Image Processing. A new Approach via Informational Entropy and Informational Divergence of non Random Functions

Guy Jumarie

Department of Mathematics, University of Quebec, Montreal Qc, Canada

By combining a maximum conditional entropy principle with a basic equation of (Shannon) information theory, one can obtain a meaningful concept of informational entropy of non random functions. When this entropy is applied to the brightness function of an image, one so has at hand a new tool which provides new approaches to some image processing problems, such as, for instance, image representation, image compression and image similarity. As a by-product, to some extent, this new modelling provides a support to the so-called monkey model of image entropy. But while the latter involves the brightness itself, here, the entropy of the brightness function is expressed in terms of the contrast of the brightness instead of the brightness itself. In this framework, a new concept of informational divergence of an image is obtained, which could be of help in image analysis.

Keywords: information, entropy of non random functions, image entropy, brightness, brightness contrast, image compression, image similarity, image encoding, Shannon information, Renyi information, image cross-entropy

1. Introduction

Basically, image processing is an information processing. Digitization, compression, restoration, segmentation and even pattern recognition can be thought of as problems dealing with information; and as a result, information theory (IT in the following) should be quite relevant. Strictly speaking, IT is not essential (it is not a prerequisite) to address these questions. For instance, let us consider image compression (see Rosenfeld et al, 1982). Several techniques are available. There is the Karhunen–Loeve compression (Loeve, 1948); Fourier, Hadamar and cosine compression (Chen et al, 1977; Katajima, 1980); predictive compression (Cutler, 1952; Graham, 1958); block truncation compression (Mitchell et al, 1978; Delp et al, 1979); and none of them explicitly refers to IT, despite the fact that the latter are more or less transparent in their respective contents.

As a matter of fact, the basic issue in introducing IT in image processing is the randomization of the problem. In substance, Shannon's information is defined in fields of probabilities, and using IT in image processing is easy if we are dealing with a family of images; but the matter is much less obvious if we have only one picture at hand.

Much earlier (see for instance Pratt, 1978) came the idea of considering the brightness function $b(x, y)$ of a picture as more or less equivalent to a bi-variate probability density, and to process it as if it were that. At the beginning, this comparison was purely formal. But later, this point of view was supported by using a physical model (referred to as the monkey model) in which photons are randomly thrown on a set of empty pixels (this is recalled in the next section), and as a result, it can now be considered that this approach has a sound support.

Nevertheless, this monkey model of image entropy gives rise to the following problem: it works as if the brightnesses of neighbouring pixels were uncorrelated (or independent), what is rather troublesome on the surface. At first glance, a proper measure of the amount of uncertainty involved in a picture should take into account the mutual dependence between neighbouring pixels, and this requirement is not met in the monkey model.

Our main purpose herein is to once more consider the problem of measuring the amount of uncertainty (which is equal to the amount of information in some instances!) contained in an image; and more exactly, we shall show how one can obtain a new model of image entropy which explicitly takes into account the pixel correlation. This will be done by merely using a principle of maximum conditional entropy combined with a basic equation of IT.

The paper is organized as follows. In the next section, we shall recall the definition of image entropy in terms of brightness function, and its derivation via the monkey model of throwing photons. Then we shall obtain a model of entropy of non random functions as a direct consequence of the basic equation of information theory. Applying this definition to brightness functions will allow us to define a new model of image entropy, which explicitly refers to the mutual dependence of pixels, via the brightness contrast. Then we shall see how this entropy provides new approaches to some problems of image processing. Lastly, by combining the monkey model with the entropy of markovian processes, we shall obtain a concept of informational divergence of image, which is directly related to the complexity of the latter.

2. Monkey Model of Image Entropy

2.1. Main Definitions

Analogously with the entropy of random variables, it has been suggested to define the entropy of an image as follows.

Definition 2.1. Let be given an image P (P for picture) defined on the domain $(x,y) \in \Omega \subset \mathbf{R}^2$ and characterized by the brightness $b(x_i, y_i)$ for all (i,j). Its image entropy is defined by

$$H(P) = -\sum_{i,j} b(x_i,y_i)\ln b(x_i,y_i), \ (x,y) \in \Omega. \tag{2.1}$$

In the continuous case, one will set

$$H(P) = -\int_\Omega b(x,y)\ln b(x,y)\,dx\,dy. \tag{2.2}$$

This definition requires a few comments. Indeed, given a random discrete scalar valued random variable X with the probability distribution $p_i, i = 1,\ldots,n$; its entropy is given by

$$H(X) = -\sum_{i=1}^{n} p_i \ln p_i; \tag{2.3}$$

but if we consider the entropy defined by the incomplete distribution $(p_j,\ldots,p_{j+m}$; then its expression is (see for instance Aczel et al, 1975)

$$H(X; x_j \le X \le X_{j+m}) = -\sum_{i=j}^{j+m} p_i \ln p_i / \sum_{i=j}^{j+m} p_i. \tag{2.4}$$

In a similar manner, for a continuous random variable with the probability density $p(x)$, the corresponding mean value of uncertainty in the interval (a,b) is

$$H(X; a \le X \le b) = -\int_a^b p(x)\ln p(x)\,dx \Big/ \int_a^b p(x)\,dx. \tag{2.5}$$

As a result, it seems that a definition of $H(P)$, which would be fully consistent with the entropy of incomplete probability distribution (equ. (2.4), should be the average

$$\overline{H}(P) = -\frac{\sum_{i,j\in\Omega} b(x_i,y_i)\ln b(x_i,y_i)}{\sum_{i,j\in\Omega} b(x_i,y_i) b(x_i,y_j)}. \tag{2.6}$$

$\overline{H}(P)$ is the mean value of uncertainty involved in the image, whilst $H(P)$ is the total uncertainty in this image.

2.2. Photon/Unit Grey-Level Allocation Model

The expression (2.1) can be obtained by considering a digital image as a probability distribution where grey-level value of pixel represents the number of photons reaching this point (Frieden, 1980). We can assume to have a number B of photons and an initial image in the form of an empty grid comprising n cells (pixels). The B photons are allocated, one at a time, among the n cells with uniform spatial probability (Frieden, 1972). This model is sometimes referred to as the "monkey model" by the analogy with a group of monkeys which randomly throw B balls in a two-dimensional array of boxes to form the image (Gull et al,

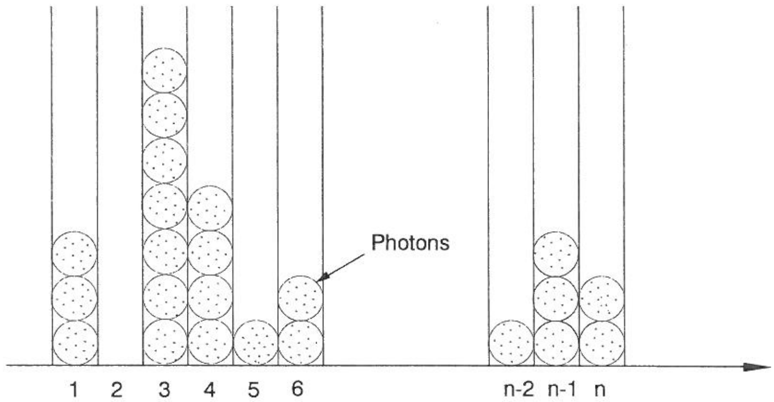

Fig. 1. Photon / unit grey-level allocation model

1978; Skilling, 1986; Jaynes, 1986). Each ball represents a unit grey-level and each box a pixel, see Fig. 1 for the one-dimensional case

The number of ways W that an image $(b_1, b_2, \ldots, b_n)$ can be formed is

$$V(b_1, \ldots, b_n) = B!/b_1 \ldots b_n, \qquad (2.7)$$

and in a classical manner, using Stirling's approximation $n! = n^n e^{-n}(2\pi n)^{1/2}$, one finds that

$$\ln \cong H(P). \qquad (2.8)$$

3. Shannon Entropy of non Random Functions

3.1. Entropy of Non Random Continuous Functions

Proposition 3.1. *Let $f : \mathbf{R}^n \to \mathbf{R}^n$, $x \to f(x)$ designate a continuously differentiable function, the Jacobian determinant of which is denoted by $f'(x)$. A measure of the amount of uncertainty it involves on the domain Ω, which is fully consistent with Shannon entropy of random variable, is the S-entropy (S holds for Shannon) of order $c \in \mathbb{R}$, defined by*

$$H_c(f) = \frac{\int_\Omega |f'(x)|^c \ln |f'(x)| \, dx}{\int_\Omega |f'(x)|^c \, dx}. \qquad (3.1)$$

PROOF. (i) Let $x \in \Omega \supset \mathbf{R}^n$ and $X' \in \mathbf{R}^n$ denote two random vectors; then the Shannon entropy of the pair (X, X') is defined by the equation

$$H(X, X') = H(X) + H(X'|X), \qquad (3.2)$$

where $H(X, X')$ is the conditional entropy of X' given X. Clearly

$$H(X'|X) = \int_\Omega p(x) H(X'|X = x) \, dx. \qquad (3.3)$$

(ii) Next, assume that the random vectors X and Y are related by the equation $Y = f(X)$, where $f(.)$ is a continuously differentiable function; then, according to a basic result of Shannon information theory, one can write $H(Y)$ in the form

$$H(Y) = H(X) + \int_\Omega p(x) \ln |f'(x)| \, dx. \qquad (3.4)$$

(iii) Analogously with the equ. (3.2), we get the following

$$H(Y) = H(X, f(.); \Omega) = H(X) + H(f(.); \Omega|x), \qquad (3.5)$$

and to define the conditional entropy of $f(.)$ on Ω given X, by the expression

$$H(f(.); \Omega|X) = \int_\Omega p(x) \ln |f'(x)| \, dx. \qquad (3.6)$$

(iv) This being so, it is well known that the equ. (3.2) yields the equality

$$H(X, X') = H(X) + H(X'), \qquad (3.7)$$

when and only when X and X' are independent. Moreover one has the inequality

$$H(X'|X) \leq H(X').$$

As a result, we can write

$$H(X') = \max_{p(x)} H(X'|X), \qquad (3.8)$$

subject to the condition that

$$H(X) = \text{constant}. \qquad (3.9)$$

(v) Analogously, we shall define the entropy of $f(.)$ on the domain Ω by the relation

$$H(f(.);\Omega) = \max_{p(x)} H(f(.);\Omega|X),$$

given the constraint (3.9). We should express that the quantity

$$\int_\Omega p(x)[\ln|f'(x)| + \lambda + \mu \ln p(x)]\,dx$$

is optimum where the Lagrange multipliers λ and μ are associated with the constraints (3.9) and the normalization condition on $p(x)$, respectively. Then we have the equation

$$\ln|f'(x) + \lambda + \mu[\ln p(x) + 1] = 0,$$

hence

$$p(x) = |f'(x)|^{-1/\mu} \exp[-(1 + (\lambda/\mu))].$$

The normalization condition on $p(x)$ provides

$$p(x)=|f'(x)|^{-1/\mu} / \int_\Omega |f'(x)|^{-1/\mu}\,dx. \quad (3.10)$$

In order to determine the constant μ, we shall refer to the constraint (3.9), namely

$$\int_\Omega p(x) \ln p(x)\,dx = -h,$$

or

$$\ln K(\mu) + \frac{\int_\Omega |f'(x)|^{-1/\mu} \ln|f'(x)|\,dx}{\mu K(\mu)} = h, \quad (3.11)$$

where $K(\mu)$ is defined by

$$K(\mu) = \int_\Omega |f'(x)|^{-1/\mu} dx. \quad (3.12)$$

The constant μ appears as an implicit function of h. Letting $-1\mu = c$ and substituting the expression (3.10) for $p(x)$ into (3.6) results in the entropy (3.1).

3.2. Entropy of Non-Random Discrete Functions

Assume that $f(.)$ is a discrete function defined on a finite grid; then the expression of the corresponding entropy will be obtained by making the substitution $\int \leftarrow \sum, f' \leftarrow \Delta f$ into (3.1).

For instance, for a one-dimensional function $f(x_i) \in \mathbf{R}$, $x_i \in \mathbf{R}$, one has

$$H_c(f_\Delta(.);\Omega) = \frac{\sum_{i=0}^{n-1} |\Delta f_i|^c \ln|\Delta f_i|}{\sum_{i=0}^{n-1} |\Delta f_i|^c}, \quad c \in \mathbf{R} \quad (3.13)$$

with the notation $\Delta f_i = f(x_{i+1} - f(x_i)$.

Notice that, exactly like with Shannon entropy of random variables, there is no direct relation between the S-entropy of continuous functions and the S-entropy of discrete functions (or discrete mappings). We are actually facing two different concepts, and here we used the formal approach only for the sake of simplification. For further details, see for instance (Jumarie, 1994).

3.3. Further Remarks and Comments

(i) Assume that $f(x)$, in the expression of $H_c(f(.);\Omega)$ is the (cumulative) distribution function

$$F(x) = \int_{-\infty}^{x} p(\xi)\,d\xi$$

of the scalar valued random variable X; then one has the equation

$$H_1(F(.);R) = -H(X). \quad (3.14)$$

On the surface, the minus sign could look troublesome; but, nevertheless, it is quite right so; the contrary would be rather puzzling. $H_1(F)$ measures the uncertainty involved in the mathematical expression of $F(.)$, whilst $H(X)$ characterizes the uncertainty related to the value of X; the magnitudes of these two uncertainties vary in opposite ways. $H(X)$ achieves its maximum value when $p(x)$ is the uniform density, in which case the graph of $F(.)$ is a straightline, that is to say, the curve which exhibits the smallest amount of complexity.

(ii) Another way to seize the informational meaning of the ln-term is as follows:

Firstly, according to the equation

$$H(kX) = H(X) + \ln|k|, k = \text{constant}. \quad (3.15)$$

In this expression, $\ln|k|$ can be thought of as the amount of uncertainty involved in the magnitude of the constant k in the sense that the number of bits we need to encode $|k|$ is exactly equal to $\ln|k|/\ln 2$. To encode k, we shall need one bit more for its sign.

Next, consider the Taylor expansion

$$f(x+Y) \cong f(x) + Y^T f'(x), \qquad (3.16)$$

where T denotes transpose operation, where Y is a "small" random variable (i.e. with a small variance). Taking the conditional entropy of both sides, given $X = x$, yields

$$H(f(x+Y)|x) \cong H(Y) + \ln|f'(x)|. \quad (3.17)$$

It can be re-writenas follows:

$$H(f(x+Y)|x) \cong H(Y) + \ln|f'(x)|. \quad (3.18)$$

to obtain the identity

$$H(f(.)|x) = \ln|f'(x)|. \qquad (3.19)$$

(iii) If we define

$$\begin{aligned} |f'|_m &= \min|f'(x)|, \quad x \in \Omega \\ |f'|_M &= \max|f'(x)|, \quad x \in \Omega \end{aligned} \qquad (3.20)$$

then it follows:

$$\begin{aligned} \lim H_c(f(.);\Omega) &= \ln|f'|_M \text{ as } c \uparrow +\infty, \\ &= \ln|f'|_m \text{ as } c \downarrow -\infty. \end{aligned} \qquad (3.21)$$

In a similarmanner, defining

$$\begin{aligned} |\Delta f|_m &= \min|\Delta f(x_i)|, \quad i = 0,1,2,\ldots,n-1, \\ |\Delta f|_M &= \max|\Delta f(x_i)|, \quad i = 0,1,2,\ldots,n-1, \end{aligned} \qquad (3.22)$$

gives the equation

$$\begin{aligned} \lim H_c(f_\Delta(.);\Omega) &= \ln|\Delta f|_M \text{ as } c \uparrow +\infty, \\ &= \ln|\Delta f|_m \text{ as } c \downarrow -\infty. \end{aligned} \qquad (3.23)$$

(iv) The parameter c in the expression of the S-entropy appears as a Lagrange parameter, and, strictly speaking, its value should be defined by the constraints. Unfortunately, in this way, we shall not obtain a formula in a closed form. So, in a general modelling, we shall rather consider c as a (structural) parameter of the definition, and we shall rather try to determine its suitable value by using other considerations and remarks.

(v) As a matter of fact, the density

$$p_c(x) = |f'(x)|^c \Big/ \int_\Omega |f'(x)|^c\, dx \qquad (3.24)$$

is proportional to the frequency with which the path generated by the point $(x, f(x))$ is scrutinized. Clear, $p_c(x)$ defines the corresponding scanning frequency.

The value $c = 0$ defines the uniform observation (each point $(x, f(x))$ is visited with the same frequency), $c = +\infty$ characterizes an observation concentrated at $(x_M, f(x_M)$, and $c = -\infty$ corresponds to the observation concentrated at $(x_m, f(x_m)$.

(vi) For the sake of consistency with Shannon entropy of random variables, we shall select the special value $c = 1$, to measure the amount of uncertainty involved in $f(.)$, and we shall set

$$H(f(.);\Omega) = H_1(f(.);\Omega).$$

(vii) $H(f_\Delta(.);\Omega)$ is the mean value of the uncertainty involved in the absolute value of each step $\Delta f(x_i)$ on Ω, so the total amount of uncertainty involved in all the n steps is $nH(f_\Delta(.);\Omega)$.

(viii) Considering $|\Delta f|_m$ as the measurement unit of the magnitude of $f(.)$, one is led to introduce the relative variation

$$|\overline{\Delta f_i}| = |\Delta f_i| / |\Delta f|_m, \qquad (3.25)$$

and to introduce the new entropy

$$H_c(|\Delta f|_m^{-1} f_\Delta(.);\Omega) = \sum_{i=0}^{n-1} |\overline{\Delta f_i}| \Big/ \sum_{i=0}^{n-1} |\overline{\Delta f_i}|^c. \qquad (3.26)$$

4. A New Approach to Entropy of Images

4.1. Main Definition

Definition 4.1. Let an image P be defined by the brightness function $b(x,y)$ on the domain Ω. As a result of the definition of entropy of non-random functions, the amount of uncertainty involved in the latter is defined by

$$H_c(b(.,.);\Omega) = \frac{\int_\Omega |b''_{xy}(x,y)|^c \ln|b''_{xy}(x,y)|\, dx\, dy}{\int_\Omega |b''_{xy}(x,y)|^c dx\, dy} \qquad (4.1)$$

In the discrete case, this entropy reads

$$H_c(b_\Delta(.,.);\Omega)=\frac{\sum_{i,j}|\Delta^2 b(x_i,y_j)|^c \ln|\Delta^2 b(x_i,y_j)|}{\sum_{i,j}|\Delta^2 b(x_i,y_j)|} \quad (4.2)$$

with

$$\Delta^2 b(x_i,y_j) = b(x_{i+1},y_{j+1}) - b(x_{i+1},y_j) - b(x_i,y_{j+1}) + b(x_i,y_j). \quad (4.3)$$

Derivation of this definition. The expression (4.1) can be supported by the following arguments.

(i) Let $f(x,y) = g_1(x)g_2(y)$, $x \in \Omega_1 \subset \mathbf{R}$, $y \in \Omega_2 \subset \mathbf{R}$ denote an $\mathbf{R}^2 \to \mathbf{R}$ function. According to the practical significance of informational entropy as a measure of uncertainty, it is quite understandable to define the entropy of $f(x,y)$ by

$$H_c(f(.,.);\Omega_1 \times \Omega_2) = H_c(g_1(.);\Omega_1) + H_c(g_2(.);\Omega_2). \quad (4.4)$$

(ii) Expliciting the right-side term of (4.4) yields

$$H_c(f(.,.);\Omega_1 \times \Omega_2) = \frac{\int_{\Omega_1}\int_{\Omega_2} |f''_{xy}(x,y)|^c \ln|f''_{xy}(x,y)|\, dx \cdot dy}{\int_{\Omega_1}\int_{\Omega_2} |f''_{xy}(x,y)|^c dxdy}. \quad (4.5)$$

(iii) We now consider this expression (4.5) as a definition which applies to any function $f(x,y)$.

4.2. Further Remarks and Comments

Definition of the "monkey model" of image entropy assumes that the individual pixels are mutually independent; in other words, that the pixels of the image could be randomly rearranged without modifying the amount of uncertainty contained in the picture, what is contrary to the intuition (Titterington, 1984). In order to reply to this criticism, Skilling (1986) suggested to replace $H(P)$ in equ. (2.1) by

$$H(P) = -\sum_{i,j} b_{ij}\ln(b_{ij})/m_{ij}), \quad b_{ij} = b(x_i,y_i), \quad (4.6)$$

where m_i is a measure which takes into account this dependence.

In the following, we wish to point out that the dependence of the pixels appears in quite a natural manner in the expression (4.1), and in order to simplify the explanation, we shall consider a line of pixels characterized by the brightness function $b(x_0), b(x_1), \ldots, b(x_n)$. We are interested in the mutual entropy $H(b_0, b_1, \ldots, b_n)$, $b_i = b(x_i)$; assuming that b_i depends upon b_{i-1} only, for every i, then one will have the equation

$$H(b_0,b_1,\ldots,b_n) = H(b_0) + H(b_1|b_0) + \ldots + H(b_n|b_{n-1}). \quad (4.7)$$

According to section 3, one has

$$H(b_{i+1}|b_i) = \int_{\mathbf{R}} p_i(x)\ln|\Delta b(x_i)|\, dx, \quad (4.8)$$

therefore

$$H(b_0,b_1,\ldots,b_n) = H(b_0) + \sum_i \int_{\mathbf{R}} p_i(x)\ln|\Delta b(x_i)|\, dx. \quad (4.9)$$

5. Image Encoding

5.1. Statement of the Problem

As a simple illustrative example, assume that we want to encode the horizontal row of pixels of the preceding subsection, and let denote the brightness of the i-th pixel. In terms of the variation of this brightness, we shall have to encode the set of data $b_0, b_1, \ldots, b_{n-1}$, or, in a similar manner the set $(|b_0|, \operatorname{sgn} b_0)$, $(|\Delta b_0)|, \operatorname{sgn}\Delta b_0)$, $\ldots$, $(|\Delta b_{n-1}|, \operatorname{sgn}\Delta b_{n-1})$.

Then, the problem is to determine the mean number of bits necessary to encode each component.

This problem refers to the H_c entropy, and we shall clarify the significance of the latter in terms of encoding.

5.2. Relation Between Entropy of Non-Random Functions and Encoding

It should be pointed out that the equ. (3.14), i.e. $H(F(.);\mathbf{R}) = -H(X)$ which connects the entropy of the (cumulative) distribution function $F(X)$ and the entropy of X, is already transparent in one of the encoding techniques already suggested by Shannon himself ((1948) (see for instance Yaglom et all 1960).

The basic idea of this method is the following. We write down the numbers x_i, $i = 1, 2, \ldots, n$ of the alphabet following the order of decreasing probabilities $p_1 \geq p_2 \geq p_3 \geq \ldots \geq p_n$ (recall that $p_i = pr(X = x_i)$ and consider the sums p_i, $i = 1, 2, \ldots, n$ defined by the expressions

$$p_1 = 0, \qquad (5.1)$$
$$p_i = p_1 + p_2 + \ldots + p_{i-1}, \; i=1, 2, \ldots, n. \qquad (5.2)$$

In the inequality $p_1 < p_2 < p_3 < \ldots < p_n$, one can consider the "$n$ numbers" $\{p_1, \ldots, p_n\}$ as an alphabet which is in a one-to-one correspondence with the initial alphabet $\{x_1, \ldots, x_n\}$. Then we have to encode this new alphabet, that is to say, to associate a suitable sequence of bits to each number p_i. In other words, encoding the sequence $\{x_i\}$ is equivalent to enconding the sequence $\{p_i\}$, and the amount of information involved should be the same.

The fact that $H_1(F(.); \mathbf{R})$ is negative should not be disturbing, and is merely due to the fact that one has the inequality $p_i < 0$ for all i, in such a manner that it can be written in the form

$$p_i = a_1 2^{-1} + a_2 2^{-2} + \ldots + a_k 2^{-k} + \ldots. \qquad (5.3)$$

Let L_i be defined by the condition

$$-\frac{\ln p_i}{\ln 2} \leq L_i \leq -\frac{\ln p_i}{\ln 2} + 1, \qquad (5.4)$$

for all i, then one can show that, if for each p_i, one retains only the first numbers $a_1, a_2. \ldots, a_{L_i}$, the sequences obtained are all different, and thus define the code.

5.3. On the Relation Between Entropy of Random Variables and Entropy of Non-Random Functions

Background on the meaning of c

The fact that $H_c(f(.); \Omega)$ is an increasing function of c gives rise to following question. Since we are interested in the encoding algorithms which involve a smaller number of bits, it would at first glance be quite relevant to use H_0 instead of H_1 for defining the corresponding code. Clearly, since one has $H_0 < H_1$, then H_1 should be disqualified in advance.

As a matter fact, this remark is not quite correct. A good encoding procedure is a trade-off between the number of bits necessary to encode the picture on the one hand, and the amount of information which is preserved (in the transformation) on the other. In other words, we should ascertain that the value $c = 0$ does not lessen the amount of information contained in the image.

Let us examine this point. If we denote by X_c the random variable whose probability density is expressed by equ. (3.24), then a simple calculation yields the relation

$$H(X_c) + cH_c(f(.); \Omega) = \ln V_c(f(.); \Omega), \qquad (5.5)$$

with the notation

$$V_c(f(.); \Omega) = \int_\Omega |f'(x)|^c dx \qquad (5.6)$$

which can be thought of as related to the total variation of $f(.)$ on Ω.

According to this equation, the value $c = 1$ appears to be a neutral value which neither creates nor destroys information. Clearly, the equation

$$H(X_1) + H_1(f(.); \Omega) = \ln V_1(f(.); \Omega) \qquad (5.7)$$

seems like a referential relation in which the right side term is absolute, in the sense that it depends upon $f(.)$ only, irrespective of any external parameter (namely c).

In contrast, the values $c > 1$ and $c < 1$, in the right side term, may increase or lower the amount of uncertainty involved in $f(.)$, depending upon the position of $|f'(x)|$ with respect to the unit. If one has $|f'(x)| > 1$ for all $x \in \Omega$, then the equ. (5.5) defines an increase of uncertainty; and when, on the contrary, one has that $|f'(x)| < 1$, then $c > 1$ reduces the uncertainty (while $c < 1$ increases it).

Identification principle

The equ. (5.5) can be re-written in the form

$$H(X_c) = -cH_c\Big(\frac{f(.)}{V_c(f(.); \Omega)}; \Omega\Big) + (1-c)\ln V_c(f(.); \Omega), \qquad (5.8)$$

which shows that, in a case when $c = 1$, there is a complete identification between the entropy of X_c and the entropy of $f(.)$. In other words, in the encoding of $f(.)$ with $c = 1$, we shall be able to duplicate the rationale of the subsection 5.2.

5.4. Application to Image Encoding

We can now get back to the problem of encoding a row of pixels, see subsection 5.1. In order to achieve this objective, we can directly duplicate the result of the preceding section.

(i) Firstly, we have shown that there is a complete identification between encoding random variables and encoding non random functions.

(ii) So, as an illustrative example, consider the brightness function pictured in Fig. 2.

According to the algorithm of subsection 5.2, we shall introduce the steps Δ_i with the additional condition $|\Delta_1| > |\Delta_2| > |\Delta_3| > |\Delta_4| > |\Delta_5|$ on the one hand, and the probability-like distribution on the other.

$$p_i = |\Delta_i| / \sum_{i=1}^{5} |\Delta_i|, \quad i = 1, \ldots, 5. \qquad (5.9)$$

We shall have to encode the numbers p_i, $i = 1, 2, 3, 4, 5$, as defined by the equ. (5.2); together with the sign of Δ_i for each i. In the following, we shall examine the problem of image compression; but before we do that, we need to define the Renyi entropy of non-random functions.

6. Renyi Entropy of non Random Functions

6.1. Continuously Differentiable Functions

Proposition 6.1.. *Let $f : \mathbf{R}^n \to \mathbf{R}^n$, $x \to f(x)$ denote a continuously differentiable function, the Jacobian determinant of which is denoted by $f'(x)$. Then the measure of the amount of uncertainty it involves on the domain Ω, and which, furthermore, is fully consistent with Renyi entropy of random variables, is the R-entropy (R stands for Renyi) of order α, $\alpha \neq 1$, $\alpha \in \mathbf{R}$ defined by*

$$H_{R,\alpha}(f(.), \Omega) = -\frac{1}{1-\alpha} \ln \frac{\int_\Omega |f'(x)|^\alpha dx}{\int_\Omega |f'(x)| dx}. \qquad (6.1)$$

Proof The proof of this result is summarized as follows

(i) The Renyi entropy of order α of the real valued n-dimensional random variable X with the probability density $p(x)$ is defined by the expression (see for instance Aczel et al, 1975):

$$H_{R,\alpha} = -\frac{1}{1-\alpha} \ln \int_{\mathbf{R}^n} p^a(x)\, dx, \qquad (6.2)$$

and one can show that the following inequalities hold,

$$H_{R,\alpha} > H(X), \; \alpha < 1, \qquad (6.3a)$$

$$H_{R,\alpha} < H(X), \; \alpha > 1, \qquad (6.3b)$$

(ii) The transformation $Y = f(X)$ yields

$$H_{R,\alpha}(Y) = -\frac{1}{1-\alpha} \ln \int_{\mathbf{R}^n} p^a(x) |f'(x)|^{1-a}\, dx, \qquad (6.4)$$

and analogously with the derivation of the entropy of $f(.)$, we can write

$$H_{R,\alpha}(Y) = H_{R,\alpha}(X) + H_{R,\alpha}(f(.); \mathbf{R}^n); \qquad (6.5)$$

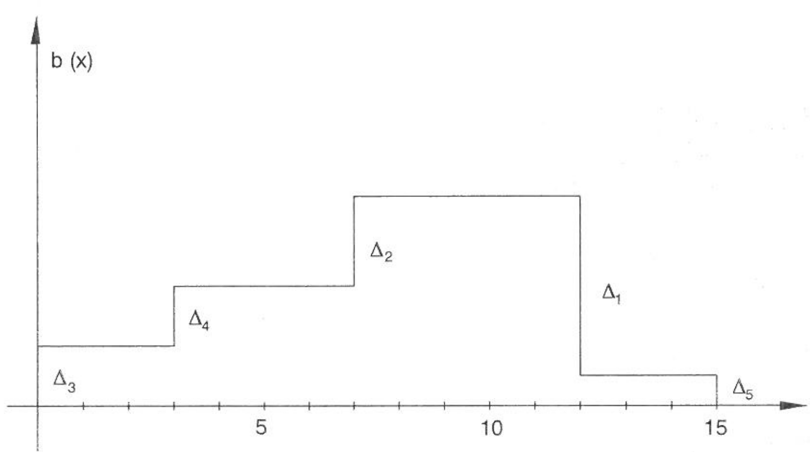

Fig. 2. Brightness of a one-dimensional row of pixels

therefore

$$H_{R,\alpha}(f(.);\mathbf{R}^n) = H_{R,\alpha}(Y) + H_{R,\alpha}(X). \quad (6.6)$$

(iii) The quantity to maximize now reads (on the domain Ω):

$$\max \frac{1}{1-\alpha} \ln \frac{\int_\Omega p^a(x)|f'(x)|^{1-\alpha}dx}{\int_\Omega p^\alpha(x)\,dx} \quad \text{w.r.t. } p, \quad (6.7)$$

subject to the constraint

$$\int_\Omega p(x)\,dx = 1, \quad (6.8)$$

$$\int_\Omega p(x)^\alpha\,dx = \text{constant}. \quad (6.9)$$

(iv) Using the Lagrange parameter technique, one obtains the equ. (6.1) as a "special case".

6.2. Discrete Functions

We shall formally obtain the entropy of discrete functions by substituting $\int \leftarrow \sum$, $f' \leftarrow \Delta f$ into the equ. (6.1).

For instance, for an $\mathbf{R} \to \mathbf{R}$ function, the Renyi entropy is

$$H_{R,\alpha}(f_\Delta(.);\Omega) = -\frac{1}{1-\alpha} \ln\left[|\Delta f_i|^a / \sum_{i=0}^{n-1} |\Delta f_i|\right]. \quad (6.10)$$

7. Image Compression

7.1. Renyi Entropy of Image

Definition 7.1. Consider the image as in the Definition 4.1. Its Renyi entropy of order $\alpha \in \mathbf{R}$, $\alpha \neq 1$, is

$$H_{R,\alpha}(b(.,.);\Omega) = -\frac{1}{1-\alpha} \ln \frac{\int_\Omega |b''_{xy}(x,y)|^a dx\,dy}{\int_\Omega |b''_{xy}(x,y)|\,dx\,dy}; \quad (7.1)$$

and in the discrete case, it is

$$H_{R,\alpha}(b(.,.);\Omega) = -\frac{1}{1-\alpha} \ln \frac{\sum_{i,j} |\Delta^2 b''(x_i,y_j)|^a}{\sum_{i,j} |\Delta^2 b''(x_i,y_j)|}. \quad (7.2)$$

In order to accomplishthese definitions, it is sufficient to duplicate the rationale of Section 4.

7.2. Application to Image Compression

Let us first note that the following properties hold:

$$\begin{aligned} H_{R,\alpha}(b(.,.);\Omega) &> H(b(.,.);\Omega),\ \alpha < 1, \\ H_{R,\alpha}(b(.,.);\Omega) &< H(b(.,.);\Omega),\ \alpha > 1, \end{aligned} \quad (7.3)$$

The second inequality can be interpreted as follows. The transformation $|b''_{xy}(x,y)|| \leftarrow |b''_{xy}(x,y)|^\alpha$, with $\alpha < 1$, lessens the uncertainty contained in the image; and as a result, the number of bits which are necessary to encode the transformed image will be lower than for the original image.

In practical implementation, we shall avoid the use of the Renyi entropy itself, and we shall proceed as follows: we shall encode $|\Delta_i^2|^\alpha$ instead of $|\Delta_i^2|$, and to this end, we shall refer to the entropy

$$H_1^\alpha(b(.,.);\Omega) = \frac{\sum_{i,j} |\Delta^2 b(x_i,y_j)|^a \ln |\Delta^2 b(x_i,y_j)|^\alpha}{\sum_{i,j} |\Delta^2 b(x_i,y_j)|^\alpha}. \quad (7.4)$$

Example

Assume that

$$(\Delta_1, \Delta_2, \Delta_3, \Delta_4, \Delta_5) = (7, 6, 5, 4, 3).$$

(i) A simple calculation yields

$$\begin{aligned} V_1(f) &= \sum_i |\Delta_i| = 25, \\ H_0(f) &= 1.57, \\ H_1(f) &= 1.65, \\ H(X_1) &= 1.57 \quad \text{see equ. (5.8))} \end{aligned}$$

The fact that $H(X_1)$ is smaller than $H_1(f)$ is quite understandable. $H(X_1)$ refers to the encoding of $|\Delta_i| / \sum_i |\Delta_i|$ for all i, so that if we want to determine $|\Delta_i|$, we need to know the value of $\sum_i |\Delta_i|$.

(ii) Regarding the Renyi entropy, in the special case when $\alpha = 0.5$, one finds that

$$H_{R;0.5}(f) = 1.63$$
$$H_{R;0.5}(X_1) = 1.58;$$

and, as expected, the following inequalities are satisfied, clearly $H_{R;0.5}(X_1) < H_{R;0.5}(f)$ and $H_{R;0.5}(f) < H_1(f)$. (

iii) Now assume that we are encoding $|\Delta_i|^\alpha$; then will be the entropy

$$H_1^{0.5}(f) = 0.80,$$

which, as expected, is lower than $H_1(f)$.

8. Similitudes of Images via Brightness Cross-Entropy

8.1. Statement of the Problem

In a 2–D geometry described by rectangular coordinates, the similitude transformation is defined by the equs $x' = kx$, $y' = ky$, $k=$ constant. It multiplies the distance between every pair of points by the same constant, referred to as the ratio of similitude. Such a change of variables transforms figures into similar patterns.

When we try to apply the same concept to images, we can do it in two different ways.

(i) We can consider some special drawings in the picture, and measure their similutude.

(ii) Or else, we can consider the image as a whole and define the corresponding measure of similitude. This point of view is particularly relevant when it is not easy to meaningfully extract some characteristic drawings from the image, which is the case, for instance, with an "impressionist" painting.

In such a case, the most direct way is to compare the brightnesses themselves, which brings us to the following definition.

Definition 8.1. Given two pictures characterized by the respective brightness functions $b_1(x,y)$ and $b_2(x,y)$ defined on the same domain Ω, we shall say that they are similar if there exists a positive constant k such that

$$b_2(x,y) = kb(x,y), \quad (x, y \in \Omega. \qquad (8.1)$$

The task now is to have a practical criterion to measure the so defined.

8.2. Cross-Entropic Variance of Brightness Functions

In the wake of thought of statistics, the first idea to obtain a measure of image similitude, is to consider the variance of the variable $b_2(x,y)/b_1(x,y)$, which provides the sought property. Indeed, if the latter is zero, then $b_2(x,y)/b_1(x,y)=$ constant. For instance, one can select the quantity

$$D(b_1,b_2) = \sum_{i,j} b_2(x_i,y_j)\left[\frac{b_2(x_i,y_j)}{b_1(x_i,y_j)}\right]^2 \Big/ \sum_{i,j} b_2(x_i,y_j) - \left[\sum_{i,j} b_2(x_i,y_j)\frac{b_2(x_i,y_j)}{b_1(x_i,y_j)} \Big/ \sum_{i,j} b_2(x_i,y_j)\right]^2. \qquad (8.2)$$

Nevertheless, with our purpose of generalizing Kullback cross-entropy (or relative entropy or divergence between probability distribution) $\sum_i q_i \ln(q_i/p_i)$, we shall rather consider the variance of $\ln|b_2/b_1|$:

$$V_H(b_2,b_1) = \sum_{i,j} b_2(x_i,y_j)\ln^2\frac{b_2(x_i,y_j)}{b_1(x_i,y_j)} \Big/ \sum_{i,j} b_2(x_i,y_j) - \left[\sum_{i,j} b_2(x_i,y_j)\ln\frac{b_2(x_i,y_j)}{b_1(x_i,y_j)} \Big/ \sum_{i,j} b_2(x_i,y_j)\right]^2. \qquad (8.3)$$

which will be referred to as the *cross-entropic variance of the brightness functions* of the two pictures to be compared. It is equal to zero when and only when, there exists a positive constant k such that $b_2(x,y) = kb_1(x,y)$.

9. Spatial Informational Divergence of Images

9.1. Preliminary Background

Divergence of random variables. Let X denote a scalar real valued random variable defined on Ω, and let $H(X)$ denote its (Shannon) entropy. Let $H_m(X)$ denote the maximum value of $H(X)$. Then the divergence $D(X)$ of X is defined by

$$D(X) = H_m(X) - H(X). \qquad (9.1)$$

$D(X)$ can be thought of as a measure of the (structural) information contained in X. When $D(X) = 0$, there is no information in the stochastic definition of X, and the larger $D(X)$ is, the more information is contained in X. $D(X)$ is not an uncertainty measure: it is a difference of information, and thus an information itself. If one considers that X defines a system, then $D(X)$ would measure the complexity of the latter.

Note that this divergence is differs from the so-called Kullback divergence between two probability distributions.

Entropy of Markovian processes. Let $X(t)$ denote a scalar real valued stochastic process which satisfies the Fokker-Planck-Kolmogorov equation

$$\partial_t p(x,t) = -\partial_x[f(x,t)p] + \frac{1}{2}\partial_{xx}[g^2(x,t)p], \tag{9.2}$$

the Shannon entropy of $X(t)$ on the time interval $(0, T)$ is

$$H(X;0,T) = H(X,(0)) + \int_0^T \int_{\mathbf{R}} p(x,t) \ln[2\pi e g^2(x,t)]^{1/2} dx\, dt. \tag{9.3}$$

This expression is a direct consequence of the equation

$$H(X(t+dt)|X(t) = x) = \ln[2\pi e g^2(x,t). \tag{9.4}$$

For further details, see Jumarie, 1997.

9.2. Application to Divergence of Image

For convenience purpose, we shall consider continuous brightness functions, but the reader will easily generalize to (discrete) pixels by making the usual substitutions.

(ii) In the first step, we shall refer to the monkey model of entropy

$$\overline{H}(P) = -\frac{\int_\Omega b(x,y) \ln b(x,y)\, dx\, dy}{\int_\Omega b(x,y)\, dx\, dy}. \tag{9.5}$$

and we shall assume that it is equivalent to the term H_m in the expression of $D(X)$ in the equation (9.1). This assumption is quite correct, since this model implicitly presupposes that the pixels are mutually independent.

(ii) Next, acccording to the equ. (3.6), the term $|\ln b''_{xy}(x,y)|$ can be understood as the density of conditional entropy. As a result, using the probabilistic significance of $b(x,y)$ in equ. (9.5), and analogously with equ. (9.3), one can define another model of image entropy in the form

$$\overline{H}(P) = \frac{\int_\Omega b(x,y) \ln |b''_{xy}(x,y)|\, dx\, dy}{\int_\Omega b(x,y)\, dx\, dy}, \tag{9.6}$$

which explicitly takes the pixel dependence into account. This entropy is exactly the parallel of $H(X;0,T)$ in the expression of $D(X)$, see equ. (9.3).

(iii) Analogously, with the equ. (9.1), we are led to the following

Definition 9.1. An informational divergence $D(P)$ of the image P with the brightness function $b(x,y)$, with is fully consistent with the monkey model of entropy on the one hand, and which takes into account the mutual dependence of pixels on the other, is given by the expression:

$$\begin{aligned} D(P) &= \overline{H}(P) - \overline{H}(P/b) \\ &= -\frac{\int_\Omega b(x,y) \ln b(x,y) |b''_{xy}(x,y)|\, dx\, dy}{\int_\Omega b(x,y)\, dx\, dy} \end{aligned} \tag{9.7}$$

In this expression, $\ln b(x,y)$ is the density of uncertainty involved in the brightness of the pixel (x,y), while the component $\ln|b''_{xy}(x,y)|$ represents the density of uncertainty contained in the contrast of the brightnesses of neighbouring pixels. $D(P)$ could be a valuable tool to analyze the image from the point of view of its informational content.

10. Concluding Remarks

In our paper, we have shown how, by combining the maximum conditional entropy principle with a basic equation of information theory, one can derive a new model of image entropy, which involves the contrast of the brightness instead of the brightness itself, therefore, a new approach to image processing.

As a by-product, the information theoretic framework described herein, provides some support to the so-called monkey model of image entropy.

The theory applies to Renyi entropy, and this allows us to introduce a γ-exponent for image compression using $b''_{xy}(x, y)|$ instead of $b\gamma(x, y)$, as suggested by other authors (Horn, 1986).

Lastly, in the same framework we have obtained a new concept of informational divergence of an image (which should not be confused with Kullback divergence between two probability distributions), which seems to be of interest in image analysis.

Acknowledgement

This paper was written, while the author, who is a French citizen from West Indies (Caribee) was living in Canada. All the help and the support came from outside the country.

References

ACZEL, J. AND DAROCZY, Z., On the Measures of Information and their Characterizations, Academic Press, New York, 1975.

CHEN, W., SMITH, C. H. AND FRALICK, S., A fast computational algorithm for the discrete cosine transform, IEEE Trans. Comm., COM–25, pp 1285–1291, 1977.

CUTLER C. C., Differential quantization of communication signals, Patent No 2, 605, 361, July 29, 2952, USA.

DELP, E. J. AND MITCHELL, O. R., Image truncation using block truncation coding, IEEE Trans. Comm., COM–27, pp 1335–1342, 1979.

FRIEDEN, B. R., Restoring with maximum likelihood and maximum entropy, J. Optical Society of America , Vol 62, No 4, pp 511–518, 1972.

FRIEDEN B. R., Statistical models for the image restoration problem, Computer Graphics and Image Processing, Vol 12, pp 40–59, 1980.

GRAHAM R. E., Predictive quantizing of television signals, IRE Wescon. Convent. Record, Pt 4, pp 147–156, 1958.

GULL, S. F. AND DANIELL, G. J., Image reconstruction from incomplete and noisy data, Nature, Vol 272, pp 686–690, 197.

HORN, B. K. P., Robot Vision, MIT Press, Massasuchetts, 1986.

JAYNES, E. T., Monkey, kangaroos and N, in Maximum Entropy and Bayesian Methods in Applied Statistics, J.H. Justices (ed.), Cambridge University Press, pp 26–58, 1986.

JUMARIE, G., Information and quantum information without probability for functions and matrices, Reviw Bulletin of the Calcutta Mathematical Society, Vol 2, No 1–2, pp 69–83, 1994.

JUMARIE, G., Order, disorder, complexity in Markovian dynamics. An approach via divergence of path entropy, Advances in Systems Science and Applications , 1997 (to appear).

KITAJIMA, H., A symmetric cosine trandform, IEEE Trans. Comp., C–29, pp 317–323, 1980.

LOEVE, M., Fonction aléatoire de second ordre, in Processus Stochastiques et Mouvement Brownien, P. Levy (ed.), Hermann, Paris, 1948.

MITCHELL, O. R., DELP, E. J. AND CARLTON, S. J., Block truncation coding: a new approach to image compression, Record IEEE Internat. Conf. Commun., Vol 1, 12B.1.1–12B.1.4, 1978.

PRATT, W., Digital Image Processing, Academic Press, New York, 1978.

ROSENFELD, A. AND KAK, A. C., Digital Picture Processing, Academic Press, New York, 1982.

SHANNON, C. E., A mathematical theory of communication I and II, Bell System Tech. J., Vol 27, pp 379–423, pp 623–656, 1948.

SKILLING, J., Theory of maximum entropy image reconstruction, in Maximum Entropy and Bayesian Methods in Applied Statistics, J.H. Justice (ed.), Cambridge University Press, Cambridge, pp 156–178, 1986.

TITTERINGTON, D. M., The maximum entropy method for data analysis, Nature, Vol 312, pp 381–382, 1984

YAGLOM, A. M. AND YAGLOM, I. M., Probability and Information (in french), NAUKA, Moscow, 1960.

Received: April, 1996
Accepted: October, 1996

Contact address:
Guy Jumarie
Department of Mathematics
University of Quebec
P.O. Box 8888 Downtown St
Montreal Qc, H3C 3P8
Canada
e-mail: jumarie.guy@uqam.ca

GUY JUMARIE West Indies, Caribee. B. Appl. Math. Paris, 1960. B. Phys. Paris, 1961. Dr. in Appl. Math. Paris 1962. Dr. Univ.,Lille, Physics, 1972. Dr. Sc. Physics, Lille, 1981. Two research books, one in relative cybernetics (Gordon and Breach) and the other one in (relative) information theory (Springer Verlag). About 280 papers. Research interest: applied control, cybernetics and systems, information theory and applications, dynamical systems, image processing, fuzzy information.