduces to shape matching, that is matching a model or pattern from a knowledge base with a shape extracted from a digital image. In its simplest terms, this means deciding between a linear, quadratic, cubic, ..., regression curve to approximate a number of data values. Such model identification problems have been solved in information-theoretic terms using the Akaike information criterion which has proven to be quite reliable in very different application contexts [Akaike, 1974, 1978; Sakamoto et al., 1986; Blais, 1987, 1988].

6. EXPERIMENTAL RESULTS

One practical experiment consisted in the evaluation of the performance of a series of averaging median filters for Landsat images. These filters are reported to adequately remove the coherent noise while preserving the fine details in the digital image [Heinonen et al., 1985; Nieminen et al., 1987; Vassiliou et al., 1988]. With the averaging median filters, the median is taken from the centre pixel value and the results from averaging the pixel values in specified directions. The principal objective of the experiment was to select the filter configuration that gives the best results in a manner as objectively as possible.

Three specific averaging median filter configurations as shown in Figure 1 and a standard median filter were applied to four Landsat images which displayed totally different characteristics. The first one was collected over the Kananaskis Valley, west of Calgary, Alberta, Canada. This is an area of rugged terrain near the eastern edge of the Rocky Mountains, including peaks, valleys and lakes. The elevation varies from 1300 m to 3500 m above mean sea level. The area is partly covered by forest. The second image shows the city of Calgary and its suburbs. It includes both urban and agricultural regions. Unlike the previous area the relief is moderate. Both images cover approximately the same area of 100 square kilometres, and they had previously been corrected for the systematic errors. The third and fourth images are the same as the first and second except for the addition of simulated white noise. The variances of the normalized noise were two and five square grey levels for the third and fourth images, respectively.

As shown in Table 1, the four previously discussed measures of Shannon-Wiener, relative, average quadratic and cubic entropies were used to evaluate each of the images and their respective filtered versions. The maximum values of the Shannon-Wiener and relative entropies agree very well as they have been shown to agree with some thirty-three different averaging median filter configurations applied to the same two Landsat images [Vassiliou et al., 1988]. In the latter reference, the optimal filter results were confirmed visually and spectrally (i.e., using spectral analysis). In the case of the average quadratic and cubic entropies, the optimal filter results are slightly different as more emphasis is implicitly given to smaller normalized frequencies. These results are also quite consistent.

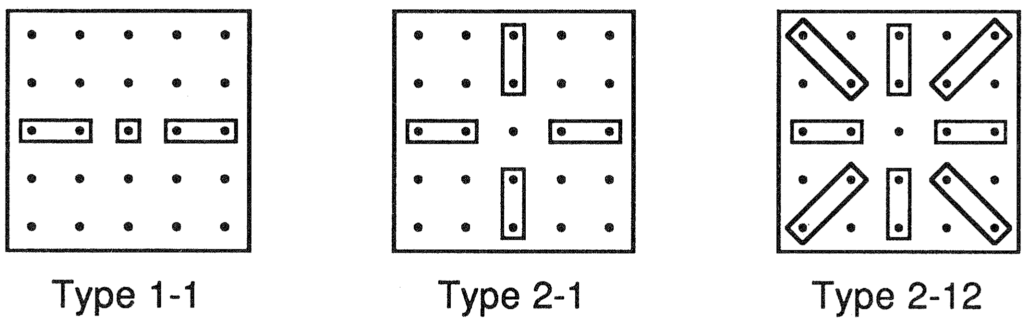

Figure 1: Averaging Median Filter Configurations

TABLE 1
Experimental Entropy Results

Test Image	Filter Configuration	Shannon-Wiener Entropy	Relative Entropy	Average Quadratic Entropy	Average Cubic
1	Original	4.55841	--	0.93787	0.99271
	1	4.53268	-0.00359	0.93739	0.99270
	2	4.54131	-0.00213	0.93746	0.99266
	3	4.54448	-0.00156	0.93740	0.99261
	4	4.37825	-0.05737	0.92154	0.98720
2	Original	5.95398	--	0.97667	0.99918
	1	6.01920	0.03006	0.97827	0.99934
	2	6.01514	0.01872	0.97766	0.99927
	3	6.00163	0.01356	0.97725	0.99923
	4	5.66144	-0.04594	0.97102	0.99876
3	Original	4.80418	--	0.95708	0.99758
	1	4.77943	-0.00360	0.95671	0.99758
	2	4.78920	-0.00171	0.95681	0.99756
	3	4.79317	-0.00133	0.95692	0.99757
	4	4.72293	-0.01591	0.95493	0.99731
4	Original	5.00152	--	0.96470	0.99857
	1	4.97898	-0.00315	0.96428	0.99854
	2	4.98808	-0.00175	0.96445	0.99855
	3	4.99174	-0.00116	0.96454	0.99856
	4	4.91716	-0.01804	0.96241	0.99835

Another series of experiments consisted in the selection of best matching strategies in the transfer of control entities from aerial triangulation diapositives to satellite imagery. In these experiments, small patches around the entities have been digitized using a digital camera and image matching techniques have been used for the correlation process. Three algorithms have so far been implemented, i.e., statistical correlation, absolute variation and adaptive least squares. This last approach seems the most appropriate in view of the geometrical differences between aerial photography and satellite imagery [Chapman et al., 1988]. The decision for a best match requires objective measures of cohesion between the corresponding images of the control entity or template [e.g., Guiasu, 1977: Chapter 20]. Using different characteristics of the images, the optimal matching can be carried out using all the available information. The maximum entropy approach to such image matching problems is currently under investigation.

7. CONCLUDING REMARKS

Objective decisions related to the problems of digital image processing require information measures and distance functions as most conventional statistics are unapplicable or unreliable. For instance, optimal filtering, template and pattern matching in digital imagery cannot be automated without a reliable adaptive methodology that can be implemented in a production environment.

Among the necessary characteristics of information measures and distance functions for such applications are coordinate system independence, limited sensitivity to noise, adaptability and reliability. The information measures should also be independent of any semantics for digital image applications. From the limited theoretical and practical investigations undertaken in this research, it appears that the relative entropy is the only information measure that satisfies these requirements.

Further research is clearly warranted on these questions as automation in digital image processing and pattern matching is very much dependent on the results. In digital photogrammetry, computer stereovision, information systems and other fields, the potential applications are unlimited.

8. ACKNOWLEDGEMENTS

This research was sponsored by the Natural Sciences and Engineering Research Council operating grant to the first author and post-doctoral fellowship to the second author. The experimental work was also partially supported by the Alberta Bureau of Surveying and Mapping, Department of Forestry, Lands and Wildlife, and a Pilot Project grant from The University of Calgary.

9. REFERENCES

Aczel, J. and Z. Daroczy [1975]: On measures of Information and Their Characterizations. Academic Press.

Aczel, J. and Z. Daroczy [1963]: Sur la caracterisation aximomatique des entropies d'ordres positifs, y comprise l'entropie de Shannon. C.R. Acad. Sci. Paris, Vol. 257, pp. 1581 - 1584.

Akaike, H. [1978]: A New Look at the Bayes Procedure. Biometrika, Vol. 65, pp. 53 - 59.

Akaike, H. [1974]: A New Look at the Satistical Model Identification. IEEE Trans. Autom. Control, AC-19, pp. 716 - 723.

Andrews, H.C. [1970]: Computer Techniques in Image Processing. Academic Press.

Bernoulli, J. [1713]: Ars Conjectandi. Paris.

Blais, J.A.R. [1988]: Information Theory and Applications. Under preparation.

Blais, J.A.R. [1987]: Information Theory and Optimal Estimation. Manuscripta Geodaetica, Vol. 12, pp. 238 - 244.

Chapman, M.A., M. Boulianne, A. Tam and J.A.R. Blais [1988]: Satellite Image Rectification Using Control Transferred from Digitized Aerial Triangulation Diapositives. Congress of the International Society of Photogrammetry and Remote Sensing, Commission IV, Working Group 1, in Kyoto, Japan.

Chen, C.H. [1977]: On the Use of Distance and Information Measures in Pattern Recognition and Applications. Pattern Recognition Theory and Applications, edited by K.S. Fu and A.B. Whinston, Noordhoff International Publishing, Leyden, pp. 45 - 60.

Dabrowski, R. [1984]: Probability Measure Representation of Norms Associated with the Notion of Entropy. Proc. Amer. Math. Soc., Vol. 90, pp. 263 - 268.

DeLuca, A. and S. Termini [1972]: A Definition of a Nonprobabilistic Entropy in the Setting of Fuzzy Sets Theory. Information and Control, Vol. 20, pp. 301 - 312.

Geman, S. and D. Geman [1984]: Stochastic Relaxation, Gibbs Distributions and Bayesian Restoration of Images. IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. PAMI-6, No. 6, pp. 721 - 741.

Guiasu, S. [1977]: Information Theory with Applications. McGraw-Hill.

Heinonen, P. and P. Neuvo [1985]: Smoothed Median Filters with FIR Substructures. Proceedings of the 1985 IEEE ICASSP Conference, pp. 49 - 52.

Korenblum, B. [1985]: On a Class of Banach Spaces Associated with the Notion of Entropy. Trans. Amer. Math. Soc., Vol. 290, pp. 527 - 553.

Kovalevsky, V.A. [1980]: Image Pattern Recognition. Springer Verlag.

Martin, N.F.G. and J.W. England [1981]: Mathematical Theory of Entropy. Addison-Wesley.

Nieminen, A., P. Heinonen and Y. Neuvo [1987]: A New Class of Detail-Preserving Filters for Image Processing. IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 9, No. 1, pp. 74 - 90.

Sakamoto, Y., M. Ishiguro and G. Kitagawa [1986]: Akaike Information Criterion Statistics. KTK Scientific Publishers, Tokyo.

Shiozacki, A. [1986]: Edge Extraction Using Entropy Operator. Computer Vision, Graphics and Image Processing, Vol. 36, pp. 1 - 9.

Vassiliou, A.A., M. Boulianne and J.A.R. Blais [1988]: On the Application of Averaging Median Filters in Remote Sensing. IEEE Transactions on Geoscience and Remote Sensing, in press.

Watanabe, S. (ed.) [1969]: Methodologies of Pattern Recognition. Academic Press.