SPATIAL ENTROPY : A TOOL FOR CONTROLLING CONTEXTUAL CLASSIFICATION CONVERGENCE

Henri MAÎTRE, Isabelle BLOCH, Marc SIGELLE

Télécom Paris, Département Images
46 rue Barrault, F-75634 Paris Cedex 13
Tel: 33 1 45 81 76 55, Fax: 33 1 45 81 37 94, Email: maitre@ima.enst.fr

ABSTRACT

A new kind of entropy is proposed, which associates spatial and radiometric properties of images. The possible use of this entropy is shown firstly to measure the effect of picture processing algorithms, then to control the evolution of iterative contextual classification algorithms like Markov Random Fields.

1. INTRODUCTION

Since the early works of Shannon and Wiener [9], entropy has been widely used to express the information content of a message. Although a more modern theory based on algorithmic information concepts has been proposed in the 60's [5], the probabilistic approach of Shannon and Wiener remains the most adequate in the domain of image processing, where statistical properties govern signal and noise. Jaynes [4] has shown that the formulation they proposed is the only one which allows to express rationally the quantity of information we could expect to gain from an emitter, as soon as this emitter is defined by a stationary emission of symbols. On the other hand, maximization of entropy is now a universally recognized optimality criterion which allows to select, from a broad class of possible solutions, the one with the highest probability [8]. Kullback [6], followed by Jauch and Báron [3] proposed a differential form to measure the gain of knowledge between two successive stages of the system, i.e. the gain of information provided by one stage of processing for instance. Following this idea, we can define a good algorithm for pattern recognition as one such that it drastically reduces entropy : starting from a stage where entropy is high (i.e. where ignorance is the rule), and providing a final result where it is as low as possible (i.e. where certainty is total).

Our approach follows this track. At first we extend the conventional definition of entropy to encompass spatial properties which are known as being of prime importance for picture processing; then, we show how it may be used to compare different iterative classification algorithms.

2. ENTROPY AND IMAGE

Following the usual information theoretic approach, an image may be seen as a source of independent pixels with values in N, each level i having a probability p_i. The image entropy S is measured as:

$$S = -\sum_{i \in N} p_i \log p_i \tag{1}$$

To extend this definition to non-independent pixels, higher-order entropy has been introduced, taking into account the joint probabilities of pairs, or of triples of pixels. In communication theory, it allows for instance to evaluate limit performances of predictive coders for lossless image coding. But, in the framework of pattern recognition and image processing, this extension is not very fruitful, since it doesn't reflect the spatial dependency between pixels as seen by the user. Thus we propose an alternative extension which allows to separate grey-level (or radiometric) entropy from spatial entropy.

3. SPATIAL ENTROPY

We will abundantly make use of a vocabulary introduced in Markov Random Field theory.

Let s denote a site (or pixel), with neighbouring pixels r belonging to the neighbourhood V of s. Every pixel is characterized by the following elements:

- its grey-level: n_s, with n_s having values in $N = \{0 ... \nu\}$;
- a label: l_s, which denotes its belonging to a class of possible objects in the image (this is the result of a recognition process) $L = \{1 ... \lambda\}$;

- what we call the *color* of the neighbourhood (i.e. a description of the context of the pixel): v_s, which takes its values in $V = \{0, ..., \lambda^{CardV}\}$. This color may be economically described by the cliques contained in the neighbourhood V.

Let q denote the probability for the site s to have the characteristic values n_s, l_s, v_s: $q = proba(n_s, l_s, v_s)$, with the normalization constraint:

$$\sum_{n_s \in N} \sum_{l_s \in L} \sum_{v_s \in V} q = 1.$$

The **global system entropy** is defined as:

$$E = -\sum_{n_s \in N} \sum_{l_s \in L} \sum_{v_s \in V} q \log q. \qquad (2)$$

For each class (or label), a class marginal probability is defined as:

$$p(n, v|l) = \frac{q(n, l, v)}{\sum_{i \in N} \sum_{j \in V} q(i, l, j)} \qquad (3)$$

which produces a **global entropy of class-l**:

$$e_l = -\sum_{n \in N} \sum_{v \in V} p(n, v|l) \log p(n, v|l). \qquad (4)$$

We may also derive the classical class histograms as:

$$p(n|l) = \sum_{v \in V} p(n, v|l) = \sum_{v \in V} \left[\frac{q(n, l, v)}{\sum_{i \in N} \sum_{j \in V} q(i, l, j)} \right], \qquad (5)$$

and neighbourhood class histogram:

$$p(v|l) = \sum_{n \in N} p(n, v|l) = \sum_{n \in N} \left[\frac{q(n, l, v)}{\sum_{i \in N} \sum_{j \in V} q(i, l, j)} \right], \qquad (6)$$

which allow us to define the **radiometric (or grey level) entropy** of class l:

$$\epsilon_l = -\sum_{n \in N} p(n|l) \log p(n|l), \qquad (7)$$

and its **spatial entropy**:

$$\zeta_l = -\sum_{v \in V} p(v|l) \log p(v|l). \qquad (8)$$

Radiometric entropy is limited to:

$$\epsilon_l \leq \log N, \qquad (9)$$

whereas spatial entropy is limited to:

$$\zeta_l \leq \log V. \qquad (10)$$

They reflect different qualities of the classification. Radiometric entropy expresses how poorly homogeneous a class is. In a perfect case, with ideal signals concentrated on a small number of grey-tones, the radiometric entropy is very small and hopefully it tends to zero when all the pixels belonging to the same class have the same grey-level. On the contrary in a more complex situation, the class will be spread nearly uniformly on the whole range of grey-tones, providing a high entropy, very close from the image radiometric entropy.

On the other hand, spatial entropy expresses how irregular the pixel context is. To have a low spatial entropy a class may be spread in another (for instance in a composite texture), or grouped in compact clusters, but this configuration must be the statistically dominant everywhere in the picture.

4. APPLICATION TO PICTURE PROCESSING

The first illustration we want to make is the use of spatial entropy to measure the effect of regularization (or smoothing) algorithms. We have chosen the domain of noise filtering by means of local filters, and we compare the efficiency of several local filters: low-pass filter (mean filter), median filter (with 2 different windows : 3x3 and 5x5 pixels), morphological filter (combination of an opening and a closing in 8 connectivity), and Nagao filter (a quasi-optimal optimal filter on adaptive windows).

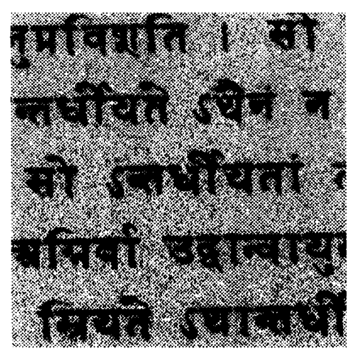

Figure 1: Original text corrupted with noise.

Our experiment is made with a highly noisy image of a text (Fig.1). The best threshold σ provides a segmentation (Fig.2) which allows one to measure the radiometric ϵ_{ink} and the spatial (in 4 connectivity) ζ_{ink} entropies of the dark component below the threshold. After applying any smoothing filter, using the same threshold σ, we may measure the effect of filtering by means of the variation of radiometric and spatial entropies. The radiometric entropy can be measured in two ways :

- first by measuring the histogram of the original image, using the mask provided by thresholding the smoothed image, denoted by ϵ'_{ink};
- or by measuring the histogram of the filtered image, using the same mask, denoted by $\epsilon"_{ink}$.

As expected, the spatial entropy after smoothing (ζ'_{ink}) strongly decreases, proving that filtering makes the contours more regular (cf. Table 1). On the contrary, a small increase of the radiometric entropy ϵ' (approximately 0.5 bit/pixel) clearly proves that the classification is degraded by the fact that some pixels, lying on the border of a class are converted to the other class because of smoothing. The difference between ϵ' and $\epsilon"$ measures the effect of correction of the grey level by filtering.

filter	ϵ	ζ	ζ'	ϵ'	$\epsilon"$	$\epsilon' - \epsilon"$
original	6.88	1.51				
median 3x3			0.81	7.43	6.57	0.86
median 5x5			0.68	7.53	6.34	1.19
mean			0.75	7.46	6.51	0.95
morpho.			0.69	7.77	6.61	1.26
Nagao			1.01	7.28	6.31	0.87

Table 1: Effect of several filters on the radiometric ϵ and the spatial ζ entropies of the dark class of image of Fig.1 after filtering.

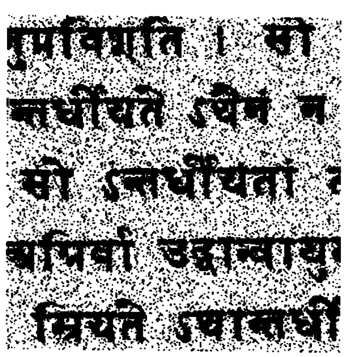

Figure 2: The original text image thresholded at $\sigma = 147$.

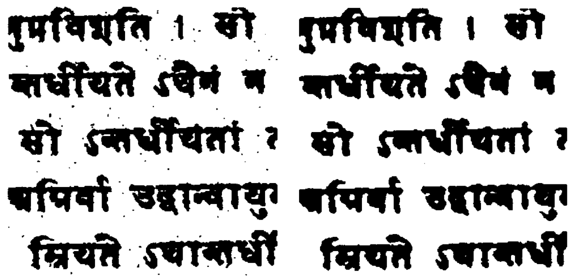

Figure 3: Result of thresholding at $\sigma = 147$ the original image smoothed by a 3x3 median filter (left), and a 5x5 median filter (right).

5. APPLICATION TO THE CONTROL OF CONTEXTUAL ITERATIVE CLASSIFICATION

Iterative contextual classification methods, like label relaxation, Markov Random Fields (MRF) or simulated annealing are often used to improve a conventional classification made on radiometric criteria only. They work as a second stage iterative re-classification which modifies the class of pixels surrounded with inconsistent context [2]. As a consequence, this second stage generally degrades the first classification from a radiometric point of view, for the sake of a better spatial coherence. They are specially appropriate when the signal is corrupted by noise (making the radiometric classification unreliable), and when strong assumptions exist on the spatial coherence of the detected areas. In this case, they are less blind than the filtering techniques which have been seen in the previous part.

The evolution of such a re-classification process is well captured by the behaviour of radiometric spatial and global class entropies: radiometric entropy usually grows up, expressing that some disorder is created in the class when introducing pixels with odd grey levels. On the contrary, spatial entropy, initially reflecting the chaos of a non-contextual classification, regularly decays. Depending on the algorithm used, global entropy may follow different evolution, either a regular decay or a more irregular one, but it usually drops down when the algorithm converges to a local or global solution. From the examination of these curves, fruitful information may be obtained when comparing algorithms. Besides the quality of the convergence point (which is immediately reflected by the asymptotic level), the speed of convergence, and its regularity are available for the user.

The demonstration is made with a detection problem in X-rays angiographic medical images where 2 classes are looked for: blood vessels which appear clearer because of contrast product, and background (bone and soft-tissues of the brain) (Fig.4). Starting with a coarse classification obtained by segmenting the histogram (Fig.5), the contextual classification is provided by a MRF, using an Ising model in 4-connectivity. The class conditional histogram (vessel or background), which are necessary to express the attachment to data in the Ising model, were learnt on training zones and modelized with normal distributions. The optimization was made using simulated annealing to guarantee an optimal convergence (Fig.6).The experiment compares several different simulated annealing schemes with different parameters (initial temperature T, and ratio of temperature decay p). The evolution of the blood-vessel class global entropy is plotted versus the number of iterations (Fig.7 and 8). Fig.7 presents the global entropy decay along iterations for class 1 (background) and class2 (blood-vessels) for several different initial temperatures and different decreases in temperature. For a similar ultimate convergence value, we see the benefit of a rather fast decay in temperature ($p = 0.80$) which provides a gain in convergence time. On Fig.8 is presented the evolution of the only spatial entropy of class 2 (blood-vessels). We see that a high initial temperature (right curves, $T = 20$) provides at first a serious disorder (high spatial entropy), which is slowly decreased to the final limit value. This disorder will make disappear nearly all information issued from the initial classification, and allows simulated annealing to be independant from the initial conditions.

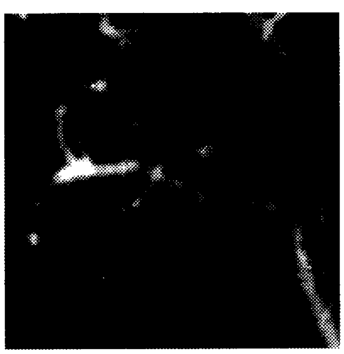

Figure 4: Original brain angiographic image: we are interested in detecting the blood-vessels which appear as clear parts because of contrast product.

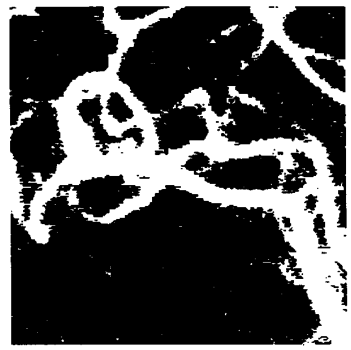

Figure 5: After histogram segmentation, a first classification is obtained, on which a contextual reclassification is made.

Figure 6: Final classification using a MRF with a simulated annealing optimization stage with $T = 4$ and $p = 0.9$.

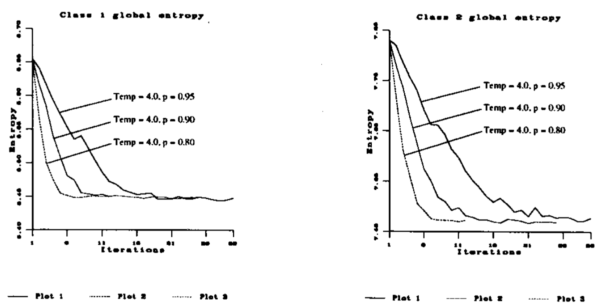

Figure 7: Global class entropies for class 1 (left) and class 2 (right) for several different parameters of the simulated annealing versus the number of iterations: the initial temperature T and the geometric decay law of temperature T.

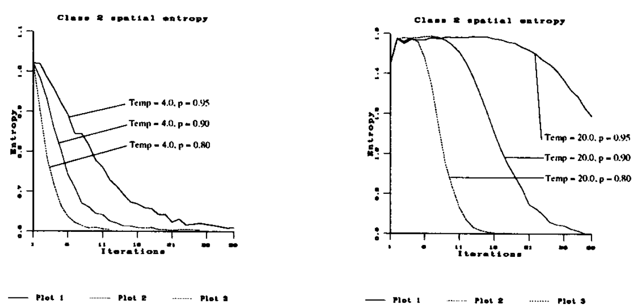

Figure 8: For class 2, the spatial entropy may increase when the initial temperature is too high.

6. CONCLUSION

We have shown the possible use of spatial entropy to control picture processing algorithms. This entropy well reflects the statistical behaviour of the context of a pixel. It provides a complementary information to the more conventional radiometric entropy which is usually measured. Alternative use of this spatial entropy have been independently proposed by means of Markov random fields for remote sensing applications [1].

7. REFERENCES

[1] M. BERTHOD, E. VOLDEN, G. GIRAUDON, *personnal communication.*

[2] S. GEMAN, D. GEMAN, *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images,* IEEE Trans on PAMI 6:721-741, 1984.

[3] J.M. JAUCH, J.M. BÁRON, *Entropy Information and Szilard's Paradox,* Helv. Phys. Acta 45, 220, 1972.

[4] E.T. JAYNES, *Information Theory and Statistical Mechanics,* Phys. Review, Vol. 106, No-4, 620-630, 1957.

[5] A. N. KOLMOGOROV, *Three approaches to the quantitative definition of information,* Information Transmission Vol 1, No 1, pp 3-11, 1965.

[6] S. KULLBACK, *Information Theory and Statistics,* Wiley NY, 1959.

[7] H. MAÎTRE, *Entropie, Information et Image,* Internal Report ENST 93D009 - Paris - 1993.

[8] MaxEnt, 12th Int. Workshop on Maximal Entropy, Paris, A.M. Djafari, G. Demoment Ed. 1992.

[9] W. SHANNON, W. WEAWER, *The Mathematical Theory of Communication,* University of Illinois Press, Urbana, 1949.