Journal of Mathematical Imaging and Vision 13, 35–56, 2000

An Analysis of Edge Detection by Using the Jensen-Shannon Divergence

JUAN FRANCISCO GÓMEZ-LOPERA
Dept. de Física Aplicada, Univ. de Granada, Campus de Fuentenueva, 18071 Granada, Spain
jfgomez@ugr.es

JOSÉ MARTÍNEZ-AROZA AND AURELIANO M. ROBLES-PÉREZ
Dept. de Matemática Aplicada, Univ. de Granada, Campus de Fuentenueva, 18071 Granada, Spain
jmaroza@ugr.es
arobles@ugr.es

RAMÓN ROMÁN-ROLDÁN
Dept. de Física Aplicada, Univ. de Granada, Campus de Fuentenueva, 18071 Granada, Spain
rroman@ugr.es

Abstract. This work constitutes a theoretical study of the edge-detection method by means of the Jensen-Shannon divergence, as proposed by the authors. The overall aim is to establish formally the suitability of the procedure of edge detection in digital images, as a step prior to segmentation. In specific, an analysis is made not only of the properties of the divergence used, but also of the method's sensitivity to the spatial variation, as well as the detection-error risk associated with the operating conditions due to the randomness of the spatial configuration of the pixels. Although the paper deals with the procedure based on the Jensen-Shannon divergence, some problems are also related to other methods based on local detection with a sliding window, and part of the study is focused to noisy and textured images.

Keywords: edge detection, image segmentation, Jensen-Shannon divergence

1. Introduction

1.1. Segmentation and Edge Detection Methods

The scientific technical literature on the treatment of computer images and vision is copious. Nevertheless, an unresolved problem is to design a universal method that can be automated—that is, that can provide good image segmentation in all cases without human intervention. This objective may in reality be unattainable, and is being replaced by an endless variety of partial solutions to specific problems (see any of the worthy texts available on this material, i.e. [1–3]). The segmentation of a digital image, at a low level, follows two basically different procedures:

1. *Growth of regions*. This is begun by seeding candidate pixels, then making regions grow within their surroundings until they meet other adjacent regions, with some halt criterion at the border. Segmentation is assured, but not the suitability of the edge.
2. *Edge detection*. Throughout the image, an exhaustive search is made for points that separate different regions. Dividing lines are assured but not their connection and, therefore, not segmentation into regions.

Examples of local methods can be found in [4, 5]. The method analysed here is of the second type. It was proposed by the authors [6] as a suitable procedure for detecting edges in several classes of images, such as

regular or natural textures, noisy regions, and other more specific images. The success of the proposed method has been shown empirically [9, 10] by simply applying it to cases. Nevertheless, the good results observed were (up to now) not theoretically sustained. This is the goal of the present work.

1.2. A Brief Description of the Method

For the sake of self-containing, our edge-detection method is briefly described here. The search for edge points is made by a sliding window that passes over the entire image, giving relative-frequency vectors of greys observed in two semi-windows and allowing decisions to be made by means of a measurement of probabilistic differences between the vectors, namely the Jensen-Shannon divergence. This is zero for equal vectors (no differences), and it takes higher values for increasing differences between the relative-eg. 1F frequency vectors. In this way a matrix of divergences is obtained, and a search for local maximums, which are finally considered as belonging to edge points, ends the algorithm.

Thus, the method locally examines histograms of grey levels in the neighbourhood of each pixel, ignoring prior characteristics of the regions to be compared. Both the behaviour of Jensen-Shannon divergence and the suitability of the operating method with regard to the appropriate choice of parameters explain the good results that were obtained in most cases.

1.3. Contents

In the present work, we perform a theoretic analysis—as thoroughly as possible—on the variations of this function under the operating conditions of the method, as well as the fate derived from the random spatial configuration of the portions delineated by the window. Our introductory work on this method [6] provides a basic description and a preliminary theoretic analysis. This preparatory work is described here almost in its entirety to facilitate comprehension of the present work.

In principle, the mathematical properties of Jensen-Shannon divergence can be studied theoretically with the usual analytical methods, although some of the properties or meanings—such as that of the probabilistic distance—prove arduous and thus are not presented here. This work concentrates on the theoretical study of the usefulness of this measure for edge detection in digital images. This objective becomes especially difficult on posing the problem of statistical confidence of the method while taking into account the random spatial configuration, compatible with the relative-frequency data, which are the arguments of the divergence function. For this, the results of the present work constitute an approximation that necessarily resorts on occasion to Monte Carlo simulations when the analytical pathway proves impractical.

The study starts necessarily with a preliminary analysis of the behaviour of Jensen-Shannon divergence as an edge-detecting function under ideal conditions, in order to examine later what degree of detection remains successful under real conditions. In addition, the study contains propositions that are valid in a more general context than outlined here; for example, the discussion on the objectivity of an edge in a given image is independent of the detection method, and the problem of the changes of phase is applicable to any technique based on the comparison of histograms in a divided sliding window.

2. Some Properties of the Jensen-Shannon Divergence

2.1. The Shannon Entropy

This is a classic concept in Information Theory [8], but for the sake of self-containing it is briefly described here.

Let $\boldsymbol{P} = \{p_j,\ j = 1, \ldots, K\}$ be a discrete probability distribution ($0 \leq p_j \leq 1\, \forall j, \sum_{j=1}^{K} p_j = 1$). The Shannon entropy of $\boldsymbol{P}$ is defined as

$$H(\boldsymbol{P}) = -\sum_{j=1}^{K} p_j \log p_j \geq 0,$$

where $0 \log 0 = 0$ is assumed by continuity. In a certain sense, $H(\boldsymbol{P})$ measures the diversity of events distributed in $\boldsymbol{P}$. $H(\boldsymbol{P})$ is zero only for degenerate distributions (one sure event), and is maximum for the uniform distribution (equiprobable events). Some basic properties are:

1. $H(\boldsymbol{P}) \geq 0\ \forall \boldsymbol{P}$,
2. $H(\boldsymbol{P}) = 0$ if and only if $\boldsymbol{P}$ is a degenerate distribution (one sure event),
3. $H(p_1, \ldots, p_n, 0) = H(p_1, \ldots, p_n)$,
4. $H(\boldsymbol{P}) \leq H(\frac{1}{n}, \ldots, \frac{1}{n})\ \forall \boldsymbol{P}$.
5. (Branching property) Let $\Pi = (\pi_{kj}), \pi_{kj} \geq 0$, $\sum_{k=1}^{n} \sum_{j=1}^{m} \pi_{kj} = 1$ be a compound probability

distribution. Denote by $\boldsymbol{P} = (p_k)$, $p_k = \sum_{j=1}^{m} \pi_{kj}$ and $\boldsymbol{Q} = (q_j)$, $q_j = \sum_{k=1}^{n} \pi_{kj}$ the corresponding side distributions, and by $\boldsymbol{P}_j = (p_{jk})$, $p_{jk} = \frac{\pi_{kj}}{q_j}$ and $\boldsymbol{Q}_k = (q_{kj})$, $q_{kj} = \frac{\pi_{kj}}{p_k}$ the associated distributions of conditional probabilities. Then

$$H(\Pi) = H(\boldsymbol{P}) + \sum_{j=1}^{m} p_j H(\boldsymbol{Q}_j)$$
$$= H(\boldsymbol{Q}) + \sum_{k=1}^{n} q_k H(\boldsymbol{P}_k).$$

As a consequence, if Π comes from two independent probabilistic experiments, then $\boldsymbol{P} = \boldsymbol{P}_j \; \forall j$, $\boldsymbol{Q} = \boldsymbol{Q}_k \; \forall k$, and thus $H(\Pi) = H(\boldsymbol{P}) + H(\boldsymbol{Q})$.

Proofs of all these and other properties can be found in [8].

2.2. *Definitions*

The *generalised Jensen-Shannon divergence* is a measure of (inverse) cohesion between probability distributions:

$$\mathrm{JS}_\pi(\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_n) = H\left(\sum_{i=1}^{n} \pi_i \cdot \boldsymbol{P}_i\right) - \sum_{i=1}^{n} \pi_i \cdot H(\boldsymbol{P}_i)$$

where $\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_n$ are probability distributions, $\boldsymbol{P}_i = \{p_{ij},\ j = 1, \ldots, K\}$, $i = 1, \ldots, n$, $\pi = \{\pi_1, \pi_2, \ldots, \pi_n \mid \pi_i > 0, \Sigma\pi_i = 1\}$ is a set of weights. When $\pi = \{1/n, 1/n, \ldots, 1/n\}$, that is, when all weights are equal, then JS_π is plainly denoted by JS.

The two terms in the right-hand side of the above equation are the entropy of the π-linear convex combination $\boldsymbol{P} = \Sigma\pi_i \boldsymbol{P}_i$, and the same linear convex combination of the respective entropies. Because of the concavity of the Shannon entropy, the Jensen-Shannon divergence has several suitable properties, which are outlined by Lin [7]. Some of these are non-negativity, finiteness, boundedness and symmetry. In addition, $\mathrm{JS}_\pi = 0$ if and only if the histograms are equal.

As opposed to other measures of Information Theory, such as relative entropy [8], this has the benefit of being generalised to any number of random variables. It also presents the advantage over other measures—such as mutual information [8]—of being a weighted measure, with the possibility of weighting its arguments with the relative sizes of samples or subregions of an image. In the rest of this section, some other properties that serve to explain its application in edge detection are presented (without mathematical proofs).

Throughout the present work, some notations and abbreviations are used, as follows:

$\boldsymbol{K}$	Grey scale of the image under study; $\boldsymbol{K} = \{0, 1, \ldots, K-1\}$, $\lvert\boldsymbol{K}\rvert = K$.
$\boldsymbol{P}(D)$	Normalised histogram of a region or portion D in the image: $\boldsymbol{P}(D) = \{p_0, p_1, \ldots, p_{K-1}\}$, where p_k indicates the relative frequency of pixels with a grey level of $k \in \boldsymbol{K}$ within D. This can be considered either as a set of relative frequencies or as a probability distribution. For brevity, the notation $\boldsymbol{P}_a$ will be used instead of $\boldsymbol{P}(a)$ when dealing with regions.
W	Window that slides over the image. Centred at a given position, it indicates the pixels which must be included to perform an operation associated with this position.
W_1, W_2	The window W is divided into two disjoint semi-windows W_1, W_2 of equal size and shape, and with an unchanging relative position within W.
N	Size of each semi-window. Thus, W is $2N$ in size.
$H(D)$	The Shannon entropy of the histogram of a portion: $H(D) = H(\boldsymbol{P}(D))$.
$\mathrm{JS}(W)$	Jensen-Shannon divergence of the window W: $\mathrm{JS}(W) = \mathrm{JS}(\boldsymbol{P}(W_1), \boldsymbol{P}(W_2))$. It is evident that its value varies with the position of W over the image.

2.3. *Bounds of Jensen-Shannon Divergence*

The JS_π divergence is an upper- and lower-bounded function: $0 \leq \mathrm{JS}_\pi \leq C(\pi)$. Is has been have previously demonstrated [11] that JS_π is a convex function and, therefore, reaches its maximum value $C(\pi)$ for a certain set of distributions, all degenerate. In addition, for certain cases, the value of $C(\pi)$ was derived: if $n \leq K$ then $C(\pi) = H(\pi)$, and if $\pi = \{1/n, 1/n, \ldots, 1/n\}$ then

$$C(\pi) = \log(n) - (r \cdot (q+1) \cdot \log(q+1) + (K - r) \cdot q \cdot \log(q))/n,$$

with q and r being the quotient and remainder, respectively, on dividing K by n. In any case, it always holds

that $C(\pi) \leq \min\{\log n, \log K\}$, and thus JS_π can be normalised between 0 and 1 with each application.

2.4. Grouping of the Grey Scale

The crucial branching property of Shannon entropy determines a similar divergence property, by which divergences corresponding to two grey scales can be related when one is a grouping of the other.

When images with a fine grey scale are treated with small windows, the number N of pixels in a semi-window proves much lower than the number K of grey levels in the scale. This gives rise to frequency vectors with many zeros, and, on calculating the divergence between two histograms under these conditions, we may find that the zeros of the two histograms do not coincide, or, more properly, that these histograms are orthogonal. Then a saturation effect of divergence arises, towards its upper bound ($\log 2 = 1$ in our case), hampering the detection of significant maximums. Fortunately, some results favour a certain robustness of the divergence against this phenomenon, as we shall see below.

Let us assume that the original grey scale $\boldsymbol{K} = \{0, 1, \ldots, K-1\}$ is branched to $\boldsymbol{K}' = \{0, 1, \ldots, K'-1\}$ in such a way that each level $k \in \boldsymbol{K}$ in the original scale corresponds to an entire variety or range of consecutive levels $\boldsymbol{R}_k = \{k'_k, k'_k + 1, \ldots, k'_k + n_k\}$ in the branched scale, and reciprocally each level $k' \in \boldsymbol{K}'$ of the branched scale comes from a single level $\kappa(k') \in \boldsymbol{K}$ of the original scale. Clearly, so that the branching is consistent, the set $\{\boldsymbol{R}_0, \boldsymbol{R}_1, \ldots, \boldsymbol{R}_{K-1}\}$ of ranges must be a partition of $\boldsymbol{K}'$. Let $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ be the respective histograms of the semi-windows W_1 and W_2 for a fixed position of the sliding window W within a given image. If we now repeat the observation of W using the branched scale (without varying the position of the sliding window), then we obtain two branched histograms $\boldsymbol{P}'_1$, $\boldsymbol{P}'_2$. These are related to $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ by means of the relationships $p_{1,k} = p'_{1,k'_k} + p'_{1,k'_k+1} + \cdots + p'_{1,k'_k+n_k}$ or, reciprocally, $p'_{1,k'} = q_{1,k'}\, p_{1,\kappa(k')}$ where $q_{1,k'} = p'_{1,k'}/p_{1,\kappa(k')}$ is the relative frequency of appearance of the k' level in the pixels of W_1 with the original level $\kappa(k') \in \boldsymbol{K}$, and verifies $q_{1,k'} \in [0,1]\, \forall k' \in \boldsymbol{K}'$, $\sum_{j=0}^{n_k} q_{1,k'_k+j} = 1\, \forall k \in \boldsymbol{K}$, giving comparable relationships for the $\boldsymbol{P}'_2$ histogram.

Thus, for each original grey level k, there are two associated probability distributions $\boldsymbol{Q}_{1,k}, \boldsymbol{Q}_{2,k}$ with $\boldsymbol{Q}_{1,k} = \{q_{1,k'_k}, q_{1,k'_k+1}, \ldots, q_{1,k'_k+n_k}\}$ and analogously for $\boldsymbol{Q}_{2,k}$. In terms of information theory, we are broadening the *alphabet*—in our case, the grey scale—for which we know the probability distributions of the broadened alphabet with respect to the original alphabet.

Theorem. *Under the above-described conditions, we have*:

$$\mathrm{JS}(\boldsymbol{P}'_1, \boldsymbol{P}'_2) = \mathrm{JS}(\boldsymbol{P}_1, \boldsymbol{P}_2) + \sum_{k=0}^{K-1} \frac{p_{1,k} + p_{2,k}}{2} \times \mathrm{JS}_{\omega_k}(\boldsymbol{Q}_{1,k}, \boldsymbol{Q}_{2,k})$$

where JS_{ω_k} *represents the Jensen-Shannon divergence between the two probability distributions, weighted with weight* $\omega_k = \{\omega_{1,k}, \omega_{2,k}\} = \{\frac{p_{1,k}}{p_{1,k}+p_{2,k}}, \frac{p_{2,k}}{p_{1,k}+p_{2,k}}\}$.

Proof: Let $\boldsymbol{P} = (\boldsymbol{P}_1 + \boldsymbol{P}_2)/2$ and $\boldsymbol{Q}_k = \omega_{1,k}\boldsymbol{Q}_{1,k} + \omega_{2,k}\boldsymbol{Q}_{2,k}$. Developing the left member of the equation and applying the branching property of Shannon entropy, we can write

$$\begin{aligned}
\mathrm{JS}(\boldsymbol{P}'_1, \boldsymbol{P}'_2) &= H((\boldsymbol{P}'_1 + \boldsymbol{P}'_1)/2) - (H(\boldsymbol{P}'_1) + H(\boldsymbol{P}'_2))/2 \\
&= H(\boldsymbol{P}) + \sum_{k=0}^{K-1} p_k H(\boldsymbol{Q}_k) \\
&\quad - \frac{1}{2}\left(H(\boldsymbol{P}_1) + \sum_{k=0}^{K-1} p_{1,k} H(\boldsymbol{Q}_{1,k})\right) \\
&\quad - \frac{1}{2}\left(H(\boldsymbol{P}_2) + \sum_{k=0}^{K-1} p_{2,k} H(\boldsymbol{Q}_{2,k})\right) \\
&= \mathrm{JS}(\boldsymbol{P}_1, \boldsymbol{P}_2) + \sum_{k=0}^{K-1} p_k (H(\boldsymbol{Q}_k) \\
&\quad - \omega_{1,k} H(\boldsymbol{Q}_{1,k}) - \omega_{2,k} H(\boldsymbol{Q}_{2,k})).
\end{aligned}$$

□

Although it is not directly useful in the present context, this result can be extended immediately to weighted divergences, as follows:

Theorem. *Let* $\pi = \{\pi_1, \pi_2\}$ *be an arbitrary distribution of weights, and let* $\mathrm{JS}_\pi(\boldsymbol{P}_1, \boldsymbol{P}_2)$ *be the corresponding weighted divergence. Under the same conditions as in the preceding theorem, we have*:

$$\mathrm{JS}_\pi(\boldsymbol{P}'_1, \boldsymbol{P}'_2) = \mathrm{JS}_\pi(\boldsymbol{P}_1, \boldsymbol{P}_2) + \sum_{k=0}^{K-1} (\pi_1 p_{1,k} + \pi_2 p_{2,k}) \times \mathrm{JS}_{\pi\omega_k}(\boldsymbol{Q}_{1,k}, \boldsymbol{Q}_{2,k})$$

where now $\pi\omega_k = \{\frac{\pi_1 p_{1,k}}{\pi_1 p_{1,k}+\pi_2 p_{2,k}}, \frac{\pi_2 p_{2,k}}{\pi_1 p_{1,k}+\pi_2 p_{2,k}}\}$.

Proof: This is achieved by direct calculation in the same way as that of the previous theorem, which now remains as a particular case of this latter theorem, for $\pi = \{1/2, 1/2\}$. □

From these results, it follows that the increase of divergence $\Delta\mathrm{JS} = \sum_{k=0}^{K-1} p_k \mathrm{JS}_{\omega_k}(\boldsymbol{Q}_{1,k}, \boldsymbol{Q}_{2,k})$ resulting from the branching of the grey scale depends on the branching distributions of the scale $\boldsymbol{Q}_{1,k}$, $\boldsymbol{Q}_{2,k}$ and of the histograms. Consequently, this increase is null only in the case that for each grey level k we have $\boldsymbol{Q}_{1,k} = \boldsymbol{Q}_{2,k}$ or that the weights in ω_k are degenerate—that is, one is null and the other is the unity. In our case, the interpretation is that the grey levels appearing simultaneously in both semi-windows must be expanded to branched levels following the same common distribution, without concern for the expansion of levels that appear only in one of the semi-windows but not in the other. If this condition is fulfilled, then the change of scale does not trigger variation in the Jensen-Shannon divergence.

Non-fulfilment of this condition may result from various factors. For example, observation at a poor spatial resolution would obscure tiny details which otherwise could add a certain chromatic richness. Another plausible case is that of an image that was acquired with reduced grey scales and that were stored at a fine scale, giving rise to a histogram containing skips and gaps with null frequency. If we now process the image in some form that slightly alters its brightness by regions, a secondary effect may arise as a strong increase in Jensen-Shannon divergence.

In conclusion, we can say that, as a general rule, it is advisable to take the most reduced grey scales possible, so long as this does not notably sacrifice detail.

2.5. Rationale for the Use of Jensen-Shannon Divergence

Certain aspects of Jensen-Shannon divergence make it the most advantageous measure, compared to other entropic measures of information theory, such as mutual information $I(X, Y)$ or relative entropy between two random variables. Some of these aspects are explained below.

Firstly, we must assume that the images of natural or artificial scenes are composed of regions which, in representing various objects, are different and independent while internally uniform. However, the presence of a correlation between samples from the semi-windows of a window does not indicate in general that both semi-windows are located in the same region. Consequently, we need not seek a measure of dependence or correlation (such as mutual information), but rather a measure of probabilistic distance which indicates whether or not the samples are plainly different.

The meaning of Jensen-Shannon divergence between probability distributions of two random variables $\mathrm{JS}(\boldsymbol{P}(X), \boldsymbol{P}(Y))$ is essentially different from that of the mutual information $I(X, Y)$ between the same variables. The joint random variable $X \otimes Y$ is also an argument of $I(X, Y)$, which implies extra information that makes the mutual information become a measure of statistical dependence between random variables. On the contrary, Jensen-Shannon divergence can be considered as a measure of pseudo-distance (e.g. it does not satisfy the triangle inequality). When applied to any number of distributions, it constitutes a certain measure of distance between all of them (or reverse cohesion). The requirement of absolute continuity does not have to be satisfied by the probability distributions involved, as it has to be by the relative entropy. On the other hand, the relative entropy is not symmetrical, has no upper bound, and does not have an adequate branching property.

The above set of properties makes the Jensen-Shannon divergence an especially adequate measure in edge detection.

3. Edge Detection Under Conditions of Uniformity

The behaviour of Jensen-Shannon divergence as an edge-detection tool is analysed initially under the ideal assumption that each region of an image satisfies the following *statistical-uniformity hypothesis*:

> "For every sample D taken from a region x, we have $\boldsymbol{P}(D) = \boldsymbol{P}_x$".

It might appear that this hypothesis is overly restrictive, from the standpoint of rigour, given that it is fulfilled only in the case of regions having pixels all of the same grey level. Nevertheless, for textured regions, the hypothesis grows asymptotically closer to fulfilment as the size of the sample increases and substantially exceeds the size of the *texel*.[1] The foregoing also occurs for regions originally uniform that undergo noise corruption, either impulsive or additive. Given that this approximate fulfilment is feasible and near the ideal case discussed, the study proceeds under the uniformity

hypothesis, without diminishing the realistic analysis of the following section.

A consequence of (but not an equivalent to) the uniformity hypothesis is the following property:

> "If a window W is moved through the interior of a region a, then its histogram $\boldsymbol{P}(W)$ remains constant and equal to the histogram $\boldsymbol{P}_a$ of the region".

Let us now consider a rectangular window W, divided into two semi-windows W_1 and W_2, which slides with W_1 leading and goes from being completely within region a to being completely inside region b, crossing the straight border that separates the two regions. At this point, the idealised analysis is made considering two cases, depending on the direction of the sliding: 1st, normal to the edge; 2nd, oblique.

3.1. *First Case: A Straight Edge Perpendicularly Crossed by the Window*

This case is partially described in [6], although the essentials are reproduced here for the reader's convenience.

While window W occupies a single region, the histograms $\boldsymbol{P}(W_1), \boldsymbol{P}(W_2)$ and $\boldsymbol{P}(W)$ are invariable and equal, in accordance with the uniformity hypothesis. As a result, the divergence $\mathrm{JS}(W)$ is null in this situation.

If the window is situated in any position on the edge, at most one of the two semi-windows will be involved in the two regions. Without diminishing generality, let us assume that we are dealing with W_1. Again, under the uniformity hypothesis, the part of W_1 situated in regions a and b contain partial histograms $\boldsymbol{P}_a$ and $\boldsymbol{P}_b$ respectively, and therefore $\boldsymbol{P}(W_1)$ will be a certain weighted mean or convex linear combination of $\boldsymbol{P}_a$ and $\boldsymbol{P}_b$, with weights quite proportional to the sizes of the sub-regions $a \cap W_1$ and $b \cap W_1$, respectively. More specifically, if $\beta \in [0, 1]$ represents the fraction of W_1 included in region b, then $\boldsymbol{P}(W_1) = (1-\beta)\boldsymbol{P}_a + \beta\boldsymbol{P}_b$, consequently giving rise to a histogram that varies with respect to the position of W. As W_2 lies within region a, we have $\boldsymbol{P}(W_2) = \boldsymbol{P}_a$ and therefore $\boldsymbol{P}(W) = (1-\frac{1}{2}\beta)\boldsymbol{P}_a + \frac{1}{2}\beta\boldsymbol{P}_b$. It is helpful now to use $W(\beta)$ and $W_1(\beta)$ to denote the particular compositions of W and W_1 respectively, for the position β. The corresponding Jensen-Shannon divergence can be expressed as a function of β as

$$\mathrm{JS}(W(\beta)) = H(W(\beta)) - \frac{1}{2}(H(W_1(\beta)) + H(W_2)), \quad \beta \in [0, 1].$$

Theorem. "*Under the foregoing conditions, if* $\boldsymbol{P}_a \neq \boldsymbol{P}_b$ *then* $\mathrm{JS}(W(\beta))$ *reaches its maximum value when* $\beta = 1$".

Proof: If we note $\boldsymbol{P}_a = \{p_{ai}\}_{i=0}^{K-1}$, $\boldsymbol{P}_b = \{p_{bi}\}_{i=0}^{K-1}$ and $d_i = p_{bi} - p_{ai}$, some straightforward calculations enable us to express

$$\begin{aligned}\mathrm{JS}(W(\beta)) = -\frac{1}{2}\sum_{i=0}^{K-1} & 2\left(p_{ai} + \frac{\beta}{2}d_i\right)\log\left(p_{ai} + \frac{\beta}{2}d_i\right) \\ & - (p_{ai} + \beta d_i)\log(p_{ai} + \beta d_i) \\ & - p_{ai}\log p_{ai}.\end{aligned}$$

To prove that $\mathrm{JS}(W(\beta))$ is a strictly increasing function with respect to β, it is sufficient to see that its derivative function

$$\begin{aligned}&\frac{dJS(W(\beta))}{d\beta} \\ &\quad = -\frac{1}{2}\sum_{i=0}^{K-1} d_i\left(\log\left(p_{ai} + \frac{\beta}{2}d_i\right) - \log(p_{ai} + \beta d_i)\right)\end{aligned}$$

vanishes for $\beta = 0$, and is strictly increasing (and therefore positive) with respect to β, by virtue of the fact that its second derivative

$$\frac{d^2\mathrm{JS}(W(\beta))}{(d\beta)^2} = \frac{1}{2\ln 2}\sum_{i=0}^{K-1}\frac{p_{ai}d_i^2}{\left(p_{ai} + \frac{\beta}{2}d_i\right)(p_{ai} + \beta d_i)}$$

is strictly positive $\forall\beta \in [0, 1]$; in fact, all the addends are non-negative—this being the reason for which $d^2\mathrm{JS}(W(\beta))/(d\beta)^2$ can become zero only if all the addends do. Under this assumption, it cannot occur that $d_i = 0\ \forall i$ after assuming two regions of different histograms. Therefore, for those d_i not equal to 0, p_{ai} would have to equal 0, and this is impossible, given that we are working with probability distributions. □

The divergence as a function of β is shown graphically in the zone corresponding to the interval $[0, 1]$ of the abscissa axis of Fig. 1, for $\boldsymbol{P}_a = \{0.5, 0.5\}$, $\boldsymbol{P}_b =$

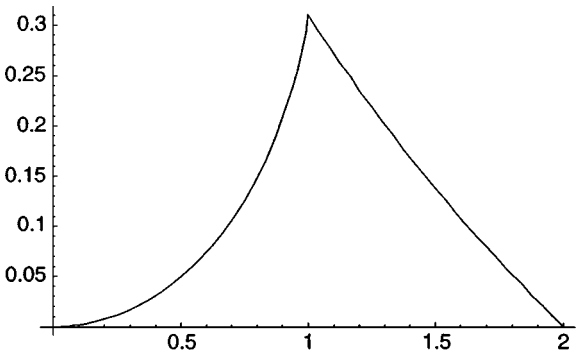

Figure 1. Jensen-Shannon divergence for a window that crosses a vertical border.

{0, 1}. The rest of the figure (interval [1, 2]) shows the divergence when the semi-window W_2 is entering b, W_1 now lying completely within b.

It bears emphasising that divergence is not a derivable function with respect to the position of W in the *centred* position, at which each semi-window lies in a different region. We can corroborate this fact by means of the same proof, deducing the impossibility of a null derivative, as illustrated in Fig. 1, where the maximum appears clearly at a point where JS is not derivable. It is noteworthy, furthermore, that the divergence is not in general symmetrical with respect to the position of the maximum. Although the figure need not be symmetrical, we can affirm that it is more asymmetrical the greater the differences are among its entropies $H(\boldsymbol{P}_a)$ and $H(\boldsymbol{P}_b)$. In the case of Fig. 1, we chose histograms of strongly differing entropies in order to show the asymmetry of the divergence.

3.2. *Second Case: A Straight Edge Obliquely Crossed by the Window*

The usefulness of studying this second case, with regard to the previous one, lies in the fact that an edge-detection algorithm cannot generally guarantee that the window will cross an edge at a perpendicular angle. Given that, edge-detection algorithms usually use several simultaneous orientations of the sliding window (typically two or four), the worst case being an edge deviation of $\pm 45^\circ$ with respect to the sliding direction. This forces us to consider, for each semi-window, a (variable) fraction of incidence in each region. Let β_1 be the fraction of the semi-window W_1 in region a, and β_2 be the fraction of the semi-window W_2 in region b (see Fig. 2). In this way, the histograms of the semi-windows W_1 and W_2 are $\boldsymbol{P}(W_1) = \beta_1 \boldsymbol{P}_a + (1-\beta_1)\boldsymbol{P}_b$ and $\boldsymbol{P}(W_2) = (1-\beta_2)\boldsymbol{P}_a + \beta_2 \boldsymbol{P}_b$, respectively. It is evident that β_1 and β_2 are not independent variables, since both depend directly on the position of W with respect to the edge. To study the variation of the Jensen-Shannon divergence as a function of β_1 and β_2, we shall distinguish three cases:

a) W_2 has not yet entered b: $1 \geq \beta_1 \geq \beta_0$, $\beta_2 = 0$;
b) W_2 has entered b, but W_1 has not yet left a: $\sqrt{\beta_1} + \sqrt{\beta_2} = \sqrt{\beta_0}$;
c) W_1 has already left a: $\beta_1 = 0$, $\beta_0 \leq \beta_2 \leq 1$;

where β_0 is the fraction of the area of the semi-window corresponding the figure of a right-angled triangle, the sides of the semi-window being the sides of the right angle, and the hypotenuse coinciding with the edge.

It is evident that a) and c) are parallel, and thus it is sufficient to study only one of them. For the case of a), the Jensen-Shannon divergence depends only on β_1, and it strictly decreases with β_1 (increasing with the movement of W from left to right), as deduced directly from the fact that its second derivative

$$\frac{\partial^2 \mathrm{JS}}{(\partial \beta_1)^2}(\beta_1, 0) = \frac{1}{2 \ln 2} \sum_{i=0}^{K-1} \times \frac{p_{ai}(p_{bi} - p_{ai})^2}{(\beta_1 p_{ai} + (1-\beta_1) p_{bi})(p_{ai} + \beta_1 p_{ai} + (1-\beta_1) p_{bi})}$$

is strictly positive $\forall \beta_1 \in [0, 1]$ and that the value of its first derivative at $\beta_1 = 1$ is zero, implying that $\frac{\partial \mathrm{JS}}{\partial \beta_1}(\beta_1, 0) < 0 \ \forall \beta_1 \in [0, 1)$. Applying reasoning analogous to that of case c), we find that the divergence strictly decreases with β_2 and consequently with the movement of W. Thus, we conclude that the divergence must reach its absolute maximum at some point in case b), corresponding to the positions of W in which the separation line between semi-windows intersects with the edge. This maximum is not always reached in the position of a centred window. A counterexample can be seen in Fig. 3, which shows the Jensen-Shannon divergence with respect to the sliding of W in the case of respective histograms $\boldsymbol{P}_a = \{1/3, 1/3, 1/3\}$, $\boldsymbol{P}_b = \{0, 0, 1\}$. However, such a deviation from the maximum is, in practice, negligible in most cases, as a chain of circumstances would have to arise to produce poor results, given that: 1) only two instead of four window orientations are used; 2) the edge deviates nearly 45° from both orientations; 3) the entropies of the regions

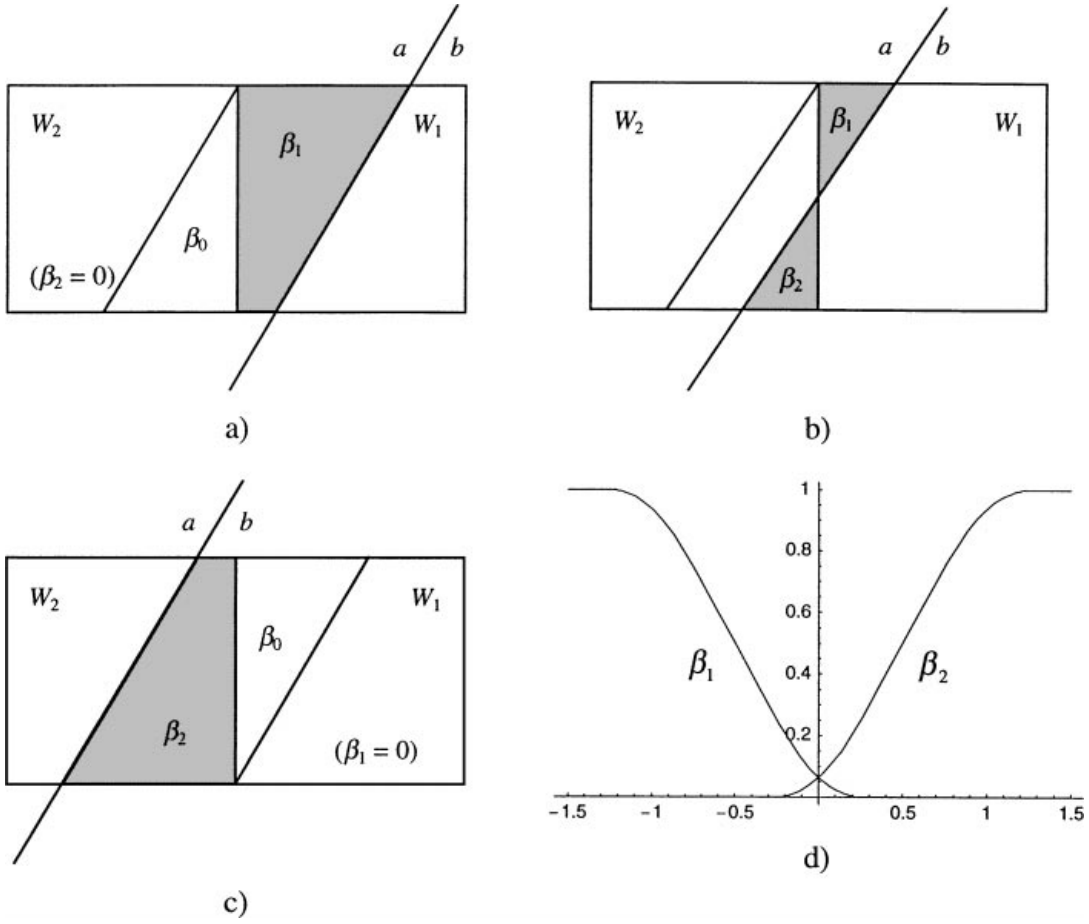

Figure 2. Variation in β_1 and β_2 when the window crosses an oblique border.

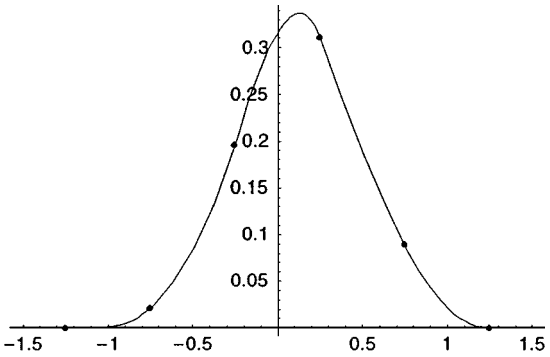

Figure 3. The Jensen-Shannon divergence according to the position of the window, on crossing an oblique edge.

strongly differ from each other (one near the upper bound and the other almost zero); 4) very large windows are being used. In the worst of cases, the deviation is always lower than $1/4$ ($1/2$ when using two orientations) of the length of the side of the semi-window.

3.3. Influence of Impulsive Noise

Here, we examine how edge detection can be influenced when an image is contaminated with impulsive noise. The problem posed consists of determining how the JS function is altered on introducing impulsive noise in variable quantities into the clean image, this being specified in position, numerical value and acuteness of the divergence maximum.

Given that we refer to a type of noise that is not spatially correlated, it is clear that, if $\boldsymbol{P}_a \neq \boldsymbol{P}_b$ in the original image, then, although the noise alters the histograms, it cannot make these coincide (on the average), except that it affects the totality of the pixels.

To illustrate this case, we shall again consider the situation described in Section 3.1. The maximum value of the divergence is $\mathrm{JS}(W(1)) = \mathrm{JS}(\boldsymbol{P}_a, \boldsymbol{P}_b)$ in the absence of noise. If we now consider that a source of noise with distribution $\boldsymbol{P}_r = \{r_0, r_1, \ldots, r_{K-1}\}$ is acting on the image with a probability $\alpha \in [0, 1]$, then the altered

histograms are (on the average) $\tilde{\boldsymbol{P}}_a = (1-\alpha)\boldsymbol{P}_a + \alpha\boldsymbol{P}_r$ and $\tilde{\boldsymbol{P}}_b = (1-\alpha)\boldsymbol{P}_b + \alpha\boldsymbol{P}_r$. The disturbed divergence $\mathrm{JS}(\tilde{\boldsymbol{P}}_a, \tilde{\boldsymbol{P}}_a)$ will always have a lower value than that of $\mathrm{JS}(\boldsymbol{P}_a, \boldsymbol{P}_b)$, as can easily be deduced on confirming that $\frac{\partial^2 \mathrm{JS}(\tilde{\boldsymbol{P}}_a, \tilde{\boldsymbol{P}}_a)}{(\partial\alpha)^2} \geq 0 \,\forall \alpha \in [0,1]$ and $\frac{\partial \mathrm{JS}(\tilde{\boldsymbol{P}}_a, \tilde{\boldsymbol{P}}_a)}{\partial\alpha}|_{\alpha=1} = 0$, implying that $\mathrm{JS}(\tilde{\boldsymbol{P}}_a, \tilde{\boldsymbol{P}}_a)$ decreases with respect to α.

The immediate result is that, despite that a maximum of JS still remains at the same position and the edge is detected at the correct position (in terms of statistical mean), the presence of noise decreases the maximum value of the divergence and consequently attenuates the slope. The former consequence affects the methods that use a threshold of the divergence, while the latter diminishes the robustness of the methods based on the detection of the divergence maximums, on making them more sensitive to the fluctuations to be studied in the following section.

4. Relaxation of the Uniformity Hypothesis: Statistical Study

The uniformity hypothesis formulated in the preceding section represents an unattainable ideal situation, and can be approximated only if certain conditions are fulfilled. Non-fulfilment of this hypothesis should be studied, since it implies an approximation to reality, a connection between continuousness and discreteness. In an image composed of pixels, a portion clearly does not generally have to present the same histogram as the region that contains it.

To make the following analysis, we must at least assume that there is an edge between regions which is perfect and objectively determined. Aside from the problem (troublesome and insurmountable) that all pixels marked as belonging to an edge necessarily lie in one of the regions, thus causing a certain ambiguity, the position of the edge could be questioned for other reasons, thereby preparing for the following subsection prior to the continuation of the analysis.

4.1. Objectively Undetermined Edge

Before studying the consequences of the absence of statistical uniformity, in this subsection we shall consider the special situation, generally improbable, in which the edge is not objectively determined, this occurring for certain configurations of adjacent texels. This lack of determination arises when there are two or more consecutive positions of window W giving the same pair of histograms observed $\boldsymbol{P}(W_1)$ and $\boldsymbol{P}(W_2)$. Any histogram based method (not only that based on the Jensen-Shannon divergence) using a sliding window would detect a multiple edge. To simplify the setting up of the problem, let us consider the following one-dimensional signal:

0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1

which admits the three following proposals to position the border:

0 0 1 0 0 1 0 0 1 0 0 1 0 0 | 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1

0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 | 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1

0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 | 1 1 0 1 1 0 1 1 0 1 1 0 1 1.

The usefulness of studying these problems of objective determination is evident because of the following points:

1. The images chosen as a test to evaluate the edge-detection methods should be free of undetermined edges.
2. The presence of an edge of double or greater width in an image may be due to indetermination, in which case any procedure of narrowing would produce a simple edge in an arbitrary position.

Below, we perform a calculation of the appearance probability of an undetermined edge. For simplification, let us assume that there is a straight vertical edge between regions of regular texture composed of texels of identical shape and size $m \times n$ (m rows, n columns) in both regions. Let $a_{ij} \in [0, \ldots, K-1]$ be the elements of the texel T_a of the region a, and analogously $b_{ij} \in [0, \ldots, K-1]$ the elements of the texel T_b of the region b (see Fig. 4). Let us call the corresponding columns $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ and $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n$.

In the hypothetical (simplified) situation in which the pixels of each region are identically and independently distributed with the probability distribution given by the histogram appropriate to its region, we have:

$$
\begin{aligned}
&\Pr\{\text{ The edge is objectively indeterminate }\} \\
&\quad = 1 - \Pr\{\boldsymbol{a}_1 \neq \boldsymbol{b}_1 \text{ and } \boldsymbol{a}_n \neq \boldsymbol{b}_n\} \\
&\quad = 1 - \Pr\{\boldsymbol{a}_1 \neq \boldsymbol{b}_1\}\Pr\{\boldsymbol{a}_n \neq \boldsymbol{b}_n\} \\
&\quad = 1 - (1 - \Pr\{\boldsymbol{a}_1 = \boldsymbol{b}_1\})(1 - \Pr\{\boldsymbol{a}_n = \boldsymbol{b}_n\}) \\
&\quad = 1 - \left(1 - \langle \boldsymbol{P}_a, \boldsymbol{P}_b \rangle^m\right)^2
\end{aligned}
$$

Region a				Region b			
a_{11}	a_{12}	…	a_{1n}	b_{11}	b_{12}	…	b_{1n}
a_{21}	a_{22}	…	a_{2n}	b_{21}	b_{22}	…	b_{2n}
⋮	⋮	⋱	⋮	⋮	⋮	⋱	⋮
a_{m1}	a_{m2}	…	a_{mn}	b_{m1}	b_{m2}	…	b_{mn}
↓	↓		↓	↓	↓		↓
$\boldsymbol{a}_1$	$\boldsymbol{a}_2$	…	$\boldsymbol{a}_n$	$\boldsymbol{b}_1$	$\boldsymbol{b}_2$	…	$\boldsymbol{b}_n$

Figure 4. Positioning of the texels for the analysis of the indetermination.

where $\langle \boldsymbol{P}_a, \boldsymbol{P}_b \rangle = \sum_{i=0}^{K-1} p_{ai} p_{bi}$ is the probability that two randomly chosen pixels, one from the region a and the other from b, have the same grey level.

This expression of the probability of the existence of objectively indeterminate edges is not completely exact, given that in addition to all the simplified hypotheses (especially that of i.i.d.), the situation may occur—although the possibility is extremely remote—that the texels are exactly equal, with the subsequent absolute absence of an edge. To correct this, the expression should stand as

$$\begin{aligned}&\Pr\{\text{ The edge is objectively indeterminate }\}\\&\quad = 1 - \left(1 - \langle \boldsymbol{P}_a, \boldsymbol{P}_b \rangle^m\right)^2 - \langle \boldsymbol{P}_a, \boldsymbol{P}_b \rangle^{mn}\\&\quad = \langle \boldsymbol{P}_a, \boldsymbol{P}_b \rangle^m \left(2 - \langle \boldsymbol{P}_a, \boldsymbol{P}_b \rangle^m - \langle \boldsymbol{P}_a, \boldsymbol{P}_b \rangle^{m(n-1)}\right).\end{aligned}$$

For the sake of illustration, let us assume that regions a and b both have uniform histograms. The probability of an objectively indeterminate edge is

$$\frac{1}{K^m}\left(2 - \frac{1}{K^m} - \frac{1}{K^{m(n-1)}}\right),$$

which clearly tends to zero when $1/K^m$ does. We can affirm as a general conclusion that the influence of the cases of objectively indeterminate edges is negligible only when the texels are of small size and the inner product of the histograms is high.

If the assumptions in the idealised study of Section 3 do not hold, then clearly the probability that the maximum of JS is not in the edge position is greater than otherwise. We shall now study the situations that can bring about this undesirable condition.

Let us restrict our study to images having regularly textured regions, each represented by its texel. Assuming that the edge is objectively determined (that is, its position is statistically distinguishable from any other), we analyse the possibility that maximum Jensen-Shannon divergence occurs in a non-centred position of the window. This may happen for certain spatial configurations of the two adjacent texels, due to the fact that, in general, the texel fractions included in the semi-windows present histograms different from those of complete texels, and therefore of those of the regions. This problem has combinatory and statistical aspects that are difficult to treat, and we shall refer to this problem as JS *fluctuation*.

4.2. Change of Phase

Let us consider the following particular case. Let us assume that an image has been constructed by establishing first an arbitrary edge and afterwards covering the two regions with the same texture, but with a relative arbitrary displacement. The resulting situation, which we call a change of phase, can objectively show a given border, even by simple observation, between two identically textured regions. This determines the possible appearance of non-null JS values at the shifted positions, while the divergence is zero at the exact position of the edge,[2] thereby impeding a correct detection. The presence of a change of phase would cause two JS maximums (not necessarily equal), one at each side of the edge. Therefore, the application of these methods to a textured image having an edge with a change of phase should produce a double line containing this edge. In practice, this is not guaranteed because of other statistical fluctuations of the spatial configuration. In addition, the reverse is not assured, either; the presence of a double line in the result is not a true indication of an edge with a change of phase, although the probability that it is may be high. The example shown in Fig. 5 is the simplest possible to illustrate the JS variations for a change of phase. This is a one-dimensional binary signal for which the texel is 01, with the change of phase displayed.

As shown in the figure, on sliding a window $2N = 2 + 2$ in size, we get a null divergence value in all positions of the window, *including the centred position*, except in the positions adjacent to the edge, in which the value is JS $= 0.31128$.

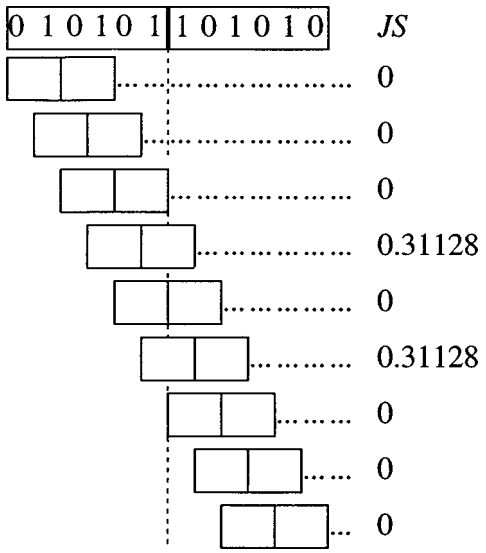

Figure 5. An example of an objectively determined edge for which the JS divergence produces a double edge.

Despite the simplifications due to the equality of texels, the JS distribution does not become sufficiently treatable. The complexity of the mathematical treatment involved in an analytical study of the problem described in order to characterise it or describe its properties, makes it enormously difficult to obtain the theoretic results. Even if the grey scale is binary—which implies a drastic simplification—the study can hardly go beyond the calculation of the probability of finding an arbitrary histogram in a semi- window crossed by an edge. This study is described in the next section. Consequently, the only plausible way to obtain the information about the importance of the problem is experimentation by a Monte Carlo simulation, the object of the following section.

4.3. Divergence Fluctuations: The Binary Case

In the general case mentioned above, in which the histogram of a texel fraction differs from the histogram of the entire texel, the problem is that the detection errors cannot be corrected, since they depend on the spatial configuration of the texel, which is unknown in the histogram based methods. Therefore, the aim is to find the JS distribution, from which the detection-error probability can be assessed.

According to the hypothesis, it can occur that only one semi-window W_i ($i = 1, 2$; it makes no difference which) overlaps an edge between regions a and b. In this way, a semi-window W_i contains a simple-texel histogram, the other a composite one. Let us study the latter. The semi-windows of which it is composed has sizes $|W_i \cap a| = (1-\beta)N$, $|W_i \cap b| = \beta N$, and the respective histograms

$$\boldsymbol{P}(W_i \cap a) = \{w_a, 1 - w_a\},$$
$$\boldsymbol{P}(W_i \cap b) = \{w_b, 1 - w_b\}$$

where (and this is precisely the reason for the present study) it needs not occur that

$$\boldsymbol{P}(W_i \cap a) = \boldsymbol{P}_a = (p_{a0}, p_{a1}),$$
$$\boldsymbol{P}(W_i \cap b) = \boldsymbol{P}_b = (p_{b0}, p_{b1}).$$

Thus, the (composite) histogram of the semi-window W_i is

$$\begin{aligned}\boldsymbol{P}(W_i) &= (1-\beta)\boldsymbol{P}(W_i \cap a) + \beta \boldsymbol{P}(W_i \cap b)\\ &= \{(1-\beta)w_a + \beta w_b, (1-\beta)(1-w_a)\\ &\quad + \beta(1-w_b)\}.\end{aligned}$$

With this, we can express the probability that $\boldsymbol{P}(W_i)$ is equal to an arbitrarily given histogram:

$$\begin{aligned}&\Pr\{\boldsymbol{P}(W_i) = \{x, 1-x\}\}\\ &\quad = \sum_w \Pr\{w_a = w\} \Pr\left\{w_b = \frac{x - (1-\beta)w}{\beta}\right\}\end{aligned}$$

with the sum extended to all the possible values of w, which are a finite quantity by virtue of the discrete nature of the grey scale and the finite number of possible partitions of W_i.

If we choose the hypergeometric distribution as the most neutral and unbiased representation of the different configurations of pixels in the regions (all are equally probable), then we can arrive without difficulty at

$$\Pr\{w_a = w\} = \frac{\binom{Np_{a0}}{(1-\beta)Nw}\binom{N(1-p_{a0})}{(1-\beta)N(1-w)}}{\binom{N}{(1-\beta)N}},$$
$$\Pr\{w_a = w^*\} = \frac{\binom{Np_{b0}}{\beta N w^*}\binom{N(1-p_{b0})}{\beta N(1-w^*)}}{\binom{N}{\beta N}};$$

expressions which can be replaced in the foregoing one with $w^* = \frac{x-(1-\beta)w}{\beta}$.

If the histogram is not binary, the above expression is given in terms of multivariate hypergeometric distribution.

The expressions obtained as sums of products of hypergeometric distributions cannot be simplified. The distribution of the Jensen-Shannon divergence, which is the objective, proves more complex still, especially due to the fact that a given value of entropy may correspond to more than one histogram, and therefore a given value of JS may also correspond to more than one pair of histograms. This lack of manageability leads us to the computer results gained by Monte Carlo simulations.

4.4. Monte Carlo Simulation

Given the breadth of the parameters involved (grey scale, size and shape of the semi-windows, sizes and shapes of the texels in each region, orientation of the edge, correlations between texels, etc.), we have chosen to conduct a series of short experiments which, though not exhaustive, can be considered representative of typical cases.

The experiments performed have started with the construction of two synthetic textures separated by a vertical rectilinear edge, in order to proceed later with the detection of this edge. All of these have the same structure, which consists of the following steps:

1. Choice of parameters of random generation of samples: shape, size and histogram of each texel.
2. Choice of parameters of edge detection: shape and size of the semi-window.
3. Generation of 10,000 valid samples and edge detection. Each sample is obtained by randomly generating the texel of each region with the specified parameters. For this, we assume the equal probability of all the possible texels. A sample is considered valid if it satisfies the following conditions:
   a) It has an objectively determined edge; the reason for this restriction is self-evident.
   b) The edge detection produces a single maximum; cases of multiple maximums hamper the rigorous evaluation of the error committed in the detection.

   Otherwise the sample is discarded.
4. Account of the results.

For all the experiments, we chose a scale of 8 grey levels, which is sufficiently broad to be representative, and sufficiently simple to allow reasonable computation time.

4.4.1. Overall Probability of Error. Here, we present the results of evaluating the probability of error in the edge detection of textured images. The problem was formulated as follows: once the parameters are fixed, to estimate the probability that maximum divergence occurs at a point on the edge of the original image. We excluded the case in which the histograms of the texels were equal, since this is studied afterwards.

Table 1 contains the numeric data of the experiments performed (A, B and C) as well as the numeric

Table 1. Results of Monte Carlo Experiments A, B and C.

	Exp. A	Exp. B	Exp. C
Parameters			
N. valid samples	10,000	10,000	10,000
Texel size	4×4	4×4	4×4
Semi-window size	4×4	7×7	4×4
Texels histograms	{0, 0, 0, 1, 13, 0, 1, 1}	{0, 0, 0, 1, 13, 0, 1, 1}	{2, 1, 1, 2, 3, 2, 3, 2}
	{2, 1, 1, 2, 3, 2, 3, 2}	{2, 1, 1, 2, 3, 2, 3, 2}	{2, 3, 2, 3, 2, 1, 1, 2}
Discards	58,236	57,370	17,957
Results (detection error)			
Ends	$-1, +1$	$-2, +2$	$-3, +3$
Mean error	$+0.0252$	-0.0618	-0.0283
Standard deviation	0.1804	0.5186	$+1.4246$

results on the distribution of the edge positions calculated in relation to the correct position. The term "discards" represent the numbers of samples that had been discarded for not being valid. Figures 6–8 graphically represent these distributions of the error in estimating the position of the edge. Below, we comment on these experiments in relation to their layout and results.

Experiment A. This experiment reveals the probability of a successful positioning of the edge when the sizes of the semi-window and texel coincide.

Figure 6 shows graphically the result of Monte Carlo simulation. In most cases, the divergence detected the edge in its correct place, and the number of times that the edge deviated was negligible. In addition, even in the few cases in which the maximum was not reached in the theoretic edge, a maximum distance of 1 pixel from the theoretic one was detected.

Experiment B. This experiment is intended to illustrate the importance of the relationship of sizes semi-window/texel. For this, we repeated Experiment A, except that now the semi-window and the texel had coprime sizes.

The graphic results of the detection can be seen in Fig. 7. Slightly more errors were made in the edge position than in the previous case, although, in general, the divergence continued to show more-than-acceptable performance.

Experiment C. This demonstrated the key role played by histograms, which in this experiment were highly similar, as opposed to the two previous cases. The graphic results are given in Fig. 8. The fact that the histograms were similar caused many more errors in edge detection than in Experiments A and B. This is logical, since in the extreme case, in which the two histograms are equal, the divergence takes a null value precisely when the window contains a semi-window in each texture (this situation will be studied in the following Section). However, this does not invalidate the use of the divergence. In fact, in most cases, the maximum is 1 or 2 pixels from the theoretic position of the edge, and in many cases the detection even gives the correct position.

4.4.2. Probability of an Unexpected Correct Estimation. In this Section, we present the results of

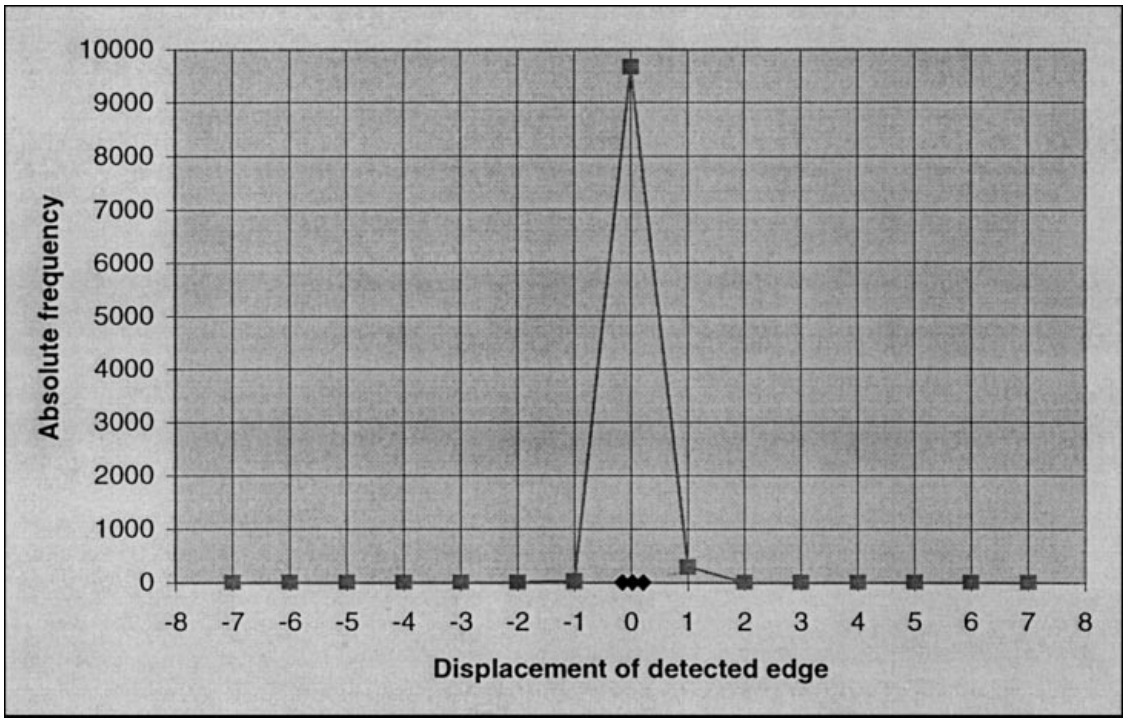

Figure 6. Results of Experiment A.

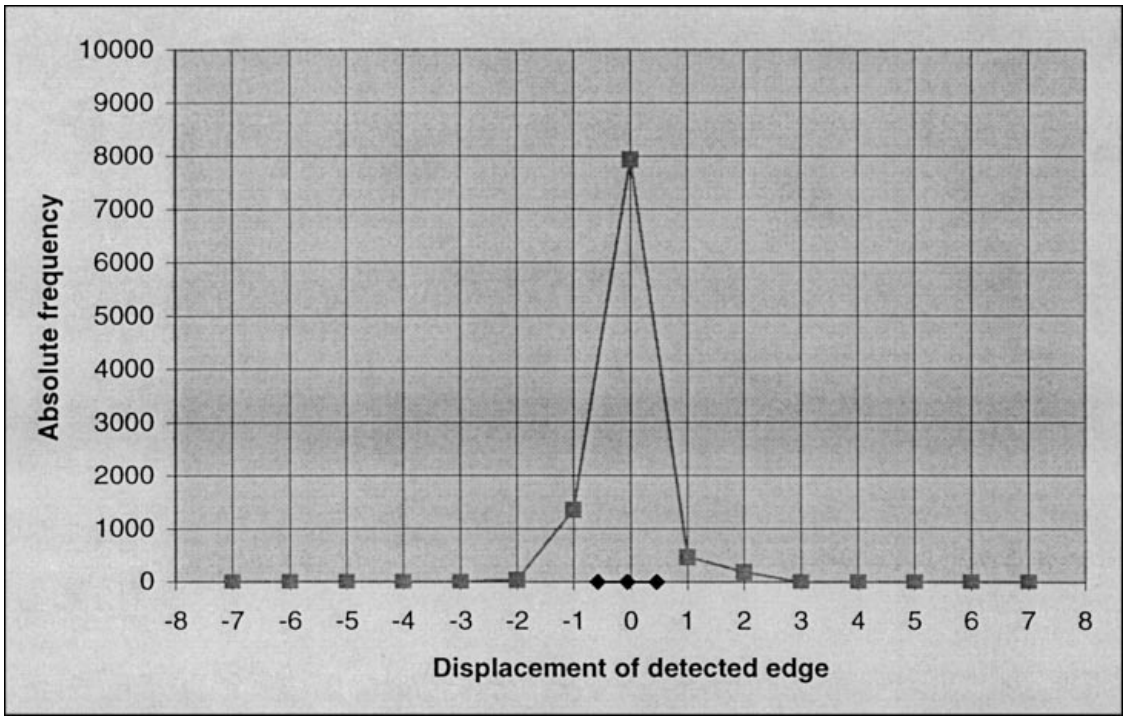

Figure 7. Results of Experiment B.

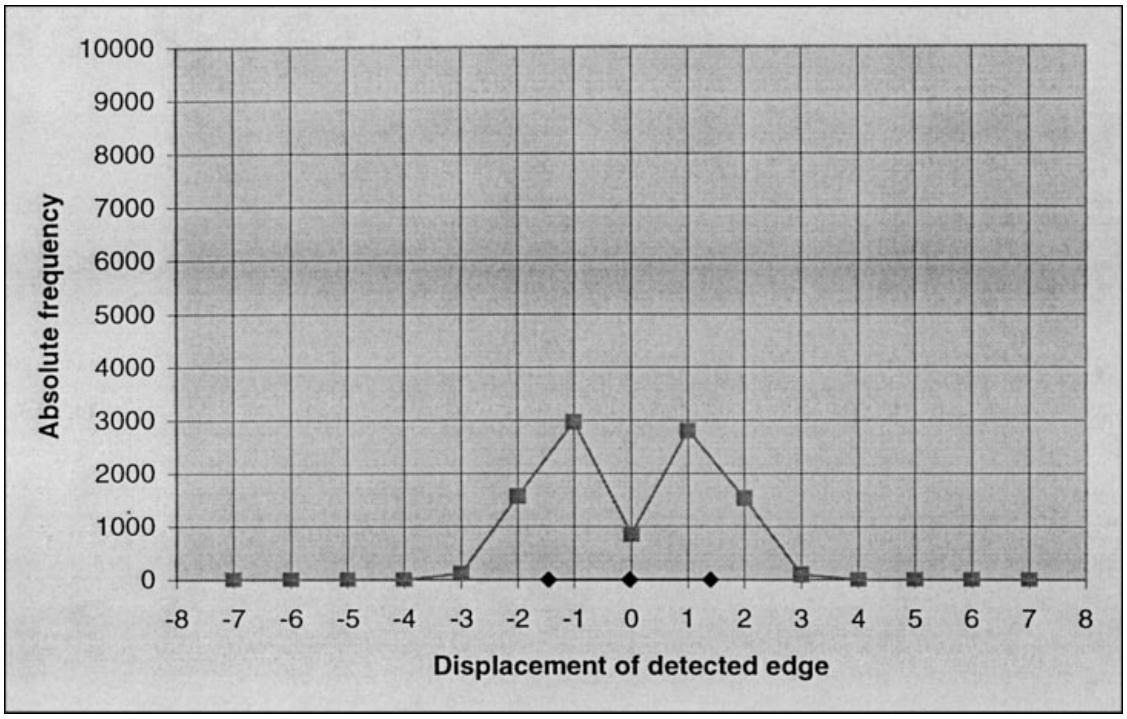

Figure 8. Results of Experiment C.

estimating the probability of a successful divergence maximum in situations in which a maximum should normally not occur in the correct position—that is, in two regions with the same histogram. Two circumstances arise: 1) the texels have the same histogram, without any other condition (fluctuations by residues); 2) the texels are identical except for a permutation in rows and/or columns (change of phase). The formulation of the problem is the following: after fixing the window and texel size, and the histogram common to both, we calculated the probability of a correct detection.

After performing several preliminary proofs with different histograms, we formulated the following conjecture: for texels larger than 2×2 with the same histogram, and using a semi-window size equal to the texel, no case of a single maximum is detected in the divergence, nor a case of a double maximum separated by two pixels. This deduction has been confirmed with texel sizes of 4×4, 5×5 and 8×8. In the case of the size 2×2, double maximums did appear at a distance of two pixels. Therefore, to achieve valid samples, the tests must be conducted with window sizes coprime with the size of the texel.

Fluctuations by residuals. In this section, we analyse the probability that the divergence is correct when the texels of the two regions have random pixel configurations but an equal histogram. If the size of the semi-window coincides with that of the texels, then the divergence will certainly have a value of zero when the window is centred. However, this is not true in the case of windows greater than the texel and of coprime size with it (having no common prime divisors), since the statistical-uniformity hypothesis is not fulfilled. To examine the possible cases, we conducted the Monte Carlo experiments D, E and F. The numeric data are given in Table 2, and the graphic results in Figs. 9–11. Below are commentaries on the layouts and on the results of each experiment.

Experiment D. The graphic results of the detection are reflected in Fig. 9. In most cases, the edge is detected at a distance of one pixel from the correct

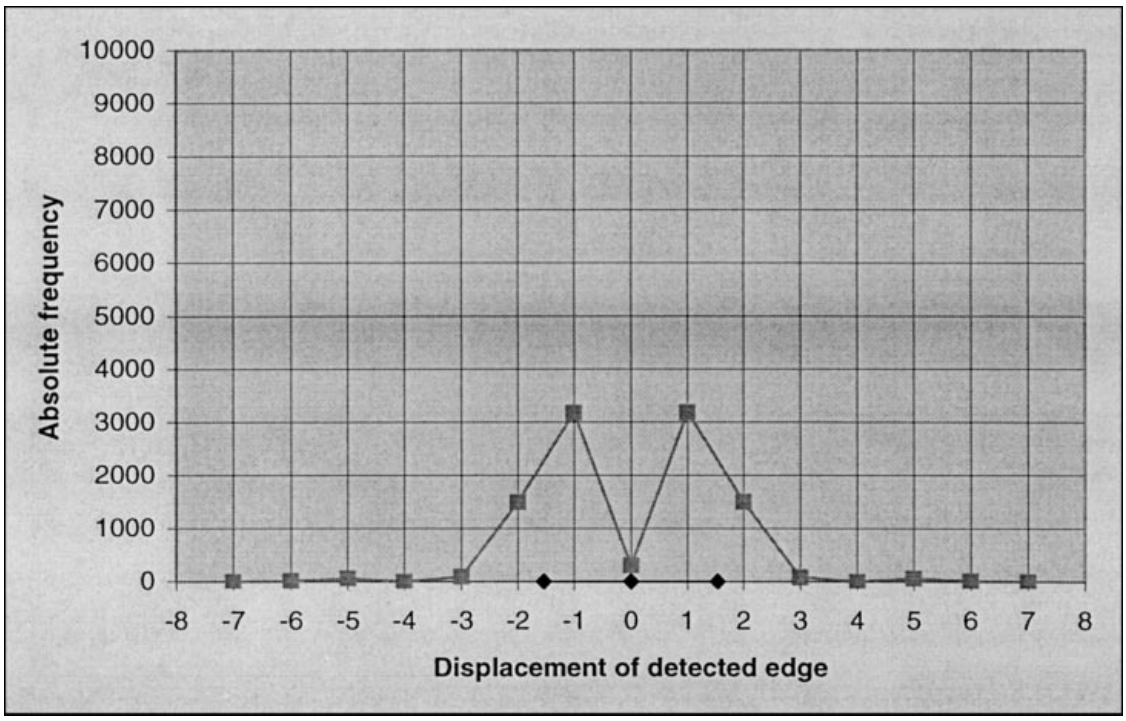

Figure 9. Results of Experiment D.

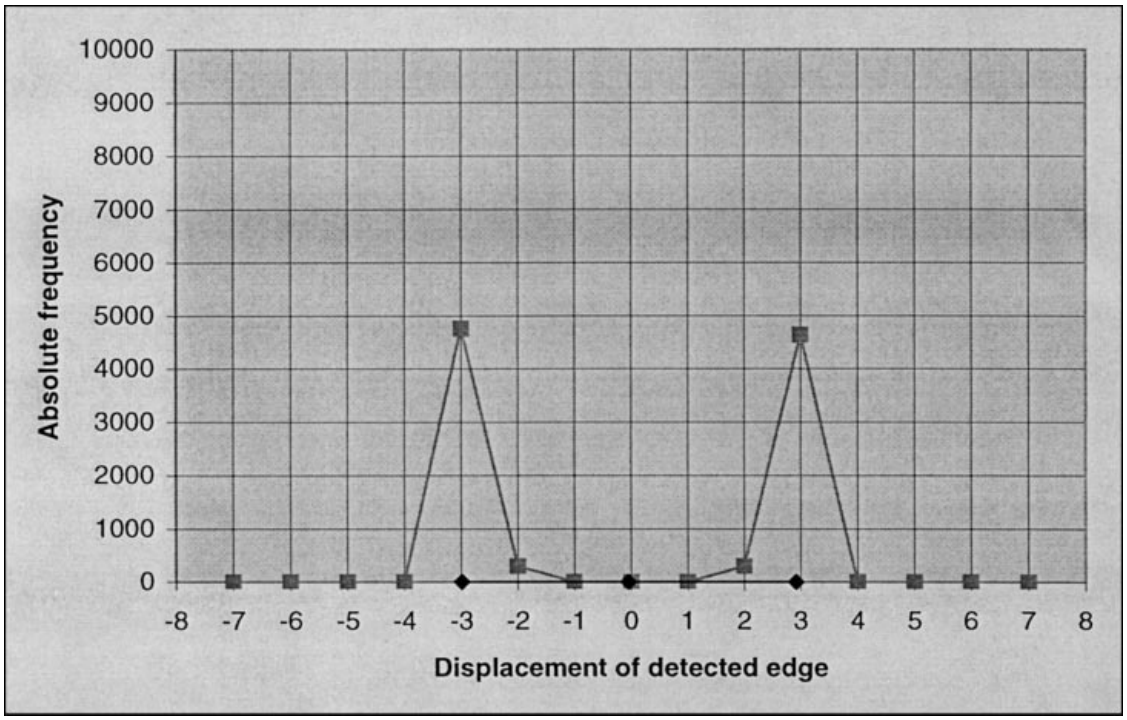

Figure 10. Results of Experiment E.

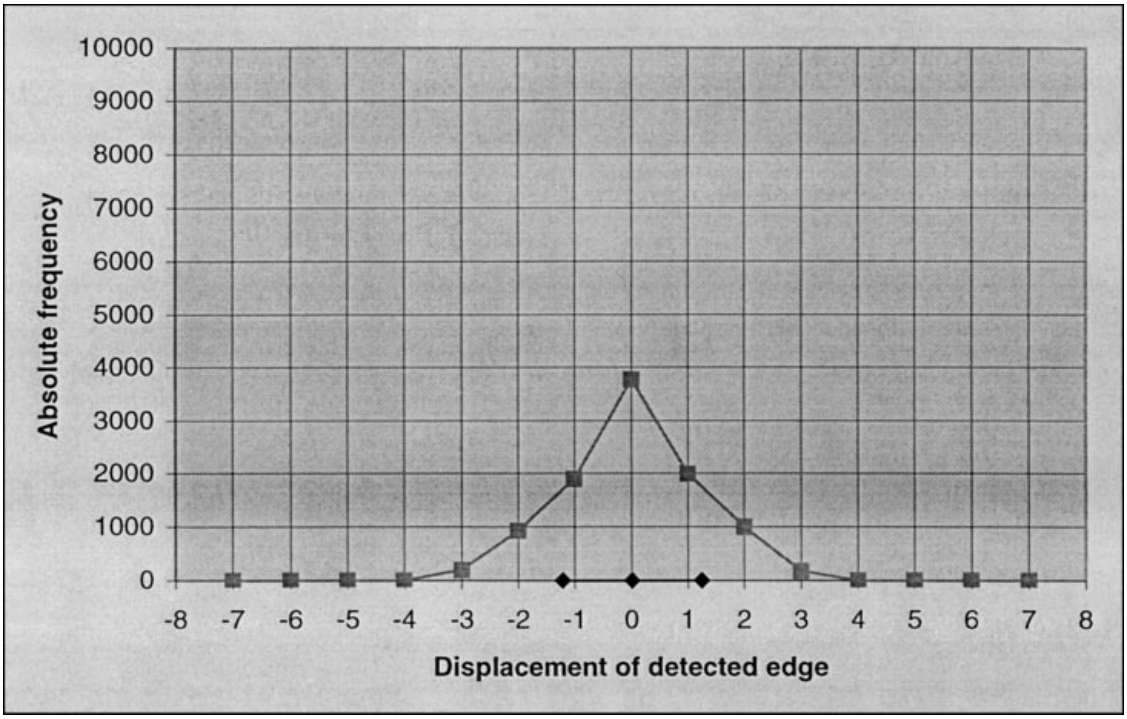

Figure 11. Results of Experiment F.

Table 2. Results of Monte Carlo Experiments D, E and F.

	Exp. D	Exp. E	Exp. F
Parameters			
N. valid samples	10,000	10,000	10,000
Texel size	4×4	4×4	4×4
Semi-window size	7×7	5×5	5×5
Texel histograms	{2, 1, 1, 2, 3, 2, 3, 2}	{2, 1, 1, 2, 3, 2, 3, 2}	{2, 1, 1, 2, 3, 2, 3, 2}
Discards	23,229	72,894	13,561
Results (detection error)			
Ends	$-7, +7$	$-3, +3$	$-3, +5$
Mean error	$+0.0035$	-0.039	$+0.0093$
Standard deviation	1.5549	2.95	1.2281

position. The fact that sometimes the edge is detected in the correct place is due to the size disparity of texels and semi-windows, or, more precisely, to the absence of common factors between the two. In fact, this numeric relationship implies the non-fulfilment of the hypothesis of statistical uniformity, and therefore the histograms observed in the two semi-windows differ somewhat from each other in the centred window position, allowing the appearance of divergence maximums.

Changes of phase. In this section, we study changes of phase, which comprise a particular case of texels with identical histograms. In fact, now the texel from one region is identical to that of the other, except in a cyclic permutation of the columns (or of the rows) of the texel.

Experiment E. Here, we show the results from the experiment on the change of phase in a direction perpendicular to the edge. The graphic results are presented in Fig. 10, in which we see that in most cases the border is detected at a distance of 3 pixels with respect to the correct position.

Experiment F. With this experiment, we analyse the results of the change of phase parallel to the edge. We used the same parameters as in the previous experiment, varying only the direction of the change of phase. The graphic results are shown in Fig. 11. In this experiment, the edge is correctly detected in most cases, as opposed to the case of the change perpendicular to the edge.

5. Some Examples

Though not the primary aim of this work, we offer some examples of application of the method analysed. We refer the reader to other publications by the same authors [9, 10].

5.1. Synthetic Noisy Images

In this section, we show the result of using Jensen-Shannon divergence to segment synthetic images contaminated by various types of noise. In Figs. 12 and 13, the two series contain synthetic images, the first being composed of bands of a uniform grey level, and the second of perfectly regular textures of texel size 4×4. All the detections were made with two 15×15 semi-windows. After this, a simple detection algorithm of local maximums was used to create the binary image. More precisely, this algorithm centres on each pixel a set of 1D-windows with the same length and different direction (typically four) and then marks the central pixel as belonging to the edge if its associated divergence is maximum at least in one of the windows. Length of the windows is a parameter, and it is selected to be a little greater than the dimensions of the windows used to calculate divergences.

5.2. Natural Scenes

This experiment, shown in Fig. 11, was performed using two 3×3 semi-windows, and applying the same detection algorithm of local maximums as in the previous

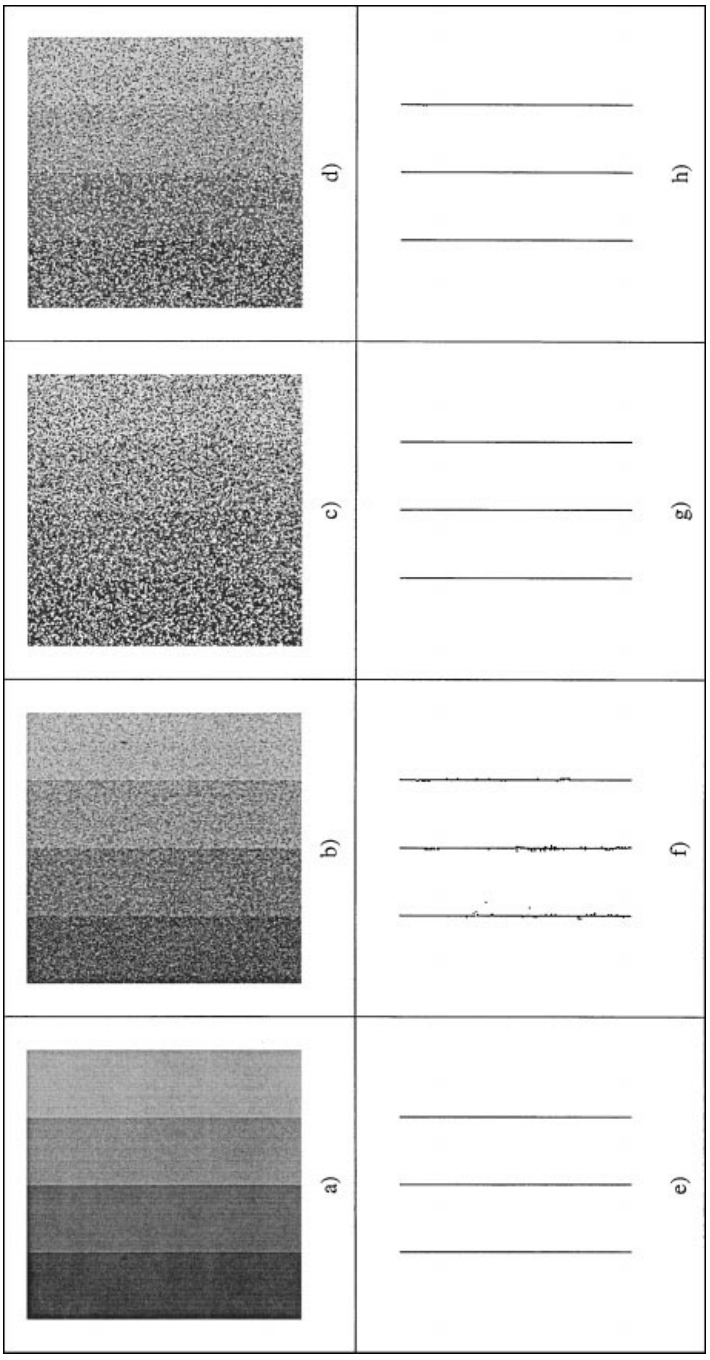

Figure 12. a) Original image of textured bands; b) Original image with Gaussian noise of $\mu = 0$ and $\sigma = 25$. c) Original image with 50% salt and pepper noise. d) Original image with 50% uniform noise. e), f), g) and h) Detected edges corresponding to the images a) to d).

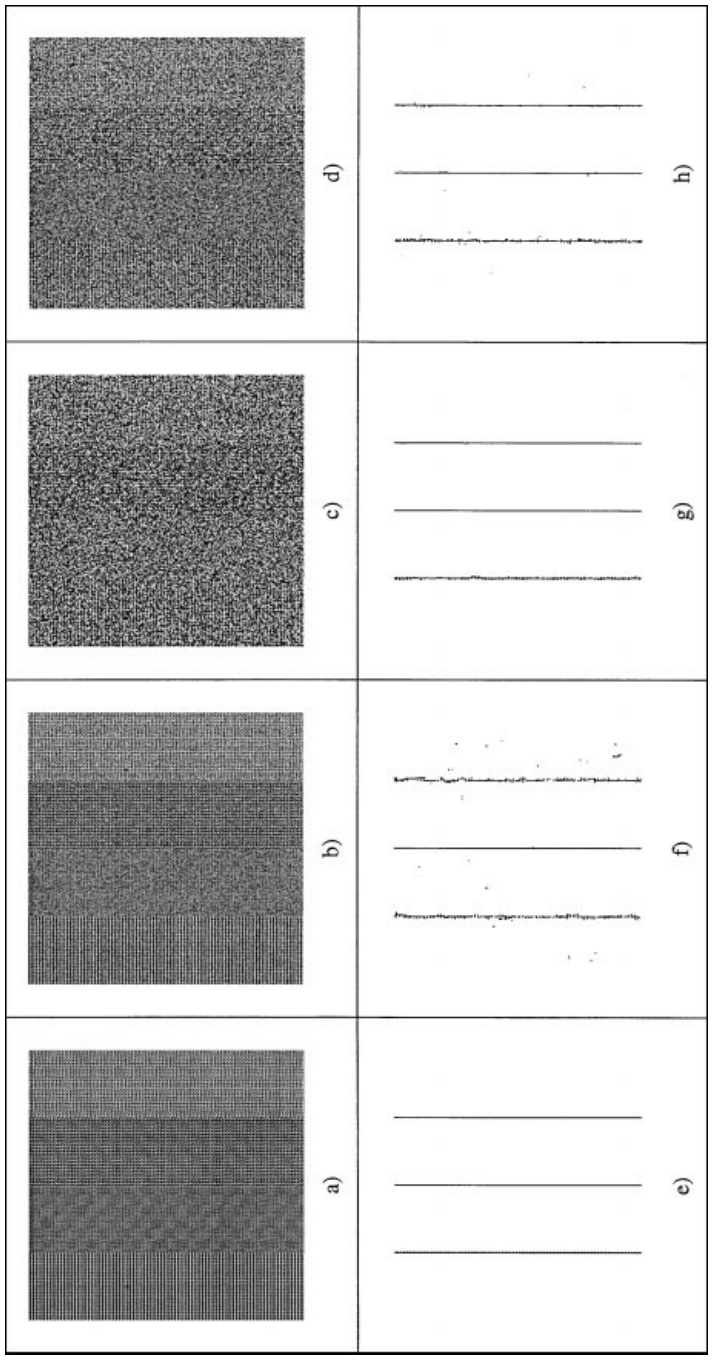

Figure 13. a) Original image of textured bands; b) Original image with Gaussian noise of $\mu = 0$ and $\sigma = 25$. c) Original image with 50% salt and pepper noise. d) Original image with 50% uniform noise. e), f), g) and h) Detected edges corresponding to the images a) to d).

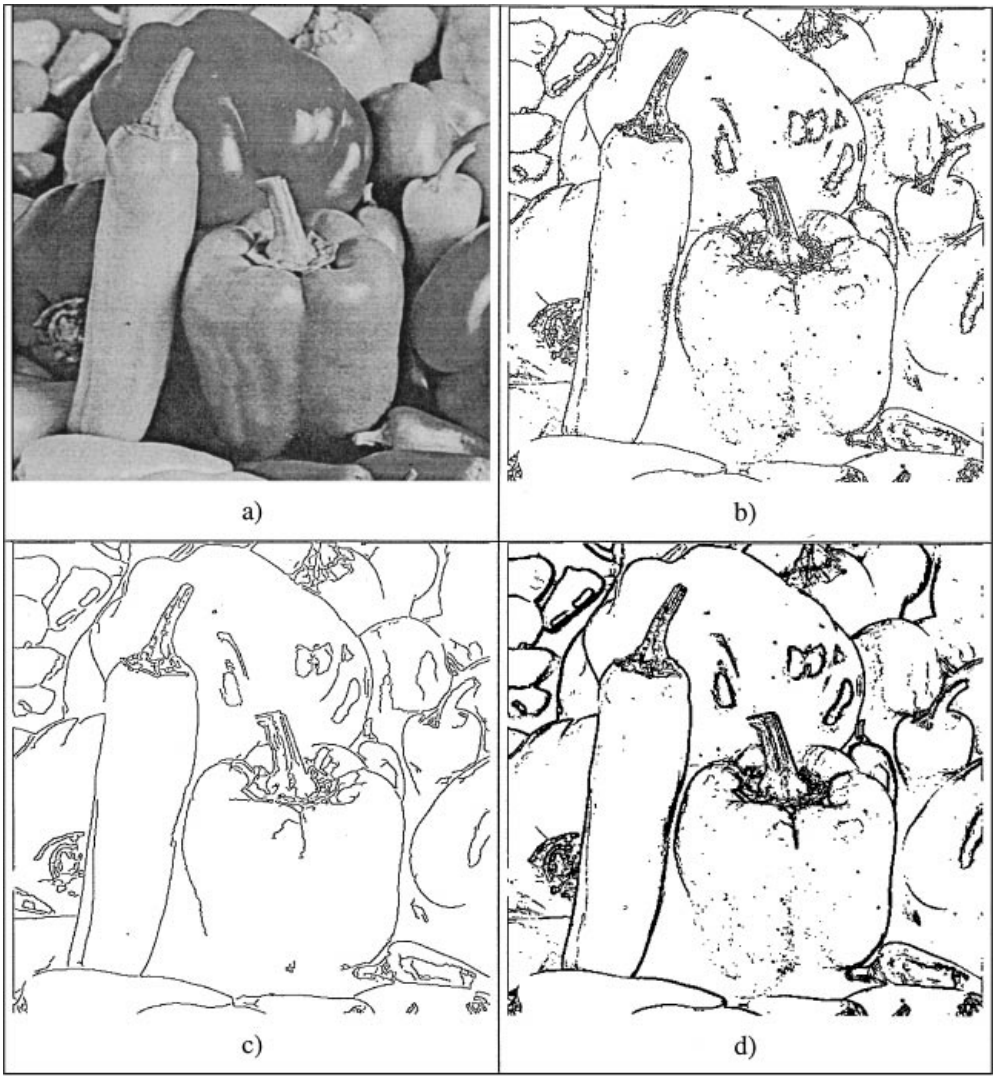

Figure 14. a) Image of a natural scene. b) Detected edges of the original image. c) Output of the Canny filter. d) Output of the Sobel filter.

experiments. Results from other well-known edge detectors have been included to allow comparison.

5.3. Satellite Images

This experiment, shown in Fig. 15, was performed using two 5×5 semi-windows, and applying the same detection algorithm of local maximums as in the previous experiments.

6. Conclusions

Here, we subject our edge-detection method to a theoretic analysis based on Jensen-Shannon divergence between grey histograms. In particular, we study the variation of the divergence with the movement of the window in the border zone between two regions. The principal results demonstrate that: 1) in general, the maximum divergence occurs for the position at which the window is centred on the edge, and the deviation (error in the position at the edge) is small and improbable; 2) the distribution of this error depends on the characteristics of the regions and of the operational parameters, appearing for the cases of most interest to the potential user; 3) the divergence maximum is a sharp peak and therefore easily detected by the function $\mathrm{JS}(x)$, even for small differences of composition between the two regions.

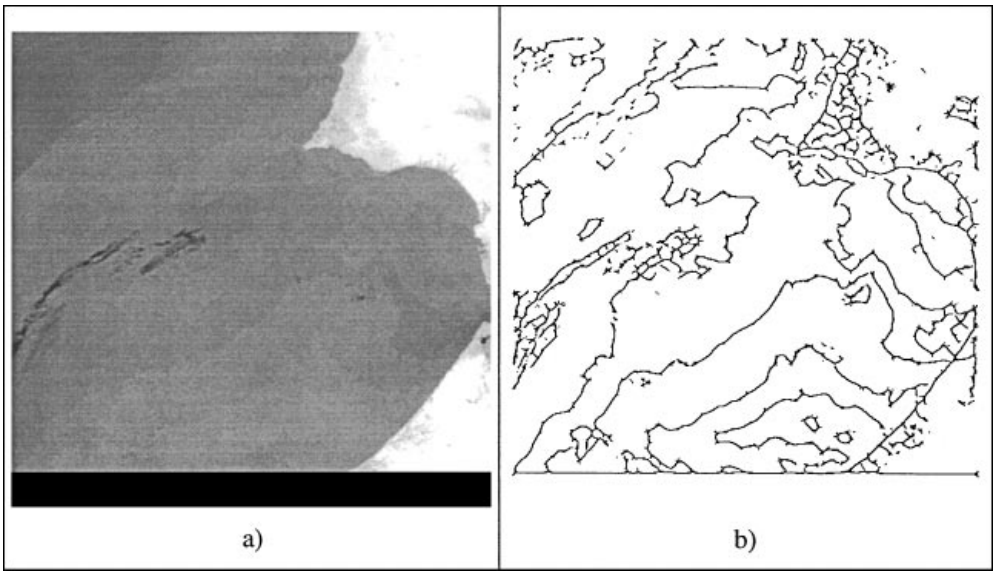

Figure 15. a) Tele-detection image obtained by the sensor AVHRR of the satellite NOAA. b) Detected edges of the original image.

Acknowledgments

This work was supported in part by grants TIC94-535 and MAR97-0464 of the Spanish government.

Notes

1. Term commonly used to designate the repeating pattern that composes a textured region.
2. Here we continue to assume that the semi-window has the same shape and size as the texel.

References

1. K.R. Castleman, *Digital Image Processing*, Prentice-Hall, 1996.
2. A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, 1986.
3. W.K. Pratt, *Digital Image Processing*, John Wiley & Sons, 1991.
4. J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. PAMI-9, No. 6, pp. 679–698, 1986.
5. D.J. Park, K.M. Nam, and R.H. Park, "Edge detection in noisy images based on the co-occurrence matrix," *Pattern Recognition*, Vol. 27, No. 6, pp. 765–774, 1994.
6. V. Barranco López, P. Luque Escamilla, J. Martínez Aroza, and R. Román Roldán, "Entropic texture-edge detection for image segmentation," *Electronic Letters*, Vol. 31, No. 11, pp. 867–869, 1995.
7. J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. On Information Theory*, Vol. 37, No. 1, pp. 145–150, 1991.
8. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.
9. V. Barranco López, P. Luque Escamilla, J. Martínez Aroza, and R. Román Roldán, "Texture segmentation based on information-theoretic edge detection methods," in *Proc. VI Symposium Nacional de Reconocimiento de Formas y Análisis de Imágenes*. ISBN: 84-605-2447-7. Córdoba, 1995.
10. Ch. Atae-Allah, J.F. Gómez Lopera, J. Martínez Aroza, and R. Román Roldán, "Entropic segmentation of oceanic images," in *Proc. Colloque International sur le Traitement d'Images et les Systèmes de Vision Artificielle* TISVA'98. Oujda, 1998. (Prix de la meilleure communication orale)
11. A. Ben Hamza, *Contribución al tratamiento de imágenes: filtros selectivos tipo mediana y propiedades de la divergencia de Jensen-Shannon*, Ph.D. Thesis, University of Granada, 1997.

Juan Francisco Gómez-Lopera was born in Granada, Spain, on June 3 1968. He received his Ph.D. degree in Physics from the University of Granada in 1995. From 1992 to 1996 he was Assistant Professor at the University of Almería, Spain and currently he is Assistant Professor at the University of Granada. His current interests involve Image Segmentation and its applications to different fields.

José Martínez-Aroza received his B.Sc. degree in Mathematics in 1979 from the University of Granada, Spain. From 1979 to 1987 he was a professor at the University of Almería, Spain. He received the Ph.D. degree in Applied Mathematics in 1990 from the University of Granada, Spain. He is associate professor at the Department of Applied Mathematics, University of Granada, Spain. His research interests lie on the field of digital image processing, more specifically on noise filtering and segmentation of homogeneous as well as textured regions.

Aureliano M. Robles-Pérez was born in Granada, Spain, in 1965. He received his B.Sc. degree in Mathematics in 1988 from the University of Granada. He is assistant professor at the Department of Applied Mathematics, University of Granada. His research interests are on mathematical aspects of image processing.

R. Román-Roldán is Professor of Applied Physics at the University of Granada (Spain). He is currently researching in two interrelated subjects, both based on the use of an information-theoretic measure, the Jensen-Shannon divergence: Entropic edge detection in digital images, and Complexity analysis in DNA sequences through entropic segmentation.