Edge Detection by Helmholtz Principle

AGNÈS DESOLNEUX, LIONEL MOISAN* AND JEAN-MICHEL MOREL
CMLA, ENS Cachan, 61 av. du président Wilson, 94235 Cachan cedex, France
desolneu@cmla.ens-cachan.fr
moisan@cmla.ens-cachan.fr
morel@cmla.ens-cachan.fr

Abstract. We apply to edge detection a recently introduced method for computing geometric structures in a digital image, without any *a priori* information. According to a basic principle of perception due to Helmholtz, an observed geometric structure is perceptually "meaningful" if its number of occurences would be very small in a random situation: in this context, geometric structures are characterized as large deviations from randomness. This leads us to define and compute edges and boundaries (closed edges) in an image by a parameter-free method. Maximal detectable boundaries and edges are defined, computed, and the results compared with the ones obtained by classical algorithms.

Keywords: image analysis, perception, Helmholtz principle, edge detection, large deviations

1. Introduction

In statistical methods for image analysis, one of the main problems is the choice of an adequate prior. For example, in the Bayesian model [9], given an observation "obs", the aim is to find the original "model" by computing the Maximum A Posteriori (MAP) of

$$P[\text{model} \mid \text{obs}] = \frac{P[\text{obs} \mid \text{model}] \times P[\text{model}]}{P[\text{obs}]}.$$

The term $P[\text{obs} \mid \text{model}]$ represents the degradation (superimposition of a gaussian noise for example) and the term $P[\text{model}]$ is called the prior. This prior plays the same role as the regularity term in the variational framework. This prior has to be fixed and it is generally difficult to find a good prior for a given class of images. It is also probably impossible to give an all-purpose prior!

In [6 and 7], we have outlined a different statistical approach, based on phenomenological observations coming from Gestalt theory [21, 27, 29]. According to a perception principle which seems to go back

to Helmholtz, every large deviation from a "uniform noise" image should be perceptible, provided this large deviation corresponds to an *a priori* fixed list of geometric structures (lines, curves, closed curves, convex sets, spots, local groups, ...). Thus, there still is an *a priori* geometric model, but, instead of being quantitative, this model is merely qualitative. Let us illustrate how this should work for "grouping" black dots in a white sheet. Assume we have a white image with black dots spread out. If some of them form a cluster, say, in the center of the image, then, in order to decide whether this cluster indeed is a group of points, we compute the expectation of this grouping event happening by chance if the dots were uniformly disributed in the image. If this expectation happens to be very low, we decide that the group in the center is meaningful. Thus, instead of looking for objects as close as possible to a given prior model, we consider a "wrong" and naive model, actually a random uniform distribution, and then define the "objects" as large deviations from this generic model. One can find in [13] a very close formulation of computer vision problems.

We may call this method Minimal A Posteriori Expectation, where the prior for the image is a uniform

---

random noise model. Indeed, the groups (geometric structures, gestalts[1]) are defined as the best *counterexamples*, i.e. the least expected. Those counterexamples to the uniform noise assumption are taken in a restricted geometric class. Notice that not all such counterexamples are valid: the Gestalt theory fixes a list of perceptually relevant geometric structures which are supposedly looked for in the perception process. The computation of their expectation in the uniform noise model validates their detection: the least expected in the uniform noise model, the more perceptually meaningful they will be.

This uniform noise prior is generally easy to define. Consider for example the case of orientations: since we do not have any reason to favour some directions, the prior on the circle S^1 will be the uniform distribution. We applied this method in a previous paper dedicated to the detection of meaningful alignments [6]. In [7] we have generalized the same method to the definition of what we called "maximal meaningful modes" of a histogram. This definition is crucial in the detection of many geometric structures or gestalts, like groups of parallel lines, groups of segments with similar lengths, etc.

It is clear that the above outlined Minimum A Posteriori method will prove its relevance in Computer Vision only if it can be applied to each and all of the gestalt qualities proposed by phenomenology. Actually, we think the method might conversely contribute to a more formal and general mathematical definition of geometric structures than just the ones coming from the usual plane geometry. Now, for the time being, we wish to validate the approach by matching the results with all of the classicaly computed structures in image analysis. In this paper, we shall address the comparison of edge and boundary detectors obtained by the Minimum A Posteriori method with the ones obtained by state of the art segmentation methods.

A main claim in favour of the Minimum A Posteriori is its reduction to a single parameter, the meaningfulness of a geometric event depending only on the difference between the logarithm of the false alarm rate and the logarithm of the image size! We just have to fix this false alarm rate and the dependance of the outcome is anyway a log-dependence on this rate, so that the results are very insensitive to a change. Our study of edge detection will confirm this result, with slightly different formulas though.

In addition, and although the list of geometric structures looked for is wide (probably more than ten in Gestalt theory), the theoretical construction will make sense if they are all deduced by straightforward adaptations of the same methodology to the different geometric structures. Each case of geometric structure deserves, however, a particular study, in as much as we have to fix in each case the "uniform noise" model against which we detect the geometric structure. We do not claim either that what we do is 100% new: many statistical studies on images propose a "background" model against which a detection is tested; in many cases, the background model is a merely uniform noise, as the one we use here. Optimal thresholds have been widely addressed for detection or image thresholding [1, 10, 19, 22] Also, many applied image analysis and engineering methods, in view of some detection, address the computation of a "false alarm rate". Our "meaningfulness" is nothing but such a false alarm rate, but applied to very general geometric objects instead of particular looked for shapes and events.

As was pointed out to us by David Mumford, our method is also related to the statistical hypothesis testing, where the asked question is: does the observation follow the prior law given by Helmoltz principle? The gestalts will be the "best proofs" (in terms of the *a priori* fixed geometric structures) that the answer to this question is no.

Let us summarize: not all geometric structures are perceptually relevant; a small list of the relevant ones is given in Gestalt theory; we can "detect" them one by one by the above explained Helmholtz principle as large deviations from randomness. Now, the outcome is not a global interpretation of the image, but rather, for each gestalt quality (alignment, parallelism, edges), a list of the maximal detectable events. The maximality is necessary, as shows the following example, which can be adapted to each other gestalt: assume we have detected a dense cluster of black dots; this means that the expectation of such a big group is very small for a random uniform distribution of dots. Now, very likely, many subgroups of the detected dots and also many larger groups will have a small expectation too. So we can add spurious elements to the group and still have a detectable group. Thus, maximality is very relevant in order to obtain the best detectable group. We say that a group or gestalt is "maximal detectable" if any subgroup and any group containing it are less detectable, that is, have a smaller expectation.

We shall address here one of the *serpents de mers* of Computer Vision, namely "edge" and boundary

"detection" [8, 9, 12, 17, 20, 25, 26, 30, 33]. We define an "edge" as a level line along which the contrast of the image is strong. We call "boundary" a closed edge. We shall in the following give a definition of meaningfulness and of optimality for both objects. Then, we shall show experiments and discuss them. A comparison with the classical Mumford-Shah segmentation method will be made and also with the Canny-Deriche edge detector. We shall give a (very simple in that case) proof of the existence of maximal detectable gestalt, applied to the edges. What we do on the edges won't be a totally straightforward extension of the method we developed for alignments in [6]. Indeed, we cannot do for edge or boundary strength as for orientation, i.e. we cannot assume that the modulus of the gradient of an image is uniformly distributed.

2.   Contrasted Boundaries

We call "contrasted boundary" any closed curve, long enough, with strong enough contrast and which fits well to the geometry of the image, namely, orthogonal to the gradient of the image at each one of its points. We will first define ε-meaningful contrasted boundaries, and then maximal meaningful contrasted boundaries. Notice that this definition depends upon two parameters (long enough, contrasted enough) which will be usually fixed by thresholds in a computer vision algorithm, unless we have something better to say. In addition, most boundary detection will, like the snake method [12], introduce regularity parameters for the searched for boundary [16]. If we remove the condition "long enough", we can have boundaries everywhere, as is patent in the classical Canny filter [2].

The considered geometric event will be: a strong contrast along a level line of an image. Level lines are curves directly provided by the image itself. They are a fast and obvious way to define global, contrast insensitive candidates to "edges" [3]. Actually, it is well acknowledged that edges, whatever their definition might be, are as orthogonal as possible to the gradient [2, 4, 8, 14, 20]. As a consequence, we can claim that level lines are the adequate candidates for following up local edges. The converse statement is false: not all level lines are "edges". The claim that image boundaries (i.e. closed edges) in the senses proposed in the literature [18, 23] also are level lines is *a priori* wrong. How wrong it is will come out from the experiments, where we compare an edge detector with a boundary detec-

tor. Surprisingly enough, we will see that they can give comparable results.

We now proceed to define precisely the geometric event: "at each point of a length l (counted in independent points) part of a level line, the contrast is larger than μ". Then, we compute the expectation of the number of occurrences of such an event (i.e. the number of false alarms). This will define the thresholds: minimal length of the level line, and also minimal contrast in order to be meaningful. We will give some examples of typical numerical values for these thresholds in digital images. Then, as we mentioned has been done for other gestalts like alignments and histograms, we will define here a notion of maximality, and derive some properties.

2.1.   Definitions

Let u be a discrete image, of size $N \times N$. We consider the level lines at quantized levels $\lambda_1, \ldots, \lambda_k$. The quantization step q is chosen in such a way that level lines make a dense covering of the image: if e.g. this quantization step q is 1 and the natural image ranges 0 to 256, we get such a dense covering of the image. A level line can be computed as a Jordan curve contained in the boundary of a level set with level λ,

$$\chi_\lambda = \{x/u(x) \leq \lambda\} \quad \text{and} \quad \chi^\lambda = \{x/u(x) \geq \lambda\}.$$

Notice that along a level line, the gradient of the image must be everywhere above zero. Otherwise the level line contains a critical point of the image and is highly dependent upon the image interpolation method. Thus, we consider in the following only level lines along which the gradient is not zero. The interpolation considered in all experiments below is the order zero interpolation (the image is considered constant on each pixel and the level lines go between the pixels).

Let L be a level line of the image u. We denote by l its length counted in independent points. In the following, we will consider that points at a geodesic distance (along the curve) larger than 2 are independent (i.e. the contrast at these points are independent random variables). Let $x_1, x_2, \ldots, x_1$ denote the l considered points of L. For a point $x \in L$, we will denote by $c(x)$ the contrast at x. It is defined by

$$c(x) = |\nabla u|(x), \tag{1}$$

where ∇u is computed by a standard finite difference on a 2×2 neighborhood [7]. For $\mu \in \mathbb{R}_+^*$, we consider

the event: for all $1 \leq i \leq l$, $c(x_i) \geq \mu$, i.e. each point of L has a contrast larger than μ. From now on, all computations are performed in the Helmholtz framework explained in the introduction: we make all computations as though the contrast observations at x_i were mutually independent. Since the l points are independent, the probability of this event is

$$P[c(x_1) \geq \mu] \times P[c(x_2) \geq \mu]$$
$$\times \cdots \times P[c(x_l) \geq \mu] = H(\mu)^l, \quad (2)$$

where $H(\mu)$ is the probability for a point on any level line to have a contrast larger than μ. An important question here is the choice of $H(\mu)$. Shall we consider that $H(\mu)$ is given by an *a priori* probability distribution, or is it given by the image itself (i.e. by the histogram of gradient norm in the image)? In the case of alignments, we took by Helmholtz principle the orientation at each point of the image to be a random, uniformly distributed variable on $[0, 2\pi]$. Here, in the case of contrast, it does not seem sound at all to consider that the contrast is uniformly distributed. In fact, when we observe the histogram of the gradient norm of a natural image (see Fig. 1), we notice that most of the points have a "small" contrast (between 0 and 3), and that only a few points are highly contrasted. This is explained by the fact that a natural image contains many flat regions (the so called "blue sky effect", [11, 32]. In the following, we will consider that $H(\mu)$ is given by the image itself, which means that

$$H(\mu) = \frac{1}{M} \#\{x/|\nabla u|(x) \geq \mu\}. \quad (3)$$

where M is the number of pixels of the image where $\nabla u \neq 0$. In order to define a meaningful event, we have to compute the expectation of the number of occurrences of this event in the observed image. Thus, we first define the number of false alarms.

Definition 1 (*Number of false alarms*).   Let L be a level line with length l, counted in independent points. Let μ be the minimal contrast of the points $x_1, \ldots, x_1$ of L. The number of false alarms of this event is defined by

$$NF(L) = N_{ll} \times [H(\mu)]^l, \quad (4)$$

where N_{ll} is the number of level lines in the image.

Notice that the number N_{ll} of level lines is provided by the image itself. We now define ε-meaningful level

lines. The definition is analogous to the definition of ε-meaningful modes of a histogram or to the definition of alignments: the number of false alarms of the event is less than ε.

Definition 2 (*ε-meaningful boundary*).   A level line L with length l and minimal contrast μ is an ε-meaningful boundary if

$$NF(L) = N_{ll} \times [H(\mu)]^l \leq \varepsilon. \quad (5)$$

The above definition involves two variables: the length l of the level line, and its minimal contrast μ. The number of false alarms of an event measures the "meaningfulness" of this event: the smaller it is, the more meaningful the event is.

Let us now proceed to define "edges". We denote by N_{llp} the number of pieces of level lines in the image.

Definition 3 (*ε-meaningful edge*).   A piece of level line E with length l and minimal contrast μ is an ε-meaningful edge if

$$NF(E) = N_{llp} \times [H(\mu)]^l \leq \varepsilon. \quad (6)$$

Here is how N_{llp} is computed: we first compute all level lines at uniformly quantized levels (grey level quantization step is 1 and generally ranges from 1 to 255. For each level line, L_i with length l_i, we compute its number of pieces, sampled at pixel rate, the length unit being pixel side. We then have

$$N_{llp} = \sum_i \frac{l_i(l_i-1)}{2}.$$

This fixes the used number of samples. This number of samples will be fair for a 1-pixel accurate edge detector. Clearly, we do *detection* and not optimization of the detected edge: in fact, according to Shannon conditions, edges have a between two or three pixels width. Thus, the question of finding the "best" edge representative among the found ones is not addressed here, but has been widely addressed in the literature [2, 4].

2.2.   Thresholds

In the following we will denote by F the function defined by

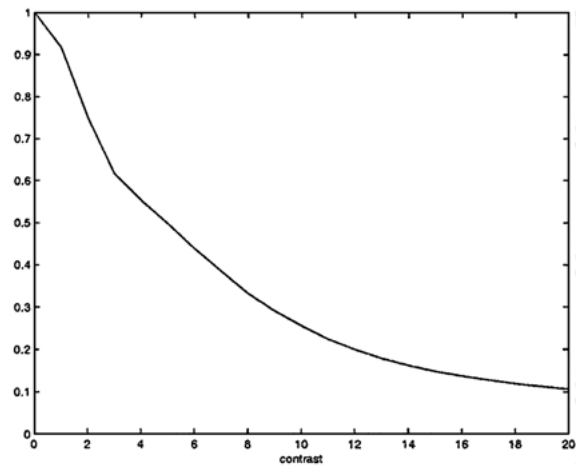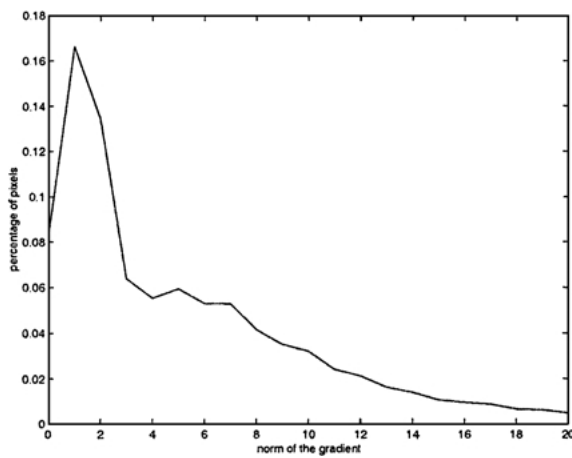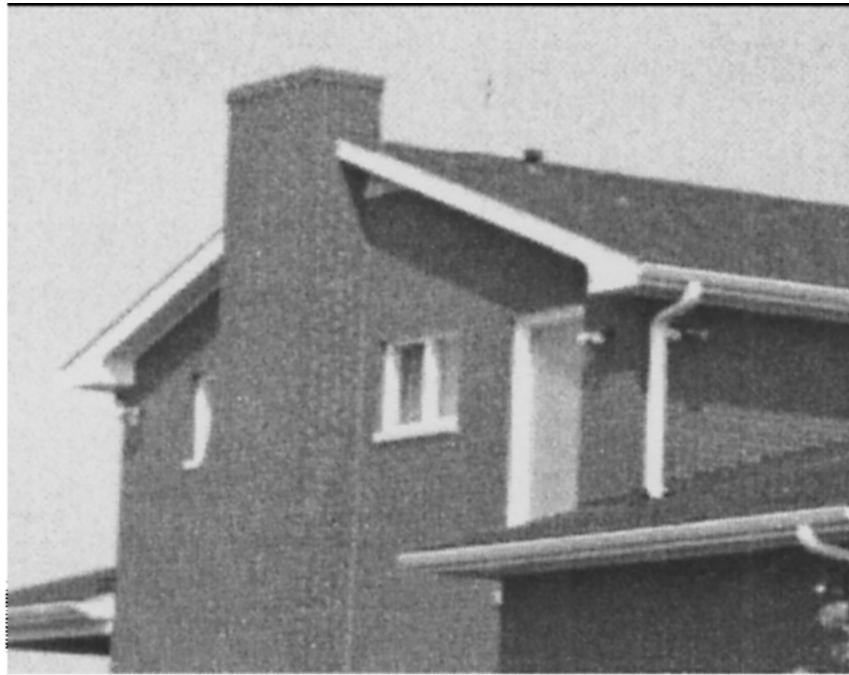$$F(\mu, l) = N_{ll} \times [H(\mu)]^l. \quad (7)$$

Figure 1.   From left to right: 1. original image; 2. histogram of the norm of the gradient; 3. its repartition function ($\mu \mapsto P[|\nabla u| \geq \mu]$).

Thus, the number of false alarms of a level line of length l and minimal contrast μ is simply $F(\mu, l)$.

Since the function $\mu \mapsto H(\mu) = P[c(x) \geq \mu]$ is decreasing, and since for all μ, we have $H(\mu) \leq 1$, we obtain the following elementary properties:

– We fix μ and $l \leq l'$, then

$$F(\mu, l) \geq F(\mu, l'),$$

which shows that if two level lines have the same minimal contrast, the more meaningful one is the longer one.

– We fix l and $\mu \leq \mu'$, then

$$F(\mu, l) \geq F(\mu', l),$$

which shows that if two level lines have the same length, the more meaningful one is the one with higher contrast.

When the contrast μ is fixed, the minimal length $l_{\min}(\mu)$ of an ε-meaningful level line with minimal contrast μ is

$$l_{\min}(\mu) = \frac{\log \varepsilon - \log N_{ll}}{\log H(\mu)}. \tag{8}$$

Conversely, if we fix the length l, the minimal contrast $\mu_{\min}(l)$ needed to become ε-meaningful is such that

$$\mu_{\min}(l) = H^{-1}([\varepsilon/N_{ll}]^{1/l}). \tag{9}$$

2.3.   Maximality

In this subsection, we address two kinds of maximality for the edges and the boundaries. Let us start with boundaries. A natural relation between closed level lines is given by their inclusion [15]. If C and C' are two different closed level lines, then C and C' cannot intersect. Let D and D' denote the bounded domains surrounded by C and C'. Either $D \cap D' = \phi$ or ($D \subset D'$ or $D' \subset D$). We can consider, as proposed by Monasse, the inclusion tree of all level lines. From now on, we work on the subtree of the detected level curves, that is, the ones for which $F(\mu, l) \leq \varepsilon$ where ε is our *a priori* fixed expectation of false alarms. (In practice, we take $\varepsilon = 1$ in all experiments.) On this subtree, we can, following Monasse, define what we shall call a *maximal monotone level curve interval*, that is, a sequence of level curves C_i, $i \in [1, k]$ such that:

– for $i \geq 2$, C_i is the unique son of C_{i-1},
– the interval is maximal (not contained in a longer one)
– the grey levels of the detected curves of the interval are either decreasing from 1 to k, or increasing from 1 to k.

We can see many such maximal monotone intervals of detected curves in the experiments: they roughly correspond to "fat" edges, made of several well contrasted level lines. The edge detection ideology tends to define an edge by a single curve. This is easily made by selecting the best contrasted edges along a series of parallel ones.

Definition 4.   We associate with each maximal monotone interval its optimal level curves, that is, the ones for which the false alarms number $F(\mu, l)$ is minimal along the interval. We call "optimal boundary map" of an image the set of all optimal level curves.

This optimal boundary map will be compared in the experiments with classical edge detectors or segmentation algorithms.

We now address the problem of finding optimal edges among the detected ones. We won't be able to proceed as for the boundaries. Although the pieces of level lines inherit the same inclusion structure as the level lines, we cannot compare two of them belonging to different level curves for detectability, since they can have different positions and lengths. We can instead compare two edges belonging to the same level curve. Our main aim is to define on each curve a set of disjoint maximally detectable edges. In the following, we denote by $NF(E) = F(\mu, l)$ the false alarm number of a given edge E with minimal gradient norm μ and length l.

Definition 5.   We call maximal meaningful edge any edge E such that for any other edge E' on the same level curve such that $E \subset E'$ (resp. $E' \subset E$) we have $NF(E') > NF(E)$ (resp. $NF(E') \geq NF(E)$).

This definition follows [6, 7] where we apply it to the definition of maximal alignments and maximal modes in a histogram.

Proposition 1.   *Two maximal edges cannot meet.*

Proof: Let E and E' be two maximal distinct and non-disjoint meaningful edges in a given level curve and μ and μ' the respective minima of gradient of the image on E and E'. Assume e.g. that $\mu' \leq \mu$. Then $E \cup E'$ has the same minimum as E' but is longer. Thus, by the remark of the preceding subsection, we have $F(\mu', l + l') < F(\mu', l')$, which implies that $E \cup E'$ has a smaller number of false alarms than E'. Thus, E' is not maximal. As a consequence, two maximal edges cannot meet.   □

3.   Experiments

3.1.   INRIA Desk Image (Fig. 2)

In this experiment, we compare our method with two other methods: Mumford and Shah image segmentation and Canny-Deriche edge detector.

In the Mumford and Shah model [17], given an observed image u defined on the domain D, one looks for the piecewise approximation v of u that minimizes the
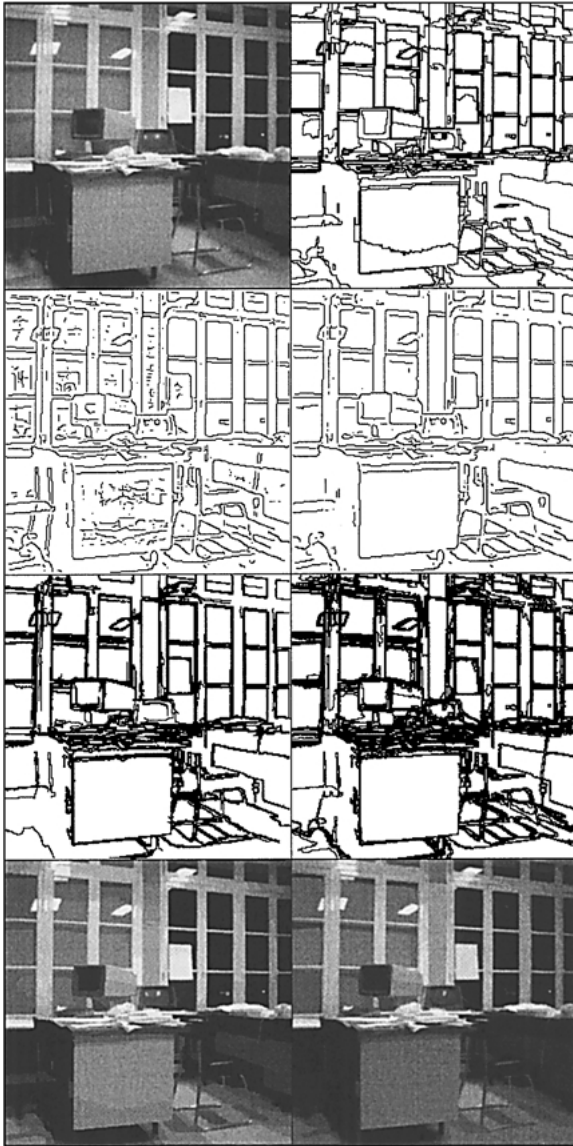
Figure 2. First row: left: original image; right: boundaries obtained with the Mumford-Shah model (1000 regions). Second row: edges obtained with Canny-Deriche edge detector, for two different threshold values (2 and 15). Third row: edges (left) and boundaries (right) obtained with our model ($\varepsilon = 1$). Fourth row: reconstruction with the Mumford-Shah model (left) and with our model (right). This last reconstruction is easily performed by the following algorithm: attribute to each pixel x the level of the smallest (for inclusion) meaningful level line surrounding x (see [15]).

functional

$$E(v) = \int_D |v - u|^2 + \lambda \text{length}(K(v)),$$

where length $(K(v))$ is the one-dimensional measure of the discontinuity set of v, and λ a parameter. Hence,

this energy is a balance between a fidelity term (the approximation error in L^2 norm) and a regularity term (the total length of the boundaries). The result v, called a segmentation of u, depends upon the parameter λ, that indicates how to weight both terms. As shown on Fig. 2, the Mumford-Shah model generally produces reasonable boundaries except in "flat" zones where spurious boundaries often appear (see the front side of the desk for example). This is easily explained: the *a priori* model is: the image is piecewise constant with boundaries as short as possible. Now, the image does not fit exactly the model: the desk in the image is smooth but not flat. The detected "wrong" boundary in the desk is necessary to divide the desk into flat regions. The same phenomenon occurs in the sky of the cheetah image (next experiment).

The Canny-Deriche filter [2, 5] is an optimization of Canny's well known edge detector, roughly consisting in the detection of maxima of the norm of the gradient in the direction of the gradient. Notice that, in contrast with the Mumford-Shah model and with our model, it does not produce a set of boundaries (ie one-dimensional structures) but a discrete set of points that still are to be connected. It depends on two parameters: the width of the impulse response, generally set to 1 pixel, and a threshold on the norm of the gradient that selects candidates for edge points. As we can see on Fig. 2, the result is very dependent on this threshold. Thus, we can consider the meaningfulness as a way to select the right edges. If Canny's filter were completed to provide us with pieces of curves, our algorithm could *a posteriori* decide which of them are meaningful. Notice that many Canny edges are found in flat regions of the image, where no perceptual boundary is present. If we increase the threshold, as is done on the right, the detected edges look perceptually more correct, but are broken.

3.2. *Cheetah Image (Fig. 3)*

This experiment compares our edge detector with the Mumford-Shah model. As before, we observe that the Mumford-Shah model produces some spurious boundaries on the background, due to the inadequacy of the piecewise constant model. This means that a more sophisticated model must be applied if we wish to avoid such spurious boundaries: the general Mumford-Shah model replaces the piece-wise constant constraint by a smoothness term (the Dirichlet integral $\int |\nabla u|^2(x)\, dx$) on each region. Now, adding this term means using a
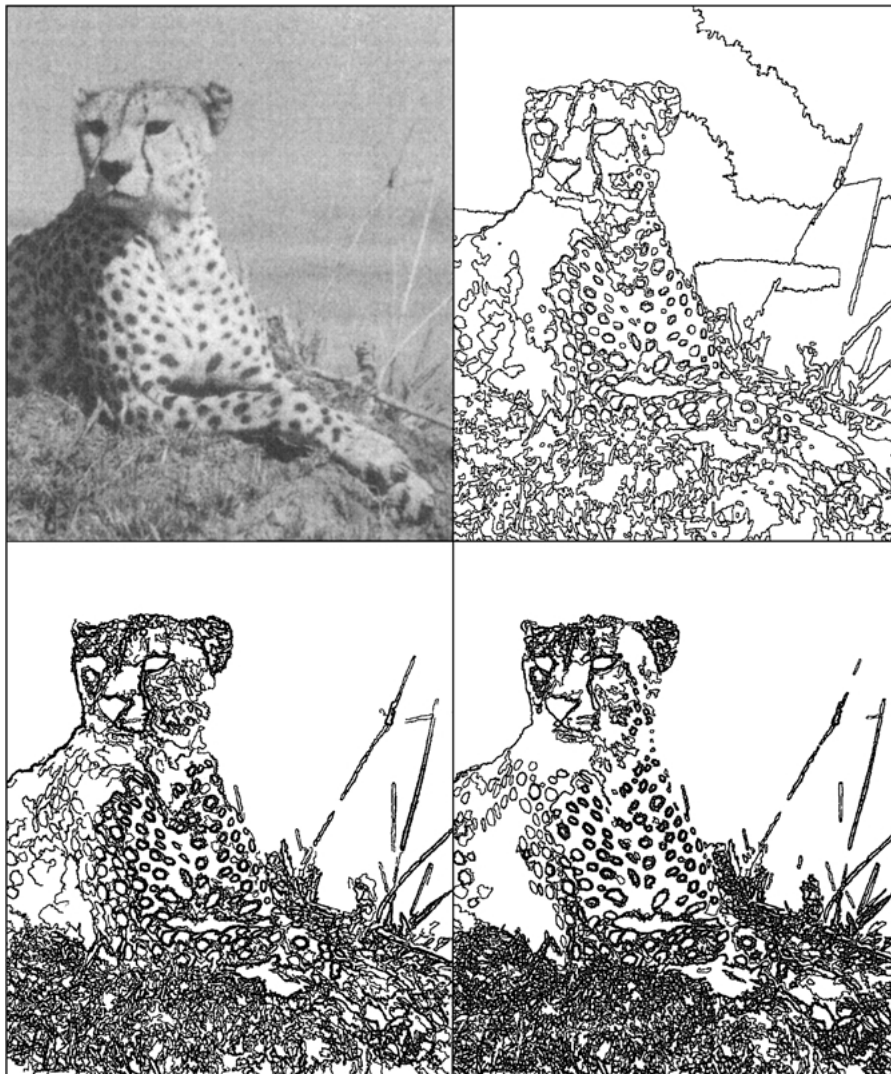
Figure 3.   First row: original image (left) and boundaries obtained with the Mumford-Shah model with 1000 regions (right). Second row: edges (left) and boundaries (right) obtained with our method ($\varepsilon = 1$).

two-parameters model since, then, the Mumford-Shah functional has three terms whose relative weights must be fixed.

3.3. DNA Image (Fig. 4)

This experiment illustrates the concept of "optimal boundaries" that we have introduce previously. When we compute the boundaries of the original image, each "spot" produces several parallel boundaries due to the important blur. With the definition of maximality we adopted, we select exactly one boundary for each spot.

3.4. Segments Image (Fig. 5)

As in the DNA experiment, the "optimal boundaries" allow to select exactly one boundary per object (here, hand-drawn segments). In particular, the number of boundaries we find (21) counts exactly the number of segments.

3.5. Noise Image (Fig. 6)

This image is obtained as a realization of a Gaussian noise with standart deviation 40. For $\varepsilon = 1$ and $\varepsilon = 10$,
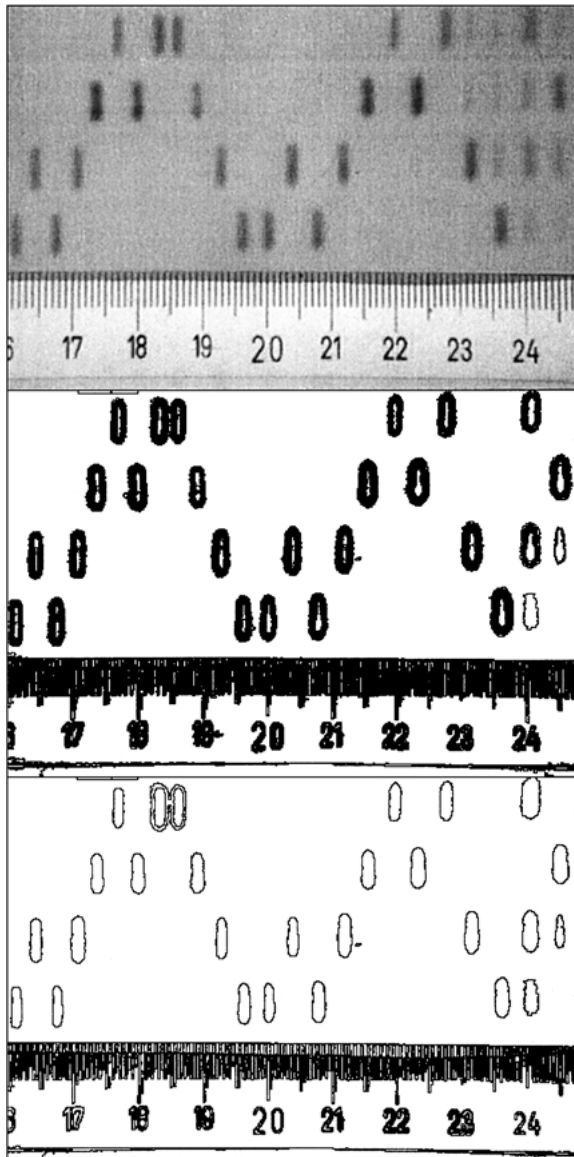
Figure 4. From top to bottom: 1. original image; 2. boundaries; 3. optimal boundaries.

Figure 5. Up: original image. Downleft: boundaries. Downright: optimal boundaries.

Figure 6. Left: an image of a Gaussian noise with standart deviation 40. Right: the meaningful boundaries found for $\varepsilon = 100$ (no boundaries are found for $\varepsilon = 1$).

no boundaries are detected. For larger values of ε, some boundaries begin to be detected: 7 for $\varepsilon = 100$ (see Fig. 6), 148 for $\varepsilon = 1000$ and 3440 for $\varepsilon = 10000$.

4. Discussion and Conclusion

In this discussion, we shall address objections and comments made to us by the anonymous referees and also by José-Luis Lisani, Yves Meyer and Alain Trouvé. In all that follows, we call respectively "boundary de-
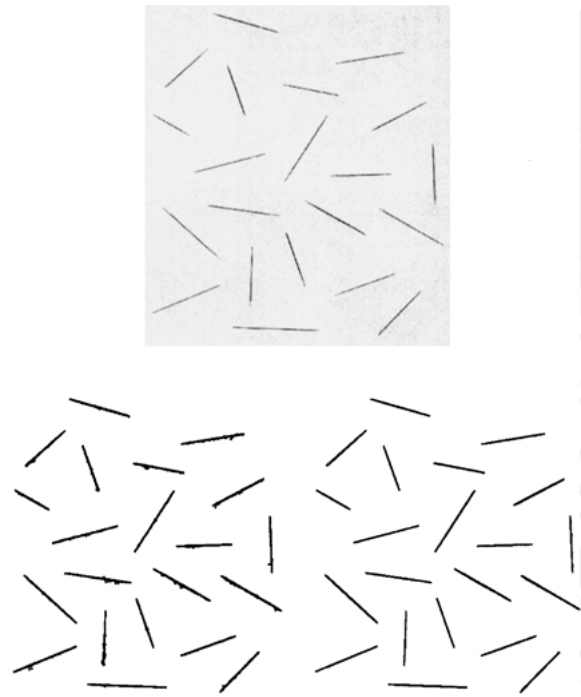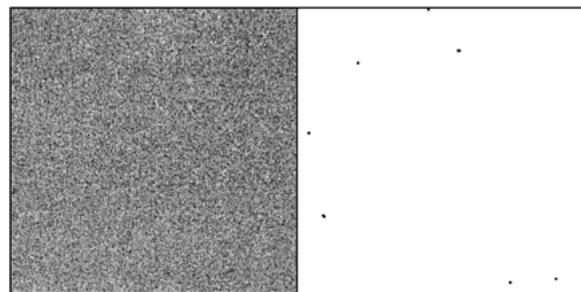
tection algorithm" and "edge detection algorithm" the algorithms we proposed. The other edge or boundary detection algorithms put into the discussion will be called by their author's names (Mumford-Shah, Canny).

4.1. Eleven Objections and their Answers

Objection 1: The blue sky effect. If a significant part of a natural image happens to be very flat, because of

a "blue sky effect", then **most** level lines of the image will be detected as meaningful. If (e.g.) one tenth of the image is a black flat region, then the histogram of the gradient has a huge peak near zero. Thus, all gradients slightly above this peak will have a probability $\frac{9}{10}$ significantly smaller than 1. As a consequence, all level lines long enough (with length larger than, say, 30 pixels) will be meaningful. In practice, this means that the image will be plagued with detected level lines with a small contrast. These detected level lines are no edges under any decent criterion?

Answer 1: If the image has a wide "blue sky", then most level lines of the ground are meaningful because any strong deviation from zero becomes meaningful. This effect can be checked on the cheetah image: the structured and contrasted ground has lots of detected boundaries (and the sky has none). This outcome can be interpreted in the following way: when a flat region is present in the image, it gives, via the gradient histogram, an indirect noise estimate. Every gradient which is above the noise gradient of the flat region becomes meaningful and this is, we think, correct.

Objection 2: Dependence upon windows. Then the detection of a given edge depends upon the window (containing the edge) on which you apply the algorithm?

Answer 2: Yes, the algorithm is global and is affected by a reframing of the image. If (e.g.) we detect edges on a window essentially containing the sky, we shall detect more boundaries (see Fig. 7) and if we compute edges in a window only containing the contrasted boundaries, it will detect less boundaries.

Question 3: How to compute edges with multiple windows? Thus, you can apply your detection algorithm on any window of the image and get more and more edges!

Answer 3: Yes, but, first, if the window is too small, no edge will be detected at all. Second, if we apply the algorithm to say, on 100 windows, we must take into account in our computations that the number of tests is increased. Thus, we must decrease accordingly the value of ε in order to avoid false detections: an easy way to do it is this: if we have 100 windows, we can take on each one $\varepsilon = 1/100$. Then the global number of false alarms over all windows
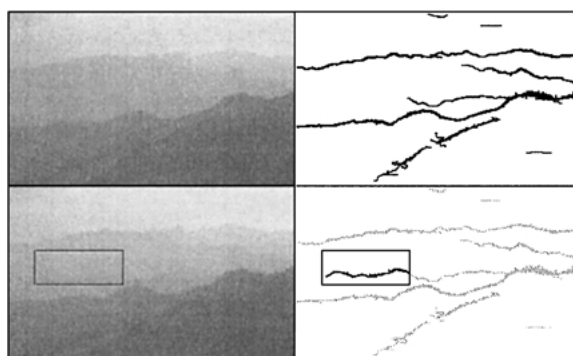

Figure 7. First row: left: original image (chinese landscape); right: maximal meaningful edges for $\varepsilon = 1$. Second row: the same algorithm, but run on a subwindow (drawn on the left image); right: the result (in black), with in light grey the edges that were detected in the full image.

remains equal to 1. Thus, a multiwindows version of the algorithm is doable and recommandable. Indeed, psychophysics and neurophysiology both advocate for a spatially local treatment of the retinian information.

Objection 4: Synthetic images where everything is meaningful. If an image has no noise at all (synthetic image), all boundaries contain relevant information. All the same, your algorithm won't detect them all?

Answer 4: Right. If a synthetic binary image is made (e.g.) of a black square with white background, then all gradients are zero except on the square's boundary. The gradient histogram has one single value, 255. (Remember that zero values are excluded from the gradient histogram). Thus, $H(255) = 1$ which means that no line is meaningful. Thus, the square's boundary won't be detected, which is a bit paradoxical! The addition of a tiny noise or of a slight blur would of course restore the detection of this square's boundary. This means that synthetic piecewise constant images fall out of the range or the detection algorithm. Now, in that case, the boundary detection is trivial by any other edge detector and our algorithm is not to be applied.

Question 5: Class of images to which the algorithm is adapted? Is there a class of images for which the Mumford-Shah functional is better adapted and another class of images where your algorithm is more adapted?

Answer 5: Our comparison of both algorithms may be misleading. We are comparing methods with different scopes. The Mumford-Shah algorithm aims at a global and minimal explanation of the image in terms of boundaries and regions. As we pointed out in the discussion of the experiments, this global model is robust but rough, and more sophisticated models would give a better explanation, provided the additional parameters can be estimated (but how?).

The detection algorithm does not aim at such a global explanation: it is a *partial detection* algorithm and not a *global explanation* algorithm. In particular, detected edges can be doubled or tripled or more, since many level lines follow a given edge. In contrast, the Mumford-Shah functional and the Canny detector attempt at selecting the *best representative* of each edge. Conversely, the detection algorithm provides a check tool to accept or reject edges proposed by any other algorithm.

Objection 6: The algorithm depends upon the quantization step. The algorithm depends upon the quantification step q. When q tends to zero, you will get more and more level lines. Thus N_{ll} and N_{llp} (numbers of level lines and pieces of level lines respectively) will blow up. Thus, you get less and less detections when q increases and, at the end, none!

Answer 6: Right again. The numbers N_{ll} and N_{llp} stand for the number of effectuated tests on the image. When the number of tests tends to infinity, the number of false alarms of Definition 1 also tends to infinity. Now, as we mentionned, q must be large enough in order to be sure that all edges contain at least one level line. Since the quantization noise is 1 and the standard deviation of noise never goes below 1 or 2, it is not likely to find any edge with contrast smaller than 2. Thus, $q = 1$ is enough, and we cannot miss any detectable edge. If we take q smaller, we shall get more spatial accuracy to the cost of less detections.

Question 7: Accuracy of the edges depends upon the quantization step. All the same, if q is not very small, you lose accuracy in the position detection. Indeed, the quantized levels do not coincide with the optimal level of the edge, as it would be found by a Canny edge detector.

Answer 7: Right again. The Canny edge detector performs two tasks in one: detecting and optimizing the edge's position at subpixel accuracy. The proposed detection algorithm does not find the optimal position of each edge. The spatial accuracy is roughly $q / \min |\nabla u|$, where the min is computed on the detected edge. In the case of the detection of optimal boundaries, we therefore get this spatial accuracy for the detected optimal boundaries. Of course, a postprocessing finding for each edge the best position in terms of detectability is possible.

Objection 8: Edges are not level lines. You claim that every edge coincides with some level line. This is simply not true!

Answer 8: If an edge has contrast kq, where q is the quantization step (usually equal to 1), then k level lines coincide with the edge, locally. Of course, one can construct long edges whose contrast is everywhere k but whose average level varies in such a way that no level line fully coincides with the edge. Now, long pieces of level lines coincide partially with it. Thus, detection of this edge by the detection algorithm is possible all the same, but it will be detected as a union of several more local edges.

Objection 9: Values of the gradient on the level lines are not independent. You chose as test set the set of all level lines. You claim that the gradient amplitudes at two different points of every edge are independent. This is, in most images, not true.

Answer 9: The independence assumption is, indeed, not a realistic assumption. It is made in order to apply the Helmholtz principle, according to which every large deviation from uniform randomness assumption is perceptible. Thus, the independence assumption is not a model for the image; it is an *a contrario* assumption against which the gestalts are detected.

Objection 10: A minimal description model would do the job as well. A minimal description model (MDL) can contain very wide classes of models for which parameters will be estimated by the MDL principle of shortest description in a fixed language [24, 28, 31]. This fixed language can be the language of Gestalt theory: explain the image in terms of lines, curves, edges, regions, etc. Then existence and

nonexistence of a given gestalt would come out from the MDL description: a "detectable" edge would be an edge which is used by the minimal description. Thus, thresholds would be implicit in a MDL model, but exist all the same.

Answer 10: A MDL model is global in nature. Until we have constructed it, we cannot make any comparison. In a MDL model, the thresholds on edges would depend on all other gestalts. Thus, we would be in the same situation as with the Mumford-Shah model: we have seen that a slight error on the region model leads to a false detection for edges. The main advantage of the proposed method relies on its lack of ambition: it is a *partial gestalt* detection algorithm, which does not require any global explanation model in order to be applied. We may compare the outcome of the algorithm with the computation in optimization theory of *feasible solutions*. Feasible solutions are not optimal. We provide *feasible, i.e. acceptable edges*. We do not provide an *optimal set of edges* as is aimed at by the other considered methods.

Objection 11: Is ε a method parameter? You claim that the method has no parameter. We have seen in the course of the discussion not less than three parameters coming out: the choice of the windows, the choice of q, and finally the choice of ε. So what?

Answer 11: We always fix $\varepsilon = 1$. Indeed, as we proved, the dependence of detectability upon ε is a Log-dependence. We also fix $q = 1$, but, here again, the q dependence would be a Log-dependence, since the number of level lines varies roughly linearly as a function of q. Finally, it is quite licit to take as many windows as we wish, provided we take $\varepsilon_k = 1/k$ where k is the number of windows. This yields a false alarm rate of 1 over all windows. Again, since the number of windows is necessarily small (they make a covering of the image and cannot be too small), we can even take $\varepsilon_k = 1$ because of the Log-dependence mentioned above. To summarize, $\varepsilon = 1$ is not a parameter. When we subdivide our set of tests in subsets on several windows, we must of course divide this value 1 by the number of sets of subtests. This does not require any user's input.

5. Conclusion

In this paper, we have tried to stress the possibility of giving a perceptually correct check for any boundary or edge proposed by any algorithm. Our method, based on the Helmholtz principle, computes thresholds of detectability for any edge. This algorithm can be applied to level lines or to pieces of level lines and computes then all detectable level lines. One cannot view the algorithm as a new "edge detector", to be added to the long list of existing ones; indeed, first, the algorithm does not select the "best" edge as the other algorithms do. Thus, it is more primitive and only yields "feasible" candidates to be an edge. Only in the case of boundary detection can it be claimed to give a final boundary detector. Now, this boundary detector may anyway yield multiple boundaries. On the other hand, the proposed method has the advantage of giving for any boundary or edge detector a sanity check.

Thus, it can, for any given edge detector, help removing all edges which are not accepted from the Helmholtz principle viewpoint. As a sanity check, the Helmholtz principle is hardly to be discussed, since it only rejects any edge which could be observed in white noise.

The number of false alarms gives, in addition, a way to evaluate the reliability of any edge and we think that the maximality criterion could also be used in conjunction with any edge detector.

Finally, we can claim that the kind of algorithm and experiments proposed here advocate for the necessity and usefulness of an intermediate layer in image analysis algorithms, where feasibility of the sought for structures is checked before any more global interpretation is attempted by a variational method.

Note

1. We choose to write gestalt(s) instead of the german original Gestalt (en). We maintain the german spelling for "Gestalt theory".

References

1. A.S. Abutaled, "Automatic thresholding of gray-level pictures using two-dimensional entropy," *Comp. Vision, Graphics and Image Processing*, Vol. 47, pp. 22–32, 1989.
2. J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 8, No. 6, pp. 679–698, 1986.

3. V. Caselles, B. Coll, and J.-M. Morel, "A Kanisza programme," *Progress in Nonlinear Differential Equations and their Applications*, Vol. 25, pp. 35–55, 1996.

4. L. Davis, "A survey of edge detection techniques," *Comp. Graphics and Image Processing*, Vol. 4, pp. 248–270, 1975.

5. R. Deriche, "Using Canny's criteria to derive a recursively implemented optimal edge detector," *Int. J. of Comp. Vision*, Vol. 1, No. 2, pp. 167–187, 1987.

6. A. Desolneux, L. Moisan, and J.-M. Morel, "Meaningful Alignments, in *Proceedings of SCTV'99*. Electronic publication available at http://www.cis.ohio-state.edu/~szhu/SCTV99.html, 1999.

7. A. Desolneux, L. Moisan, and J.-M. Morel, "Maximal meaningful events and applications to image analysis," preprint CMLA (http://www.cmla.ens-cachan.fr), 2000.

8. R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, 1973.

9. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, pp. 721–741, 1984.

10. G. Guy and G. Medioni, "Inferring global perceptual contours from local features," *Int. J. of Comp. Vision*, Vol. 20, No. 1, pp. 113–133, 1996.

11. J. Huang and D. Mumford, "Statistics of natural images and models," *Comp. Vision and Pattern Recognition*, Vol. 1, pp. 541–547, 1999.

12. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. of Comp. Vision*, Vol. 1, No. 4, pp. 321–331, 1988.

13. D. Lowe, *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers, The Netherlands, 1985.

14. A. Martelli, "Edge detection using heuristic search methods," *Comp. Graphics Image Processing*, Vol. 1, pp. 169–182, 1972.

15. P. Monasse, "Représentation morphologique d'images numériques et aplication au recalage d'images," Thèse de doctorat, Univ. Paris Dauphine, 2000.

16. J.-M. Morel and S. Solimini, *Variational Methods In Image Segmentation*, Birkhäuser, 1994.

17. D. Mumford and J. Shah, "Boundary detection by minimizing functionals," in *IEEE Conf. on Comp. Vision and Pattern Recognition*, San Francisco, 1985, pp. 22–25.

18. T. Pavlidis, "A critical survey of image analysis methods," in *IEEE Proc. of the 8th Int. Conf. on Pattern Recognition*, Paris, 1986, pp. 502–511.

19. T. Pun, "Entropic thresholding, a new approach," *Comp. Graphics and Image Processing*, Vol. 16, pp. 210–239, 1981.

20. A. Rosenfeld and M. Thurston, "Edge and curve detection for visual scene analysis," *IEEE Trans. Comput.*, Vol. 20, pp. 562–569, 1971.

21. M. Wertheimer, "Untersuchungen zur Lehre der Gestalt, II," *Psychologische Forschung*, Vol. 4, pp. 301–350, 1923.

22. J.S. Weszka, "A survey of threshold selection techniques," *Comp. Graphics and Image processing*, Vol. 7, pp. 259–265, 1978.

23. S.W. Zucker, "Region growing: Childhood and Adolescence (Survey)," *Comp. Graphics and Image Processing*, Vol. 5, pp. 382–399, 1976.

24. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, Viley, NY, 1991.

25. K.S. Fu and J.K. Mui, "A survey on image segmentation," *Pattern Recognition*, Vol. 13, pp. 3–16, 1981.

26. R.M. Haralick and L.G. Shapiro, "Image segmentation techniques," *Comp. Vision Graphics and Image Processing*, Vol. 29, pp. 100–132, 1985.

27. G. Kanizsa, *La Grammaire du Voir*, Editions Diderot, arts et sciences, 1994.

28. Y. Leclerc, "Constructing simple stable descriptions for image partitioning," *Int. J. of Comp. Vision*, Vol. 3, pp. 73–102, 1989.

29. W. Metzger, *Gesetze des Sehens*, Waldemar Kramer, Frankfurt Germany, 1975.

30. P. Parent and S.W. Zucker, "Trace inference, curvature consistency and curve detection," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 2, No. 8, pp. 823–839, 1989.

31. J. Rissanen, "A universal prior for integers and estimation by Minimum Description Length," *Annals of Statistics*, Vol. 11, No. 2, pp. 416–431, 1983.

32. D.L. Ruderman, "The statistics of natural images," *Network Computation in Neural Systems*, Vol. 5, pp. 517–548, 1994.

33. A. Sha'Ashua and S. Ullman, "Structural saliency: The detection of globally salient structures using a locally connected network," in *Proceedings of the 2nd Int. Conf. on Comp. Vision*, 1988, pp. 321–327.

34. A.P. Witkin and J.P. Tenenbaum, "On the role of structure in vision," Beck, Hope and Rosenfeld, Eds. (New York: Academic Press), pp. 481–543, 1983.

Agnès Desolneux is born in France in 1974. During the period 1994–1998, she has studied applied mathematics at the Ecole Normale Supérieure in Paris. From 1997 to 2000, she prepared a Ph.D. thesis about statistical methods in image analysis, at the CMLA (Cachan, France), and defended it in December 2000.

Lionel Moisan was born in Bègles, France, in 1970. From 1990 to 1995, he studied Mathematics and Computer Science in the Ecole Normale Supérieure de Paris. In 1997, he received the Ph.D. degree in applied mathematics from Ceremade, Paris-Dauphine University. He

has been teaching Mathematics in the Ecole Normale Supérieure de Cachan from 1995 to 1998. Since 1998, he is working at CMLA as a CNRS researcher. His fields of interest concern image processing and analysis with variational, PDE-based, and more recently statistical approaches.

Jean-Michel Morel got his Ph.D. in Applied Mathematics from the University Pierre et Marie Curie in 1980 and his Doctorat d'Etat in 1985. He has been Associate Professor of Applied Mathematics at Université Paris-Dauphine between 1985 and 1997. He is currently Professor of Mathematics at the Ecole Normale Supérieure de Cachan and Foreign Assistant at the University of Balearic Islands. His current interests are centered on the mathematical foundation of visual perception.