The Annals of Statistics
2003, Vol. 31, No. 6, 1822–1851

MAXIMAL MEANINGFUL EVENTS AND APPLICATIONS TO IMAGE ANALYSIS[1]

BY AGNÈS DESOLNEUX, LIONEL MOISAN AND JEAN-MICHEL MOREL

CMLA, ENS Cachan

We discuss the mathematical properties of a recently introduced method for computing geometric structures in a digital image without any a priori information. This method is based on a basic principle of perception which we call the Helmholtz principle. According to this principle, an observed geometric structure is perceptually "meaningful" if the expectation of its number of occurrences (in other words, its number of false alarms, NF) is very small in a random image. It is "maximal meaningful" if its NF is minimal among the meaningful structures of the same kind which it contains or is contained in. This definition meets the gestalt theory requirement that parts of a whole are not perceived. We explain by large-deviation estimates why this definition leads to an a priori knowledge-free method, compatible with phenomenology. We state a principle according to which maximal structures do not meet. We prove this principle in the large-deviations framework in the case of alignments in a digital image. We show why these results make maximal meaningful structures computable and display several applications.

1. Introduction. Digital images can represent any kind of outdoor or indoor scenes and can also be the result of astronomical, biological, medical,... experiments. Thus, no a priori model of any kind can be available for dealing with such undetermined data. Now, phenomenology tells us a different story. The Gestalt school pointed out the existence of stable, robust perceptions common to all humans. This school fixed qualitative "grouping processes" which should be universal in action on every image. According to gestalt theory, "grouping" is the main process in our visual perception (see [16], [22] and [35]). Whenever points (or previously formed visual objects) have a characteristic in common, they get grouped and form a new, larger visual object, a "gestalt." Some of the main grouping characteristics are color constancy, "good continuation," alignment, parallelism, common orientation, convexity and closedness (for a curve), among others. In addition, the grouping principle is recursive. For example, if points have been grouped into lines, then these lines may again be grouped according for example, to parallelism. The gestalt theory leads one to conjecture about the existence of parameter-less methods to compute these qualitative geometric primitives in *any* image.

Received May 2001; revised March 2003.

[1]Supported by Office of Naval Research Grant N00014-97-1-0839.

AMS 2000 subject classifications. 33B20, 62H15, 62H35, 62M40, 68U10, 68T45, 91E30.

Key words and phrases. Image analysis, perception, alignment, tail of the binomial distribution, rare events, large deviations.

These qualitative grouping principles can be formalized by the so-called Helmholtz principle. This principle attempts to describe when perception decides to group objects according to some quality (color, alignment, etc.). It can be stated in the following way. Assume that objects $O_1, O_2, \dots, O_n$ are present in an image. Assume that k of them, say $O_1, \dots, O_k$, have a common feature, say same color, same orientation and so on. We are then facing the dilemma: is this common feature happening by chance or is it significant? To answer this question, we make the following mental experiment: we assume that the considered quality has been randomly and uniformly distributed on all objects $O_1, \dots, O_n$. Notice that this quality may be spatial (e.g., position, orientation). Then we (mentally) assume that the observed position of objects in the image is a random realization of this uniform process, and ask the question: is the observed repartition probable or not? The Helmholtz principle states that if the expectation in the image of the observed configuration $O_1, \dots, O_k$ is very small, then the grouping of these objects makes sense, is a gestalt.

The Helmholtz principle can be illustrated by the psychophysical experiment of Figure 1. On the left, we display roughly 400 segments whose directional accuracy (computed as the width–length ratio) is about 12 degrees. Assuming that the directions and the positions of the segments are independent, uniformly distributed, we can compute the expectation of the number of alignments of four segments or more. (We say that segments are aligned if they belong to the same line, up to the given accuracy.) The expectation of such alignments in this case is about 2.5. Thus, we can expect two or three such alignments of four segments and we found them by computer. Do you see them? On the right, we performed the same experiment with about 30 segments, with accuracy (width–length ratio) equal to 7 degrees. The expectation of a group of four aligned segments is $1/250$. Most observers detect them immediately. This leads us to the following general definition of perceptual events (cf. [6]).

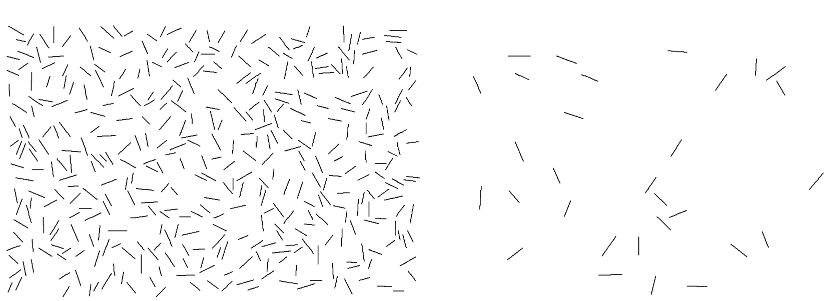

FIG. 1. *Illustration of Helmholtz principle in human perception. A group of four aligned segments exists in both images, but it can hardly be seen in the left image because such a configuration is not exceptional in view of the total number of segments. In the right image, we immediately perceive the alignment as a large deviation from randomness that could hardly happen by chance.*

DEFINITION 1 (ε-meaningful event). We say that an event of type "such configuration of points has such property" is ε-meaningful if the expectation in an image of the number of occurrences of this event is less than ε.

When $\varepsilon \leq 1$, we talk about meaningful events. If the Helmholtz principle is true, we perceive events if and only if they are meaningful in the preceding sense. The alignment on the right side of Figure 1 is meaningful while the left side of the figure contains no meaningful alignment of four segments.

Our plan is as follows. In Section 2, we explain our definition of meaningful alignments. Section 3 defines the "number of false alarms" and checks the consistency of this definition. In Section 4, we prove asymptotic (as $l \to \infty$) and nonasymptotic estimates about the meaningfulness of the following observation: "k well-aligned points in a segment of length l." Two main points are:

- We prove the $\log \varepsilon$ and $\log N$ dependence of this definition. This explains why we can always fix $\varepsilon = 1$ in practice and therefore get rid of this parameter.
- We show that detection is possible with a $\sqrt{l}$ excess of alignments in a segment of length l.

In Section 5, we prove that "detectability increases with resolution," a mathematical sanity check. Section 6 introduces "maximal meaningfulness." This is, to our knowledge, a first mathematical theory of the "masking" phenomenon in gestalt theory and a very practical tool for singling out the best explanation for each detectable alignment. Section 6 develops several mathematical and numerical arguments in favor of our main conjecture: two maximal meaningful segments on the same line have an empty intersection. In particular, it is proved that this conjecture is true in the large-deviations framework. In Section 7, we address the other remaining method parameter, namely, the precision p. We prove that it is useless to decrease this parameter, the roughest value being enough to get all detections. In Section 8, we end with joint numerical experiments, identifying maximal alignments in a digital image. Section 9 is devoted to an a posteriori bibliographical discussion and points out several extensions.

2. Definition of meaningful segments.

2.1. *The discrete nature of applied geometry.* Perceptual and digital images are the result of a convolution followed by a spatial sampling, as described in Shannon–Whittaker theory. From the samples, a continuous image may be recovered by Shannon interpolation, but the samples by themselves contain all of the image information. From this point of view, one could claim that no absolute geometric structure is present in an image, for example, no straight line, no circle, no convex set and so on. We claim, in fact, the opposite and our definition to follow will explain in which sense we can be "sure" that a line is present in a digital image.

Let us first explain what the basic local information is that we can dispose of in a digital image.

Let us consider a gray-level image of size N (i.e., a regular grid of N^2 pixels). At each point x, or pixel, of the discrete grid, we have a gray level $u(x)$ which is quantized and therefore inaccurate. We may compute at each point the normalized (unit norm) gradient, which is the simplest local contrast invariant information (local contrast invariance is a necessary requirement in image analysis and perception theory; see [22] and [35]). The direction vector $\mathrm{dir}(i, j)$ we attach to each point (i, j) of the image is simply obtained by a rotation of $\pi/2$ of the normalized gradient, so that it represents the local direction of the level line ($u =$ constant) passing through the current point. A simple finite-differences scheme gives

$$\mathrm{dir}(i, j) = \frac{1}{\|\vec{D}\|}\vec{D},$$

where

$$\vec{D} = \tfrac{1}{2}\begin{pmatrix} -u(i, j+1) - u(i+1, j+1) + u(i, j) + u(i+1, j) \\ u(i+1, j) + u(i+1, j+1) - u(i, j) - u(i, j+1) \end{pmatrix}.$$

Then we say that two points X and Y have the same direction with precision $1/n$ if

$$\bigl|\mathrm{Angle}\bigl(\mathrm{dir}(X), \mathrm{dir}(Y)\bigr)\bigr| \leq \frac{\pi}{n}.$$

In agreement with psychophysics and numerical experimentation, we consider that n should not exceed 16.

According to the Helmholtz principle, we treat the direction at all points in an image as a uniformly distributed random variable. In the following, we assume that $n > 2$ and we set $p = 1/n$ ($p < 1/2$); p is the accuracy of the direction. We interpret p as the probability that two independent points have the "same" direction with the given accuracy p. In a structure-less image, when two pixels are such that their distance is larger than 2, the directions computed at the two considered pixels should be independent random variables.

From now on, we do the computation, by the Helmholtz principle, as though each pixel had a direction which is uniformly distributed, two points at a distance larger than $q = 2$ having independent directions. Let A be a segment in the image made up of l independent pixels (this means that the distance between two consecutive points of A is 2 and so the real length of A is $2l$). We are interested in the number of points of A having their direction aligned with the direction of A. Such points of A will simply be called *aligned points of A*.

The question is: what is the minimal number $k(l)$ of aligned points that we must observe on a segment of length l so that this event becomes meaningful when it is observed in an image?

2.2. *Definition of meaning.* Let A be a straight segment with length l and let $x_1, x_2, \dots, x_l$ be the l (independent) points of A. Let X_i be the random variable whose value is 1 when the direction at pixel x_i is aligned with the direction of A, and 0 otherwise. We then have the following Bernoulli distribution for X_i:

$$\mathrm{P}[X_i = 1] = p \quad \text{and} \quad \mathrm{P}[X_i = 0] = 1 - p.$$

The random variable representing the number of x_i having the "good" direction is

$$S_l = X_1 + X_2 + \cdots + X_l.$$

Because of the independence of the X_i, the law of S_l is given by the binomial distribution

$$\mathrm{P}[S_l = k] = \binom{l}{k} p^k (1-p)^{l-k}.$$

When we consider a segment of length l, we want to know whether it is ε-meaningful or not among all the segments of the image (and not only among the segments having the same length l). Let $m(l)$ be the number of oriented segments of length l in an $N \times N$ image. We define the total number of oriented segments in an $N \times N$ image as the number of pairs (x, y) of points in the image (an oriented segment is given by its starting point and its ending point) and so we have

$$\sum_{l=1}^{l_{\max}} m(l) \simeq N^4.$$

DEFINITION 2 (ε-meaningful segment). A segment of length l is ε-meaningful in an $N \times N$ image if it contains at least $k(l)$ points having their direction aligned with the one of the segments, where $k(l)$ is given by

$$k(l) = \min\left\{k \in \mathbb{N}, \mathrm{P}[S_l \geq k] \leq \frac{\varepsilon}{N^4}\right\}. \tag{1}$$

Let us develop and explain this definition. For $1 \leq i \leq N^4$, let e_i be the following event: "the ith segment is ε-meaningful" and let χ_{e_i} denote the characteristic function of the event e_i. We have

$$\mathrm{P}\big[\chi_{e_i} = 1\big] = \mathrm{P}\big[S_{l_i} \geq k(l_i)\big],$$

where l_i is the length of the ith segment. Notice that if l_i is small we may have $\mathrm{P}[S_{l_i} \geq k(l_i)] = 0$. Let R be the random variable representing the exact number of e_i occurring simultaneously in a trial. Since $R = \chi_{e_1} + \chi_{e_2} + \cdots + \chi_{e_{N^4}}$, the expectation of R is

$$\mathrm{E}[R] = \mathrm{E}\big[\chi_{e_1}\big] + \mathrm{E}\big[\chi_{e_2}\big] + \cdots + \mathrm{E}\big[\chi_{e_{N^4}}\big] = \sum_{l=0}^{l_{\max}} m(l)\mathrm{P}[S_l \geq k(l)].$$

We compute here the expectation of R but not its law because it depends a lot on the relations of dependence between the e_i. The main point is that segments may intersect and overlap, so that the e_i events are not independent, and may even be strongly dependent.

By definition, we have

$$\mathrm{P}[S_l \geq k(l)] \leq \frac{\varepsilon}{N^4} \qquad \text{so that } \mathrm{E}[R] \leq \sum_{l=1}^{l_{\max}} \frac{\varepsilon}{N^4} m(l) \leq \varepsilon.$$

This means that the expectation of the number of ε-meaningful segments in an image is less than ε.

This notion of ε-meaningful segments has to be related to the classical "α-significance" in statistics, where α is simply ε/N^4. The difference which leads us to have a slightly different terminology is the following: we are not in a position to assume that the segments detected as ε-meaningful are independent in any way. Indeed, if, for example, a segment is meaningful it may be contained in many larger segments, which are also ε-meaningful. Thus, it will be convenient to compare the number of detected segments to the expectation of this number. This overcomes a difficulty raised by Stewart in [34] (see also the discussion in the last section). This is not exactly the same situation as in failure detection, where the failures are somehow disjoint events.

3. Number of false alarms.

3.1. *Definition.*

DEFINITION 3 (Number of false alarms). Let A be a segment of length l_0 with at least k_0 points having their direction aligned with the direction of A. We define the number of false alarms of A as

$$\mathrm{NF}(k_0, l_0) = N^4 \cdot \mathrm{P}\big[S_{l_0} \geq k_0\big] = N^4 \cdot \sum_{k=k_0}^{l_0} \binom{l_0}{k} p^k (1-p)^{l_0-k}. \tag{2}$$

The number $\mathrm{NF}(k_0, l_0)$ of false alarms of the segment A represents an upper bound of the expectation in an image of the number of segments of probability less than that of the considered segment.

3.2. *Properties of the number of false alarms.*

PROPOSITION 1. *The number of false alarms* $\mathrm{NF}(k_0, l_0)$ *has the following properties*:

1. $\mathrm{NF}(0, l_0) = N^4$, *which proves that the event for a segment to have more than zero aligned points is never meaningful.*

2. $\mathrm{NF}(l_0, l_0) = N^4 \cdot p^{l_0}$, *which shows that a segment such that all of its points have the "good" direction is ε-meaningful if its length is larger than* $(-4 \log N + \log \varepsilon)/\log p$.
3. $\mathrm{NF}(k_0 + 1, l_0) < \mathrm{NF}(k_0, l_0)$. *This can be interpreted as saying that if two segments have the same length* l_0, *the "more meaningful" one is the one that has the more "aligned" points*.
4. $\mathrm{NF}(k_0, l_0) < \mathrm{NF}(k_0, l_0 + 1)$. *This property can be illustrated by the following diagram of a segment* (*where* $\bullet$ *represents a misaligned point and* $\to$ *represents an aligned point*):

$$\to\to\ \bullet\ \to\to\ \bullet\bullet\ \to\to\to\to\to\ \bullet$$

If we remove the last point (*on the right*), *which is misaligned*, *the new segment is less probable and therefore more meaningful than the considered one*.
5. $\mathrm{NF}(k_0 + 1, l_0 + 1) < \mathrm{NF}(k_0, l_0)$. *Again*, *we can illustrate this property*:

$$\to\to\ \bullet\ \to\to\ \bullet\bullet\ \to\to\to\to\to\to$$

If we remove the last point (*on the right*), *which is aligned*, *the new segment is more probable and therefore less meaningful than the considered one*.

This proposition is an easy consequence of the definition and properties of the binomial distribution (see, e.g., [10]). If we consider a segment of length l (made up of l independent pixels), then the expectation of the number of points of the segment having the same direction as that of the segment is simply the expectation of the random variable S_l, that is,

$$\mathrm{E}[S_l] = \sum_{i=1}^{l} \mathrm{E}[X_i] = \sum_{i=1}^{l} \mathrm{P}[X_i = 1] = p \cdot l.$$

We are interested in ε-meaningful segments, which are the segments such that their number of false alarms is less than ε. These segments have a small probability (less than ε/N^4), and since they represent alignments (deviation from randomness), they should contain more aligned points than the expected number computed above. That is the main point of the following proposition.

PROPOSITION 2. *Let A be a segment of length $l_0 \geq 1$ containing at least k_0 points having the same direction as that of A. If* $\mathrm{NF}(k_0, l_0) \leq p \cdot N^4$ (*which is the case when A is meaningful because N is very large and thus $pN^4 < 1$*), *then*

$$k_0 \geq p l_0 + 1 - p.$$

This is a "sanity check" for the model. In the following, we will write $P(k, l)$ for $\mathrm{P}[S_l \geq k]$. Proposition 2 will be proved by Lemma 2, where we will extend the discrete function $P(k, l) = \mathrm{P}[S_l \geq k]$ to a continuous domain.

4. Orders of magnitude and asymptotic estimates for meaningfulness. In this section, we shall give precise asymptotic and nonasymptotic estimates of the thresholds $k(l)$, which roughly say that

$$k(l) \simeq pl + \sqrt{Cl \log \frac{N^4}{\varepsilon}},$$

where $2p(1-p) \le C \le 1/2$. Some of these results are illustrated in Figure 2. These estimates are not necessary for the algorithm [because $P(k,l)$ is easy to compute], but they provide an interesting order of magnitude for $k(l)$. In particular, we will see how the theory of large deviations and other inequalities concerning the tail of the binomial distribution can provide us with a sufficient condition for meaningfulness. In what follows, the precision $p < 1/2$ is fixed. We start with an estimate of the smallest length l of a detected alignment.

The first simple necessary condition we can get is a threshold on the length l. For an ε-meaningful segment, we have

$$p^l \le \mathrm{P}[S_l \ge k(l)] \le \frac{\varepsilon}{N^4},$$

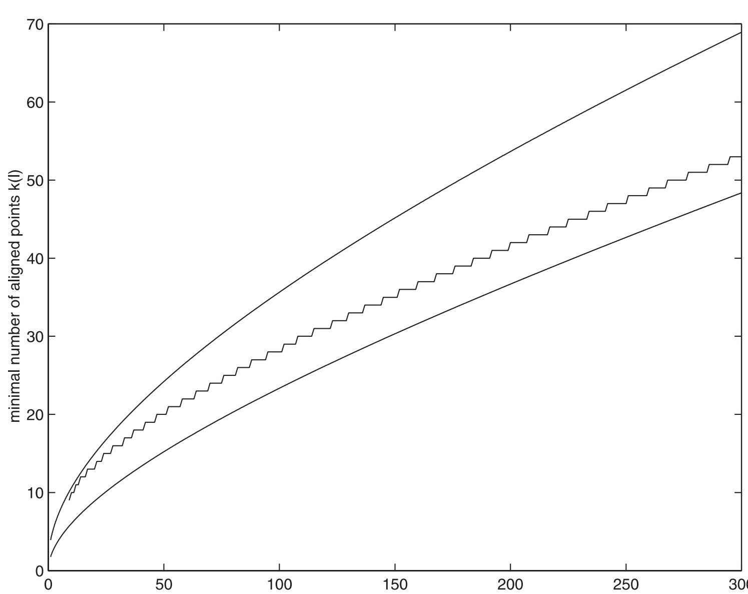

FIG. 2. *Estimates for the threshold of meaningfulness* $k(l)$. *The middle (step-case) curve represents the exact value of the minimal number of aligned points* $k(l)$ *to be observed on a* 1*-meaningful segment of length* l *in an image of size* 512×512 *for a direction precision of* $1/16$. *The upper and lower curves represent estimates of this threshold obtained by Propositions* 3 *and* 4.

so that

$$l \geq \frac{-4\log N + \log \varepsilon}{\log p}. \tag{3}$$

Let us give a numerical example: if the size of the image is $N = 512$, and if $p = 1/16$ (which corresponds to 16 possible directions), the minimal length of a 1-meaningful segment is $l_{\min} = 9$.

PROPOSITION 3 (Sufficient condition of ε-meaningfulness). *Let S be a segment of length l, containing at least k aligned points. If*

$$k \geq pl + \sqrt{\frac{4\log N - \log \varepsilon}{h(p)}}\sqrt{l},$$

where

$$h(p) = \frac{1}{1-2p}\log\frac{1-p}{p},$$

then S is ε-meaningful.

PROOF. This immediately follows from Hoeffding's inequality [14]. $\square$

Let us now work on asymptotic estimates of $P(k,l)$ when l is "large." Our aim is to get the asymptotic behavior of the threshold $k(l)$ when l is large. The problem is that if l tends to ∞, we also have to consider that N tends to ∞ (because, since l is the length of a segment in an $N \times N$ image, necessarily $l \leq \sqrt{2}N$).

PROPOSITION 4 [Asymptotic behavior of $k(l)$]. *When $N \to +\infty$ and $l \to +\infty$ in such a way that $l/(\log N)^3 \to +\infty$, one has*

$$k(l) = pl + \sqrt{2p(1-p)l\left(\log\frac{N^4}{\varepsilon} + O(\log\log N)\right)}. \tag{4}$$

PROOF. This directly follows from the following estimate (see, e.g., [10]): if $\alpha(l) \to +\infty$ and $\alpha(l)^6/l \to 0$ as $l \to +\infty$, then

$$\mathrm{P}\big[S_l \geq pl + \alpha(l)\sqrt{lp(1-p)}\,\big] \sim \frac{1}{\sqrt{2\pi}}\int_{\alpha(l)}^{+\infty} e^{-x^2/2}\,dx. \qquad \square$$

PROPOSITION 5 (Necessary condition for meaningfulness). *We assume that $0 < p \leq 1/4$ and N are fixed. If a segment $S = (k,l)$ is ε-meaningful, then*

$$k \geq pl + \alpha(N)\sqrt{lp(1-p)}, \tag{5}$$

where $\alpha(N)$ is uniquely defined by

$$\frac{1}{\sqrt{2\pi}}\int_{\alpha(N)}^{+\infty} e^{-x^2/2}\,dx = \frac{\varepsilon}{N^4}. \tag{6}$$

PROOF. This proposition is a direct consequence of Slud's theorem [33]: if $0 < p \le 1/4$ and $pl \le k \le l$, then

$$\mathrm{P}[S_l \ge k] \ge \frac{1}{\sqrt{2\pi}} \int_{\alpha(k,l)}^{+\infty} e^{-x^2/2}\, dx \qquad \text{where } \alpha(k,l) = \frac{k - pl}{\sqrt{lp(1-p)}}.$$

The assumption $0 < p \le 1/4$ is not a strong condition since it is equivalent to considering that the number of possible oriented directions is larger than 4. □

5. Properties of meaningful segments.

5.1. *Continuous extension of the binomial tail.* We first extend the discrete function $P(k,l)$ to a continuous domain (see [10]).

LEMMA 1. *The map*

$$\tilde{P} : (k,l) \mapsto \frac{\int_0^p x^{k-1}(1-x)^{l-k}\, dx}{\int_0^1 x^{k-1}(1-x)^{l-k}\, dx} \tag{7}$$

is continuous on the domain $\{(k,l) \in \mathbb{R}^2,\ 0 \le k \le l < +\infty\}$, *decreasing with respect to* k, *increasing with respect to* l, *and for all integer values of* k *and* l *one has* $\tilde{P}(k,l) = P(k,l)$.

PROOF. The continuity results from classical theorems on the regularity of parameterized integrals. Notice that the continuous extension of $\tilde{P}$ when $k = 0$ is $\tilde{P}(0,l) = 1$. The other properties are obtained by simple differentiation. □

The following property is a good example of the interest in the continuous extension of P. This yields a proof of Proposition 2.

LEMMA 2. *If* $l \ge 1$, *then* $p \le \tilde{P}(p(l-1)+1, l) < 1/2$.

The right-hand side of this inequality is a known result: it has been proved by Kaas and Buhrman [15]; indeed, they showed that, for the binomial distribution, the median and the mean are distant by no more than $\max(p, 1-p)$, which is $1-p$ in our case.

PROOF OF LEMMA 2. It is sufficient to prove that if $k - 1 = p(l-1)$, then

$$\begin{aligned} \frac{p}{1-p} \int_p^1 x^{k-1}(1-x)^{l-k}\, dx &\le \int_0^p x^{k-1}(1-x)^{l-k}\, dx \\ &< \int_p^1 x^{k-1}(1-x)^{l-k}\, dx. \end{aligned} \tag{8}$$

For that purpose, we write $f(x) = x^{k-1}(1-x)^{l-k}$ and we study the map

$$g(x) = \frac{f(p-x)}{f(p+x)}.$$

A simple computation gives that, up to a positive multiplicative term, $g'(x)$ can be written

$$2x^2\big(k-1-(1-p)(l-1)\big) - 2p(1-p)\big(k-1-p(l-1)\big),$$

and since $k-1 = p(l-1)$ and $p < 1/2$, we have $g' < 0$ on $]0, p]$. Hence, $g(x) < g(0) = 1$ on $]0, p]$, which implies

$$\begin{aligned}\int_0^p f(x)\,dx = \int_0^p f(p-x)\,dx &< \int_0^p f(p+x)\,dx\\ &= \int_p^{2p} f(x)\,dx < \int_p^1 f(x)\,dx,\end{aligned}$$

and the right-hand side of (8) is proved.

For the left-hand side, we follow the same reasoning with the map

$$g(x) = \frac{f(p-x)}{f(p+(1-p)/px)}.$$

After a similar computation, we obtain that $g' \geq 0$ on $]0, p]$, so that $f(p-x) \geq f(p+(1-p)/px)$ on $]0, p]$. We integrate this inequality to obtain

$$\begin{aligned}\int_0^p f(x)\,dx = \int_0^p f(p-x)\,dx &\geq \int_0^p f\left(p + \frac{1-p}{p}x\right)dx\\ &= \frac{p}{1-p}\int_p^1 f(x)\,dx,\end{aligned}$$

which proves the left-hand side of (8). $\square$

5.2. *Density of meaningful segments.* In general, it is not easy to compare $P(k,l)$ and $P(k',l')$ by performing simple computations on k, k', l and l'. Assume that we have observed a meaningful segment $S = (k,l)$ in an $N \times N$ image. We increase the resolution of the image in such a way that the new image has size $\lambda N \times \lambda N$, with $\lambda > 1$, and the considered segment is now $S_\lambda = (\lambda k, \lambda l)$ (we admit that the "density" of aligned points on the segment is scale invariant). Our aim is to compare the number of false alarms of S and of S_λ, that is, to compare

$$N^4 \tilde{P}(k,l) \quad \text{and} \quad (\lambda N)^4 \tilde{P}(\lambda k, \lambda l).$$

The result is given by the following proposition, and it shows that

$$\mathrm{NF}(S_\lambda) < \mathrm{NF}(S).$$

This is a consistency check for our model, since otherwise it would turn out that to get a better view does not increase the detection!

THEOREM 1. *Let $S = (k, l)$ be a 1-meaningful segment of an $N \times N$ image (with $N \geq 6$). Then the function defined for $\lambda \geq 1$ by*

$$\lambda \mapsto (\lambda N)^4 \tilde{P}(\lambda k, \lambda l)$$

is decreasing.

This theorem has the following corollary, which gives a way to compare the "meaningfulness" of two segments of the same image.

COROLLARY 1. *Let $A = (k, l)$ and $B = (k', l')$ be two 1-meaningful segments of an $N \times N$ image (with $N \geq 6$) such that*

$$\frac{k'}{l'} \geq \frac{k}{l} \quad \textit{and} \quad l' > l.$$

Then B is more meaningful than A, that is, $\mathrm{NF}(B) < \mathrm{NF}(A)$.

PROOF. Indeed, we can take $\lambda = l'/l > 1$, so that $k' \geq \lambda k$. We then have, by Theorem 1, $(\lambda N)^4 \tilde{P}(k', l') \leq N^4 \tilde{P}(k, l)$ and therefore $N^4 \tilde{P}(k', l') < N^4 \tilde{P}(k, l)$, that is, $\mathrm{NF}(B) < \mathrm{NF}(A)$. $\square$

COROLLARY 2. *The concatenation of two meaningful segments is more meaningful than the least meaningful of both.*

PROOF. Let $A = (k, l)$ and $B = (k', l')$ be two meaningful segments lying on the same line. We assume that A and B are consecutive, so that $A \cup B$ is simply a $(k + k', l + l')$ segment. Then, since

$$\frac{k + k'}{l + l'} \geq \min\left(\frac{k}{l}, \frac{k'}{l'}\right),$$

we deduce, owing to the above corollary, that $\mathrm{NF}(A \cup B) < \max(\mathrm{NF}(A), \mathrm{NF}(B))$. $\square$

The next lemma is useful to prove Theorem 1. We give its proof in the Appendix.

LEMMA 3. *Define, for $p < r \leq 1$, $B(r, l) = \tilde{P}(rl, l)$. Then one has*

$$\frac{1}{B} \frac{\partial \log B}{\partial l} < \frac{1}{l} - r \log \frac{r}{p} - (1 - r) \log \frac{1 - r}{1 - p}.$$

PROOF OF THEOREM 1. Let us define $r = k/l$ (note that, since S is 1-meaningful, we have $r > p$). We first claim that

$$r \log \frac{r}{p} + (1 - r) \log \frac{1 - r}{1 - p} \geq \frac{3 \log N}{l}. \tag{9}$$

Indeed, using Stirling's formula, we get

$$\begin{aligned}\frac{1}{N^4} \geq P(k,l) &\geq \binom{l}{k} p^k (1-p)^{l-k} \\ &\geq \frac{2e^{-1/6}}{\sqrt{2\pi l}} \exp\biggl(-l\biggl(r \log \frac{r}{p} + (1-r)\log\frac{1-r}{1-p}\biggr)\biggr),\end{aligned}$$

so that

$$l\biggl(r \log \frac{r}{p} + (1-r)\log\frac{1-r}{1-p}\biggr) \geq \log\left(\frac{CN^4}{\sqrt{l}}\right) \qquad \text{with } C = \frac{2e^{-1/6}}{\sqrt{2\pi}}.$$

Since $l \leq \sqrt{2}N$ (because l is the length of a segment of an $N \times N$ image) and $N \geq 6$, we have

$$\begin{aligned}\log\left(\frac{CN^4}{\sqrt{l}}\right) &\geq 3\log N + \log\sqrt{\frac{N}{l}} + \log\left(C\sqrt{N}\right) \\ &\geq 3\log N + \log\left(\frac{C\sqrt{6}}{2^{1/4}}\right) \geq 3\log N,\end{aligned}$$

so that (9) holds.

Now, let f be the function defined for $\lambda \geq 1$ by $f(\lambda) = (\lambda N)^4 \tilde{P}(\lambda k, \lambda l) = (\lambda N)^4 B(r, \lambda l)$. If we compute the derivative of f and use Lemma 3, we get

$$\begin{aligned}(\log f)'(\lambda) &= \frac{4}{\lambda} + l\frac{\partial \log B}{\partial l}(r, \lambda l) \\ &< \frac{4}{\lambda} + l\biggl(\frac{1}{\lambda l} - r\log\frac{r}{p} - (1-r)\log\frac{1-r}{1-p}\biggr) \\ &< \frac{5}{\lambda} - 3\log N,\end{aligned}$$

which is negative owing to the hypothesis $N \geq 6$. $\square$

REMARK. For the approximation of $\tilde{P}(k,l)$ given by the Gaussian law

$$G(k,l) = \frac{1}{\sqrt{2\pi}} \int_{\alpha(k,l)}^{+\infty} e^{-x^2/2}\,dx,$$

where

$$\alpha(k,l) = \left(\frac{k}{l} - p\right)\sqrt{\frac{l}{p(1-p)}},$$

we immediately have the result that $G(k', l') < G(k, l)$ when $k'/l' \geq k/l > p$ and $l' > l$.

6. Maximal meaningful segments.

6.1. *Definition.* Suppose that on a straight line we have found a meaningful segment S with a very small number of false alarms [i.e., $\mathrm{NF}(S) \ll 1$]. Then, if we add some "spurious" points at the end of the segment, we obtain another segment with probability higher than that of S and still having a number of false alarms less than 1, which means that this new segment is still meaningful (see the following diagram):

$$\to\to\to\to\to\to\to\to\to\to\to\to\to\to\to\to\to\to \bullet\bullet\bullet\bullet$$

In the same way, it is likely to happen in general that many subsegments of S having a probability higher than that of S will still be meaningful (see Section 8, where this problem obviously occurs for the "pencil strokes" image). These remarks justify the introduction of the following notion of "maximal segment."

DEFINITION 4 (Maximal segment). A segment A is maximal if:

(1) it does not contain a strictly more meaningful segment: $\forall\, B \subset A, \mathrm{NF}(B) \geq \mathrm{NF}(A)$;
(2) it is not contained in a more meaningful segment: $\forall\, B \supset A, \mathrm{NF}(B) > \mathrm{NF}(A)$.

Then we say that a segment is *maximal meaningful* if it is both maximal and meaningful. This notion of "maximal meaningful segment" is linked to what gestaltists called the "masking phenomenon." According to this phenomenon, most parts of an object are "masked" by the object itself except the parts that are significant from the point of view of the construction of the whole object.

PROPOSITION 6 (Properties of maximal segments). *Let A be a maximal segment. Then*:

(1) *the two endpoints of A have their direction aligned with the direction of A*;
(2) *the two points next to A* (*one on each side*) *do not have their direction aligned with the direction of A.*

This is an easy consequence of Proposition 1.

6.2. *A conjecture about maximality.* Up to now, we have established some properties that permit us to characterize or compare meaningful segments. We now study the structure of maximal segments and give some evidence that two distinct maximal segments on the same straight line have no common point.

CONJECTURE 1. If $(l, l', l'') \in [1, +\infty)^3$ and $(k, k', k'') \in [0, l] \times [0, l'] \times [0, l'']$, then

$$\begin{aligned} &\min\big(p, \tilde{P}(k, l), \tilde{P}(k + k' + k'', l + l' + l'')\big) \\ &\qquad < \max\big(\tilde{P}(k + k', l + l'), \tilde{P}(k + k'', l + l'')\big). \end{aligned} \tag{10}$$

Let us state immediately some relevant consequences of Conjecture 1.

COROLLARY 3 (Union and intersection). *If A and B are two segments on the same straight line*, *then*, *under Conjecture* 1,

$$\min\big(pN^4, \mathrm{NF}(A \cap B), \mathrm{NF}(A \cup B)\big) < \max\big(\mathrm{NF}(A), \mathrm{NF}(B)\big).$$

This is a direct consequence of Conjecture 1 for integer values of k, k', k'', l, l' and l''. Numerically, we checked this property for all segments A and B such that $|A \cup B| \leq 256$. For $p = 1/16$, we obtained

$$\begin{aligned} &\min_{|A\cup B| \leq 256} \frac{\max(\mathrm{NF}(A), \mathrm{NF}(B)) - \min(pN^4, \mathrm{NF}(A \cap B), \mathrm{NF}(A \cup B))}{\max(\mathrm{NF}(A), \mathrm{NF}(B)) + \min(pN^4, \mathrm{NF}(A \cap B), \mathrm{NF}(A \cup B))} \\ &\quad \simeq 0.000754697 \cdots > 0, \end{aligned}$$

this minimum (which is independent of N) being obtained for $A = (23, 243)$, $B = (23, 243)$ and $A \cap B = (22, 230)$ [as before, the pair (k, l) we attach to each segment represents the number of aligned points (k) and the segment length (l)].

THEOREM 2 (Maximal segments are disjoint under Conjecture 1). *Suppose that Conjecture* 1 *is true. Then any two maximal segments lying on the same straight line have no intersection.*

Notice that this property applies to maximal segments and not only to maximal meaningful segments.

PROOF OF THEOREM 2. Suppose that one can find two maximal segments $(k + k', l + l')$ and $(k + k'', l + l'')$ that have a nonempty intersection (k, l). Then, according to Conjecture 1, we have

$$\begin{aligned} &\min\big(p, P(k, l), P(k + k' + k'', l + l' + l'')\big) \\ &\qquad < \max\big(P(k + k', l + l'), P(k + k'', l + l'')\big). \end{aligned}$$

If the left-hand side term is equal to p, then we have a contradiction since one of $(k + k', l + l')$ or $(k + k'', l + l'')$ is strictly less meaningful than the segment $(1, 1)$ it contains. If not, we have another contradiction because one of $(k + k', l + l')$ or $(k + k'', l + l'')$ is strictly less meaningful than one of (k, l) or $(k + k' + k'', l + l' + l'')$. $\square$

REMARK. The numerical checking of Conjecture 1 ensures that, for $p = 1/16$ (but we could have checked for another value of p), two maximal meaningful segments with total length less than 256 are disjoint, which is enough for most practical applications.

The following proposition shows that Conjecture 1 can be deduced from a stronger (but simpler) conjecture: the concavity in a particular domain of the level lines of the natural continuous $\tilde{P}$ extension of P involving the incomplete Beta function.

PROPOSITION 7. *If the map* $(k,l) \mapsto \tilde{P}(k,l)$ *defined in Lemma* 1 *has negative curvature on the domain* $D_p = \{(k,l) \in \mathbb{R}^2,\ p(l-1)+1 \le k \le l\}$, *then Conjecture* 1 *is true*.

It is equivalent to say that the level curves $l \mapsto k(l,\lambda)$ of $\tilde{P}$ defined by $\tilde{P}(k(l,\lambda),l) = \lambda$ are concave, that is, satisfy

$$\forall\,(k_0,l_0) \in D_p, \qquad \frac{\partial^2 k}{\partial l^2}\bigl(l_0, \tilde{P}(k_0,l_0)\bigr) < 0.$$

PROOF OF PROPOSITION 7. We first prove that if $k-1 > p(l-1)$ and $\mu > 0$, then the map

$$x \mapsto \tilde{P}(k+\mu x, l+x)$$

has no local minimum at $x = 0$. Call this map f. Then it is sufficient to prove that either $f'(0) \neq 0$ or $[f'(0) = 0$ and $f''(0) < 0]$. If $f'(0) = 0$, then

$$\mu = -\frac{\tilde{P}_l}{\tilde{P}_k}(k,l),$$

so that

$$f''(0) = \mu^2 \tilde{P}_{kk} + 2\mu \tilde{P}_{kl} + \tilde{P}_{ll} = \operatorname{curv}(\tilde{P})(k,l) \cdot \frac{(\tilde{P}_k^2 + \tilde{P}_l^2)^{3/2}}{\tilde{P}_k^2} < 0.$$

We can now prove Proposition 7. Because the inequality we want to prove is symmetric in k' and k'', we can suppose that $k''/l'' \ge k'/l'$. If $k+k'-1 \le p(l+l'-1)$, then $\tilde{P}(k+k', l+l') > p$ and we are finished. Thus, in the following we assume $k+k'-1 > p(l+l'-1)$. Let us define the map

$$f(x) = \tilde{P}\bigl(k + x(k'+k''), l + x(l'+l'')\bigr) \qquad \text{for } x \in [0,1].$$

Note that, for $x_0 = l'/(l'+l'') \in \,]0,1[$, we have

$$k + x_0(k'+k'') = k + \frac{l'}{l'+l''}(k'+k'') \ge k + \frac{l'}{l'+l''}\left(k' + \frac{k'l''}{l'}\right) = k + k',$$

which implies that $\tilde{P}(k+k', l+l') \geq f(x_0)$. Hence, it is sufficient to prove that

$$\min\bigl(p, f(0), f(1)\bigr) < f(x_0).$$

The set

$$S = \bigl\{x \in [0,1], k + x(k'+k'') - 1 - p\bigl(l + x(l'+l'') - 1\bigr) > 0\bigr\}$$

is a connected segment that contains x_0 because

$$k + x_0(k'+k'') - 1 \geq k + k' - 1 > p(l+l'-1) = p\bigl(l + x_0(l'+l'') - 1\bigr).$$

Moreover, S contains 0 or 1 because the linear function involved in the definition of S is either 0 or vanishes only once. Since f has no local minimum on S, we conclude as announced that

$$f(x_0) > \min_{x \in S} f(x) = \min_{x \in \partial S} f(x) \geq \min\bigl(p, f(0), f(1)\bigr),$$

since if $x \in \partial S \cap]0,1[$, then $f(x) \geq p$ owing to Lemma 2. $\square$

All numerical computations we have realized so far for the function $\tilde{P}(k,l)$ have been in agreement with Conjecture 1. Concerning theoretical results, we shall see in the next section that this conjecture is asymptotically true. For now, the following results show that Conjecture 1 is satisfied for the Gaussian approximation of the binomial tail (correct for small deviations, that is, $k \simeq pl + C\sqrt{l}$) and also for the large-deviation estimate.

PROPOSITION 8. *The approximation of $P(k,l)$ given by the Gaussian law*

$$G(k,l) = \frac{1}{\sqrt{2\pi}} \int_{\alpha(k,l)}^{+\infty} e^{-x^2/2}\,dx, \qquad \textit{where } \alpha(k,l) = \frac{k - pl}{\sqrt{lp(1-p)}}$$

has negative curvature on the domain D_p.

PROOF. The level lines $G(k,l) = \lambda$ of $G(k,l)$ can be written in the form

$$k(l,\lambda) = pl + c(\lambda)\sqrt{l},$$

with $c > 0$ on the domain $\{k > pl\}$. Hence, we have

$$\frac{\partial^2 k}{\partial l^2}(l,\lambda) = -\frac{c(\lambda)}{4l^{3/2}}$$

and, consequently, $\operatorname{curv}(G) < 0$ on D_p. $\square$

THEOREM 3. *The large-deviation estimate of $P(k,l)$ (see, e.g.,* [4]*) given by*

$$H(k,l) = \exp\left[-k \log \frac{k}{pl} - (l-k) \log \frac{l-k}{(1-p)l}\right] \tag{11}$$

has negative curvature on the domain $\{pl \leq k \leq l\}$.

PROOF. The level lines of $H(k,l)$ are defined by

$$k(l,\lambda)\log\frac{k(l,\lambda)}{pl}+\big(l-k(l,\lambda)\big)\log\frac{l-k(l,\lambda)}{(1-p)l}=\lambda.$$

We keep λ fixed and just write $k(l,\lambda)=k(l)$. If we compute the first derivative of the above equation, we get, after simplification,

$$\begin{aligned}&k'(l)\log k(l)-k'(l)\log(pl)\\&\quad+\big(1-k'(l)\big)\log\big(l-k(l)\big)-\big(1-k'(l)\big)\log\big((1-p)l\big)=0.\end{aligned}$$

Now, again by differentiation,

$$k''(l)\log\frac{(1-p)k(l)}{p(l-k(l))}-\frac{1}{l}+\frac{k'(l)^2}{k(l)}+\frac{(1-k'(l))^2}{l-k(l)}=0,$$

which is equivalent to

$$k''(l)\log\frac{(1-p)k(l)}{p(l-k(l))}=-\frac{\big(k(l)-k'(l)l\big)^2}{lk(l)(l-k(l))},$$

so that $H(k,l)$ has negative curvature on the domain $pl\le k\le l$. $\square$

6.3. *Partial results about Conjecture* 1. In this section, we shall give an asymptotic proof of Conjecture 1. In all the following, we assume that p and r satisfy $0<p<r<1$ and $p<1/2$. The proof relies on the two following technical propositions.

PROPOSITION 9 (Precise large-deviation estimate). *Let*

$$\begin{aligned}&D(rl+1,l+1)\\&\quad=\frac{p(1-p)}{(r-p)\sqrt{2\pi lr(1-r)}}\exp\left[-l\left(r\log\frac{r}{p}+(1-r)\log\frac{1-r}{1-p}\right)\right].\end{aligned}\tag{12}$$

Then, for any positive p,r,l such that $p<r<1$ and $p<1/2$, one has

$$\begin{aligned}\frac{1-4r/((r-p)^2l(1-p))}{1+1/(r(1-r)\sqrt{2\pi lr(1-r)})}&\le\frac{\tilde P(rl+1,l+1)}{D(rl+1,l+1)}\\&\le\frac{1}{1-2/\sqrt{2\pi lr(1-r)}}.\end{aligned}\tag{13}$$

In particular, one has

$$\tilde P(rl+1,l+1)\underset{l\to+\infty}{\sim}D(rl+1,l+1)$$

uniformly with respect to r in any compact subset of $]p,1[$.

PROPOSITION 10. *For any* $\lambda \in [0, 1]$ *and* $l > 0$, *there exists a unique* $k(l, \lambda)$ *such that*

$$\tilde{P}\big(k(l,\lambda)+1, l+1\big) = \lambda. \tag{14}$$

Moreover, one has

$$\begin{aligned}&\frac{\partial^2 k}{\partial l^2}(l, \tilde{P}(rl+1, l+1)) \\ &\underset{l\to+\infty}{\sim} -\frac{(r\log(r/p)+(1-r)\log((1-r)/(1-p)))^2}{lr(1-r)(\log(r(1-p)/(1-r)p))^3}\end{aligned} \tag{15}$$

uniformly with respect to r *in any compact subset of* $]p, 1[$.

We shall not prove these results here: the proof is given in [5]; for more precise results, see [23]. It is interesting to notice that (15) remains true when $k(l, \lambda)$ is defined not from $\tilde{P}$ but from its estimate D given by (12). In the same way, one can prove that

$$\frac{\partial k}{\partial l}\big(l, \tilde{P}(rl+1, l+1)\big) \underset{l\to+\infty}{\longrightarrow} \frac{\log((1-p)/(1-r))}{\log(r(1-p)/(1-r)p)}$$

is satisfied by both definitions of $k(l, \lambda)$. This proves that (12) actually gives a very good estimate of $\tilde{P}$, since it not only approximates the values of $\tilde{P}$ but also its level lines up to second order.

THEOREM 4 (Asymptotic proof of Conjecture 1). *There exists a continuous map* $L :]p, 1[\to \mathbb{R}$ *such that* $(k, l) \mapsto \tilde{P}(k, l)$ *has negative curvature on the domain*

$$D_p^L = \big\{(rl+1, l+1),\ r \in]p, 1[,\ l \in [L(r), +\infty[\big\}.$$

PROOF. Define $k(l, \lambda)$ by (14). Owing to Proposition 10, the function

$$r \mapsto \frac{\partial^2 k}{\partial l^2}\big(l, \tilde{P}(rl+1, l+1)\big)\frac{lr(1-r)(\log(r(1-p)/(1-r)p))^3}{(r\log(r/p)+(1-r)\log((1-r)/(1-p)))^2}$$

tends to -1 as l goes to ∞, and the convergence is uniform with respect to r in any compact subset of $]p, 1[$. Thus, we deduce that the map

$$r \mapsto l(r) = \inf\big\{l_0 > 0, \forall l \geq l_0, \operatorname{curv} \tilde{P}(rl+1, l+1) < 0\big\}$$

is bounded on any compact subset of $]p, 1[$. Now, defining $L(r)$ as a continuous upper bound for $l(r)$ yields the desired result. For example, one can take

$$L(r) = \sup_{n\in\mathbb{Z}} d_n(r),$$

where d_n is the unique linear function passing through the points

$$\Big(a_{n-1}, \max_{t\in[a_{n-2},a_n]} l(t)\Big) \quad \text{and} \quad \Big(a_n, \max_{t\in[a_{n-1},a_{n+1}]} l(t)\Big),$$

and $(a_n)_{n\in\mathbb{Z}}$ an increasing sequence such that

$$\lim_{n\to-\infty} a_n = p \quad \text{and} \quad \lim_{n\to+\infty} a_n = 1. \qquad \square$$

7. About the precision p. In this section, we address the problem of the choice of the precision p. We show that it is useless to increase artificially the precision: this yields no better detection rates.

We consider a segment S of length l. We can assume that the direction of the segment is $\theta = 0$. Suppose that among the l points we observe k aligned points with given precision p (i.e., k points having their direction in $[-p\pi, +p\pi]$). Now, what happens if we change the precision p into $p' < p$? Knowing that there are k points with direction in $[-p\pi, +p\pi]$, we can assume (by the Helmholtz principle) that the average number of points having their direction in $[-p'\pi, p'\pi]$ is $k' = kp'/p$. The aim now is to compare

$$\mathcal{B}(l,k,p) \quad \text{and} \quad \mathcal{B}(l,k',p'),$$

where $\mathcal{B}(l,k,p) = \tilde{P}(k,l)$ for precision p [in the notation $P(k,l)$, we have omitted the precision p because it was fixed]. Since we are interested in meaningful segments, we will only consider the case

$$\lambda = \frac{k}{lp} = \frac{k'}{lp'} > 1.$$

We then have to study the function $p \mapsto \mathcal{B}(l, \lambda lp, p)$ and check that it is decreasing. This will be computed with the large-deviation estimate.

PROPOSITION 11. *If we consider the large-deviation estimate of* $\log \mathcal{B}(l,k,p)$ (*see also Theorem* 3), *given by*

$$G(l,\lambda lp,p) = -l\left(\frac{\lambda lp}{l}\log\frac{\lambda lp}{lp} + \left(1-\frac{\lambda lp}{l}\right)\log\frac{1-\lambda lp/l}{1-p}\right),$$

then $G(l,k,p) < G(l,k',p')$.

PROOF. We can easily prove that the function

$$p \mapsto \lambda p \log\lambda + (1-\lambda p)\log\frac{1-\lambda p}{1-p}$$

is increasing (for $\lambda > 1$). Consequently, $p \mapsto G(l,\lambda lp,p)$ decreases. $\square$

The consequences of Proposition 11 are as follows:

- If the observed alignment at precision $p' < p$ is meaningful, then the "original" alignment at precision p is more meaningful.
- The previous argument shows that we must always take the precision as coarse as possible.

REMARK. A natural question is: is $p \mapsto \mathcal{B}(l, \lambda lp, p)$ also decreasing?

8. Applications and experimental results. In all the following experiments, the direction at a pixel of an image is computed on a 2×2 neighborhood with the method described in Section 2.1 ($q = 2$), and the precision used is $p = 1/16$. The algorithm used to find the meaningful segments is the following. For each one of the four sides of the image, we consider for each pixel of the side the lines starting at this pixel, and having an orientation multiple of $\pi/200$. On each line, we compute the meaningful segments. For each segment, let l be its length counted in independent pixels (which means that the real length of the segment is $2l$). Then, among the l points, we count the number k of points having their direction aligned with the direction of the segment (with precision p). Finally, we compute $P(k, l)$: if it is less than ε/N^4, we say that the segment is ε-meaningful. Notice that $P(k, l)$ can simply be tabulated at the beginning of the algorithm using the relation

$$P(k+1, l+1) = pP(k, l) + (1 - p)P(k+1, l).$$

It must be made clear that we applied *exactly* the same algorithm to all presented images, which have very different origins. The only parameter of the algorithm is precision. We fixed it equal to $1/16$ in all experiments; this value corresponds to the very rough accuracy of 22.5 degrees; this means that, for example, two points can be considered as aligned with, say the 0 direction, if their angles with this direction are up to ± 11.25 degrees. It is clear that these bounds are very rough, but in agreement with the more pessimistic estimates for vision accuracy in psychophysics and with numerical experience as well. Moreover, in all experiments, we only keep the meaningful segments having in addition the property that their endpoints have their direction aligned with that of the segment.

IMAGE 1 [Pencil strokes (see Figure 3)]. This digital image was first drawn with a ruler and a pencil on a standard A4 white sheet of paper, and then scanned into a 478×598 digital image (left); note that a blur effect (about two pixels wide) is introduced by the scanner. Two pairs of pencil strokes are aligned on purpose. We display in the first experiment all ε-meaningful segments for $\varepsilon = 10^{-3}$ (middle image). Three phenomena occur, which are very apparent in this simple example, but will be perceptible in all further experiments.

1. Too-long meaningful alignments. We commented on this above; clearly, the pencil stroke boundaries are very meaningful, thus generating larger meaningful segments which contain them.

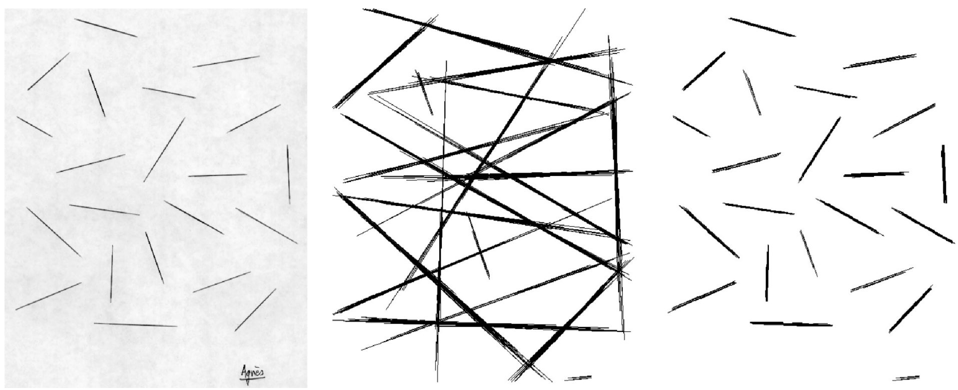

FIG. 3. *Pencil strokes image*. Left: *the original image*. Middle: *the ε-meaningful alignments for $\varepsilon = 10^{-3}$*. Right: *maximal meaningful alignments*.

2. Multiplicity of detected segments. On both sides of the strokes, we find several parallel lines (reminder: the orientation of lines is modulo 2π). These parallel lines are due to the blur introduced by the scanning process. Classical edge detection theory would typically select the best, in terms of contrast, of these parallel lines.
3. Lack of accuracy of the detected directions. We do not check that the directions found along a meaningful segment are distributed on both sides of the line direction. Thus, it is to be expected that we detect lines which are actually slanted with respect to the edge's "true" direction. Typically, a blurry edge will generate several parallel and more or less slanted alignments. It is not the aim of the actual algorithm to filter out this redundant information; indeed, we do not know at this point whether the detected parallel or slanted alignments are due to an edge or not: this must be the object of a more complex algorithm. Everything indicates that an edge is not an elementary phenomenon in gestalt.

We display in the second experiment for this image all maximal meaningful segments (right image), which shows for each stroke two bundles of parallel lines on each side of the stroke.

IMAGE 2 [Uccello's painting (see Figure 4)]. This image (left) is a result of the scan of Uccello's painting "Presentazione della Vergine al tempio" (from the book *L'opera completa di Paolo Uccello*, Classici dell'arte, Rizzoli). The size of this image is 467×369. The right image displays all maximal ε-meaningful segments found with $\varepsilon = 10^{-6}$. Notice how maximal segments are detected on the staircase in spite of the occlusion by the child up the stairs. All remarks made in Image 1 apply here (parallelisms due to the blur, etc.).

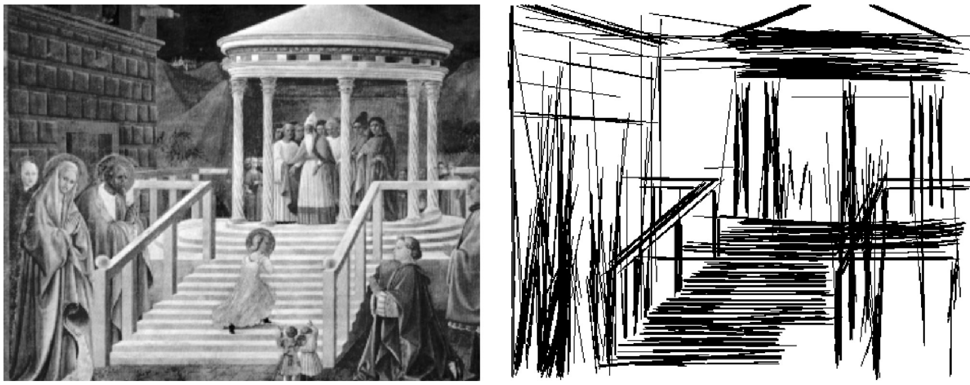

FIG. 4. *Uccello's painting*. Left: *the original image*. Right: *maximal ε-meaningful segments for* $\varepsilon = 10^{-6}$.

IMAGE 3 [The church of Valbonne (Figure 5, left)]. On the right, we display the set of maximal ε-meaningful segments of the original image for $\varepsilon = 10^{-3}$. We can again here make the same remarks as for Images 1 and 2.

FIGURE 6. We first add Gaussian white noise with standard deviation 50 to the original "pencil strokes" image (which has mean value 238 and standard deviation 17.7); thus, this corresponds to a signal-to-noise ratio (SNR) of 0.35. In the middle of Figure 6, we display the maximal ε-meaningful alignments for $\varepsilon = 1$: there are no false detections (except one segment which is too long) and on the right of Figure 6, we display the maximal ε-meaningful alignments for $\varepsilon = 100$:

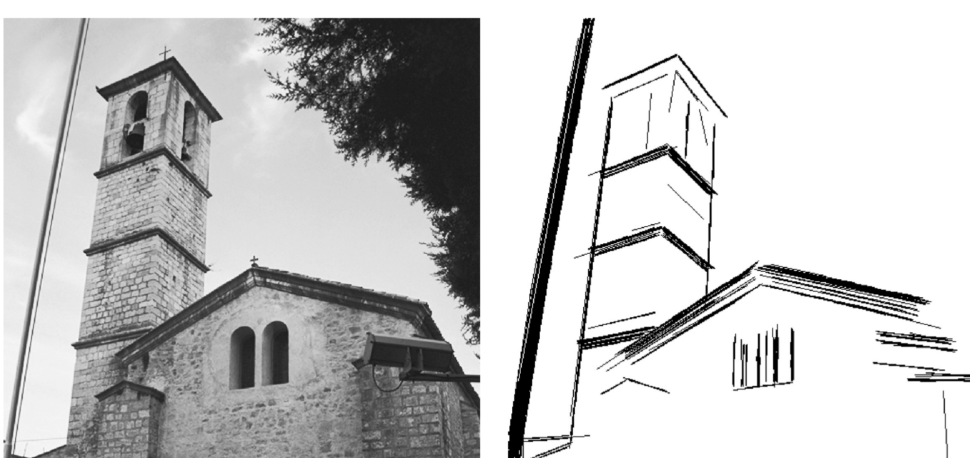

FIG. 5. Left: *the original* "*church of Valbonne*" *image* (*source*: *INRIA*). Right: *maximal ε-meaningful alignments found for* $\varepsilon = 10^{-3}$.

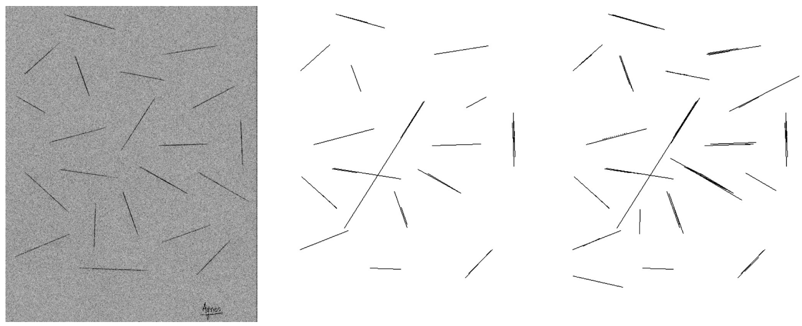

FIG. 6. Left: *the pencil strokes image corrupted by additive Gaussian noise with standard deviation* 50, *yielding a signal-to-noise ratio of* 0.35. Middle: *the maximal* ε*-meaningful alignments for* $\varepsilon = 1$ (*no false detection found*). Right: *the maximal* ε*-meaningful alignments for* $\varepsilon = 100$ (*one false detection occurs at the bottom left*).

one observes a false detection (at the bottom left). This experiment shows that the choice of ε is sharp.

9. Discussion, comparison with related works and conclusions. In this discussion, we start by comparing the results and method presented here with Stewart's seminal work. In continuation, we shall enlarge the discussion and focus it on the main question: the relationship to variational methods and high-level methods. We shall finally mention several extensions of this work.

The method we have presented in this paper can be viewed as a systematization of Stewart's MINPRAN method [34]. The method was presented as a new paradigm, but applied to the three-dimensional (3D) alignment problem. It is worth describing the method to explain what we added to it. Stewart's data are l points disseminated in a 3D cube. His question is: how to find planes along which those dots accumulate, that is, 3D alignments? When a hypothesis of 3D alignment on a plane P is generated, let us call r the distance to the plane and consider the event: "at least k points among the l randomly fall at a distance less than r from P." The probability of the event is, calling z_0 the maximal distance to the plane and setting $p = r/z_0$,

$$\mathcal{B}(l, k, p) = \sum_{i=k}^{l} \binom{l}{i} p^i (1-p)^{l-i}.$$

Then, for a given plane P, Stewart computes the minimal probability of alignment over all r's, that is,

$$H(P, l) = \min_r \mathcal{B}\left(l, k_{P,r}, \frac{r}{z_0}\right),$$

where $k_{P,r}$ is the number of observed dots at distance less than r from P. A number S of hypothesized planes being fixed, "MINPRAN accepts the best fit from S samples as correct if

$$\min_{1\le j\le S} H(P_j, l) < H_0,$$

where H_0 is a threshold based on the probability P_0 that the best fit to l uniformly distributed outliers is less than H_0. Intuitively, P_0 is the probability MINPRAN will hallucinate a fit where there is none. Thus, for a user-defined value of P_0 (e.g., $P_0 = 0.05$) we establish our threshold value H_0.... To make the analysis feasible, we assume the S fits and their residuals are independent. Strictly speaking, this assumption is not correct since the point set is the same for all fits. It is reasonable for relatively small values of S.... Doing this requires several parameters to be specified by the user. These parameters... are the estimated maximum fraction of true outliers..., the minimum number of points to be allowed in a fit, and the estimated maximum number of correct fits."

We see that Stewart's method starts exactly as we propose. Stewart actually addresses but does not solve the two problems we intended to overcome. One is the generation of the set of samples, which generates in Stewart's method at least three user parameters, and the second one is the severe restriction about the *independence of samples*. We actually solved both difficulties simultaneously by introducing the number of samples as an implicit parameter of the method (computed from the image size and Shannon's principles) and by replacing in all calculations the "probability of hallucinating a wrong event" by the "expectation of the number of such hallucinations," namely, what we call the number of false alarms. The application of our method to Stewart's problem is briefly explained, in the 2D case, in [8].

The method we presented here can be viewed as a complement to the so-called "Hough transform" in image analysis (see [20]). Let us first describe the basic Hough transform. We assume that the image under analysis is made up of dots which may create aligned patterns or not. We then compute for each straight line in the image the number of dots lying on the line. The result of the Hough transform is then a map associating with each line its number of dots. Then "peaks" of the Hough transform may be computed: they indicate the lines which have more dots. Which peaks are significant? Clearly, a threshold must be used. For today's technology, this threshold is generally given by a user or learned. The work of Kiryati, Eldar and Bruckstein [17] and Shaked, Yaron and Kiryati [32] is, however, very close to what we develop here: these authors prove by large-deviation estimates that lines in an image detected by the Hough transform could be detected as well in an undersampled image without increasing significantly the false alarm rate. They view this method as an accelerator tool, while we develop it here as a geometric definition tool. Actually, in a slightly different framework, our results of Section 5.2 arrive at the same conclusion.

The presence of ε and $p = 1/n$ in our definitions seems to contradict our notion of an a priori knowledge-free theory. Now, it does not, since we have proved that the ε-dependency of meaningfulness is low (it is, in fact, a $\log \varepsilon$-dependency). Our term ε-meaningful is related to the classical p-significance in statistics; as we have seen, we must use expectations in our estimates and not probabilities. We refer to [6] for a complete discussion of this definition. Concerning p, one can, as has been done in an application of our work by Almansa, Desolneux and Vamech [1], proceed as follows: these authors consider q possible values for p, test them all and simply divide ε by q. In that way, no a priori knowledge on the precision p is required.

Algorithms of Computer Vision belong to two classes. Many are application directed and contain a lot of a priori knowledge. They are usually classified as high-level vision algorithms, as opposed to low-level vision algorithms. Low-level vision algorithms assume, as in gestalt theory, that some general principles can deliver basic structures of images (see [21]). For instance, the Hough transform belongs to low-level vision. Now, in many papers, a priori knowledge is used even to find basic structures such as curves and boundaries (see the papers by Sha'Ashua and Ullman [31] and Yuille, Coughlan, Wu and Zhu [36], the extension field of Guy and Medioni [13] and the Parent–Zucker curve detector [26]). These methods actually contain parameters which are fixed by the user according to some a priori knowledge of the type of images. These methods do not yield an *existence proof* for the found structures.

Of course, the variational framework is well adapted to high-level vision. The general idea is to minimize with respect to the model M a functional of the kind

$$F(M, u_0) + R(M),$$

where u_0 is the given image defined on a domain $\Omega \subset \mathbb{R}^2$, $F(M, u_0)$ is a fidelity term and $R(M)$ is a regularity term containing the a priori knowledge. Let us give two examples: the Mumford–Shah model (see [24] and [25]) and the Bayesian model (see [12]). Another possibility, which turns out to be a significant improvement of Bayesian methods, is the minimal description length (MDL) method introduced by Rissanen [28] and first applied in image segmentation by Leclerc [18]. If the task is to recognize, say fire, water, clouds and grass, it is clear that the meaningfulness method developed here is not relevant and we clearly need a priori knowledge, learned or explicitly given in a Bayesian model. Now, the texture discrimination problem, namely, the question of whether grass and clouds look the same or not, might be treated by meaningfulness arguments. To summarize, what we propose here has nothing to do with "real world" object recognition, but only with the detection of basic structures.

Notice that all of the work described in the present paper does not address at all the restoration problem. Indeed, the restoration problem assumes, like the variational methods, a prior on the class of images, on the noise, on the blurring kernel and so on. Chu, Glad, Godtliebsen and Marron [3] have formulated an important and practical problem: how to smooth the noise out of image data while at the same time preserving unsmooth features such as jumps, spikes and edges? This is also addressed by Qiu [27] who proposes fitting discontinuous regression surfaces in the presence of noisy data. The a priori model for the image is a piecewise C^1 function. We can also mention the paper of Rudin, Osher and Fatemi [30] and the seminal paper by Donoho and Johnstone [9]. All of these papers address restoration of a signal with an a priori functional model (Besov, piecewise $C^1, \dots$) in the presence of (known) noise.

Since the submission of the present paper, new applications supporting the generality of the method presented here have been developed. We have defined in a similar framework several gestalts as maximal meaningful events, namely, meaningful edges and boundaries in [7], maximal meaningful histogram modes and meaningful spatial groups of objects (the so-called "proximity" gestalt) in [8]. Cao [2] applied the very same method to detect "good curves" in the sense of gestalt theory.

APPENDIX

PROOF OF LEMMA 3. We first write the Beta integral in terms of the Gamma function

$$\int_0^1 t^{x-1}(1-t)^{y-1}\,dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

Owing to (7), this yields

$$B(r,l) = \frac{\Gamma(l+1)}{\Gamma(rl)\Gamma((1-r)l+1)} \int_0^p x^{rl-1}(1-x)^{(1-r)l}\,dx. \tag{16}$$

We now use the expansion

$$\frac{d\log\Gamma(x)}{dx} = -\gamma - \frac{1}{x} + \sum_{n=1}^{+\infty}\left(\frac{1}{n} - \frac{1}{x+n}\right), \tag{17}$$

where γ is Euler's constant. Using (16) and (17), we obtain

$$\begin{aligned}\frac{1}{B}\frac{\partial B}{\partial l} &= -\gamma - \frac{1}{l+1} + \sum_{n=1}^{+\infty}\left(\frac{1}{n} - \frac{1}{l+1+n}\right)\\ &\quad - r\left[-\gamma - \frac{1}{rl} + \sum_{n=1}^{+\infty}\left(\frac{1}{n} - \frac{1}{rl+n}\right)\right]\\ &\quad - (1-r)\left[-\gamma - \frac{1}{(1-r)l+1} + \sum_{n=1}^{+\infty}\left(\frac{1}{n} - \frac{1}{(1-r)l+1+n}\right)\right]\\ &\quad + \frac{\int_0^p (r\log x + (1-r)\log(1-x))x^{rl-1}(1-x)^{(1-r)l}\,dx}{\int_0^p x^{rl-1}(1-x)^{(1-r)l}\,dx}.\end{aligned}$$

The function $x \mapsto r \log x + (1-r)\log(1-x)$ is increasing on $]0, r[$, and we have $p < r$, so

$$\frac{\int_0^p (r\log x + (1-r)\log(1-x)) x^{rl-1}(1-x)^{(1-r)l}\,dx}{\int_0^p x^{rl-1}(1-x)^{(1-r)l}\,dx} \\ \leq r\log p + (1-r)\log(1-p).$$

Then

$$\frac{1}{B}\frac{\partial B}{\partial l} \leq \frac{1}{l} + \sum_{n=1}^{+\infty}\left(\frac{r}{rl+n} + \frac{1-r}{(1-r)l+n} - \frac{1}{l+n}\right) + r\log p + (1-r)\log(1-p).$$

Now, let us consider the function

$$f : x \mapsto \frac{r}{rl+x} + \frac{1-r}{(1-r)l+x} - \frac{1}{l+x},$$

defined for all $x > 0$. Since $0 < r \leq 1$, we have $rl + x \leq l + x$ and $(1-r)l + x \leq l + x$, so that $f(x) \geq 0$ and

$$f'(x) = -\frac{r}{(rl+x)^2} - \frac{1-r}{((1-r)l+x)^2} + \frac{1}{(l+x)^2} \leq 0.$$

We deduce that, for N an integer larger than 1,

$$\sum_{n=1}^{N} f(n) \leq \int_0^N f(x)\,dx.$$

A simple integration gives

$$\int_0^N f(x)\,dx = r\log\left(1 + \frac{rl}{N}\right) + (1-r)\log\left(1 + \frac{(1-r)l}{N}\right) - \log\left(1 + \frac{l}{N}\right) \\ - r\log r - (1-r)\log(1-r).$$

Finally,

$$\sum_{n=1}^{+\infty}\left(\frac{r}{rl+n} + \frac{1-r}{(1-r)l+n} - \frac{1}{l+n}\right) \leq -r\log r - (1-r)\log(1-r),$$

which yields

$$\frac{1}{B}\frac{\partial B}{\partial l} \leq \frac{1}{l} - r\log r - (1-r)\log(1-r) + r\log p + (1-r)\log(1-p). \qquad \square$$

Acknowledgments. We thank Richard Lau for his personal interest and encouragement. We thank Jean Bretagnolle, Nicolas Vayatis, Frédéric Guichard, Isabelle Gaudron-Trouvé, Guillermo Sapiro and the anonymous reviewers for their valuable suggestions and comments.

REFERENCES

[1] ALMANSA, A., DESOLNEUX, A. and VAMECH, S. (2003). Vanishing point detection without any a priori information. *IEEE Trans. Pattern Anal. Machine Intelligence* **25** 502–507.

[2] CAO, F. (2002). Contrast invariant detection of good continuations, corners and terminators. INRIA Research Report 4542.

[3] CHU, C. K., GLAD, I. K., GODTLIEBSEN, F. and MARRON, J. S. (1998). Edge preserving smoothers for image processing (with discussion). *J. Amer. Statist. Assoc.* **93** 526–556.

[4] DEMBO, A. and ZEITOUNI, O. (1993). *Large Deviations Techniques and Applications*. Jones and Bartlett, Boston.

[5] DESOLNEUX, A., MOISAN, L. and MOREL, J.-M. (1999). Meaningful alignments. Preprint. Available at www.cmla.ens-cachan.fr/.

[6] DESOLNEUX, A., MOISAN, L. and MOREL, J.-M. (2000). Meaningful alignments. *Internat. J. Comput. Vision* **40** 7–23.

[7] DESOLNEUX, A., MOISAN, L. and MOREL, J.-M. (2001). Edge detection by Helmholtz principle. *J. Math. Imaging Vision* **14** 271–284.

[8] DESOLNEUX, A., MOISAN, L. and MOREL, J.-M. (2003). A grouping principle and four applications. *IEEE Trans. Pattern Anal. Machine Intelligence* **25** 508–513.

[9] DONOHO, D. L. and JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1224.

[10] FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications* **1**, 3rd ed. Wiley, New York.

[11] FISCHLER, M. A. and BOLLES, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM* **24** 381–395.

[12] GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intelligence* **6** 721–741.

[13] GUY, G. and MEDIONI, G. (1996). Inferring global perceptual contours from local features. *Internat. J. Comput. Vision* **20** 113–133.

[14] HOEFFDING, W. (1963). Probability inequalities for sum of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[15] KAAS, R. and BUHRMAN, J. M. (1980). Mean, median and mode in binomial distributions. *Statist. Neerlandica* **34** 13–18.

[16] KANIZSA, G. (1994). *La grammaire du voir*. Diderot, Paris.

[17] KIRYATI, N., ELDAR, Y. and BRUCKSTEIN, A. M. (1991). A probabilistic Hough transform. *Pattern Recognition* **24** 303–316.

[18] LECLERC, Y. (1989). Constructing simple stable descriptions for image partitioning. *Internat. J. Comput. Vision* **3** 73–102.

[19] LOWE, D. (1985). *Perceptual Organization and Visual Recognition*. Kluwer, Dordrecht.

[20] MAÎTRE, H. (1985). Un panorama de la transformation de Hough. *Trait. Signal* **2** 305–318.

[21] MARR, D. (1982). *Vision*. Freeman, New York.

[22] METZGER, W. (1975). *Gesetze des Sehens*. Waldemar Kramer, Frankfurt.

[23] MOISAN, L. (2003). Asymptotic estimates and inequalities for the tail of the binomial distribution. Unpublished manuscript.

[24] MUMFORD, D. and SHAH, J. (1985). Boundary detection by minimizing functionals. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[25] NITZBERG, N., MUMFORD, D. and SHIOTA, T. (1993). *Filtering, Segmentation and Depth. Lecture Notes in Comput. Sci.* **662**. Springer, Berlin.

[26] PARENT, P. and ZUCKER, S. W. (1989). Trace inference, curvature consistency and curve detection. *IEEE Trans. Pattern Anal. Machine Intelligence* **11** 823–839.

[27] QIU, P. (1998). Discontinuous regression surfaces fitting. *Ann. Statist.* **26** 2218–2245.
[28] RISSANEN, J. (1983). A universal prior for integers and estimation by minimum description length. *Ann. Statist.* **11** 416–431.
[29] ROUSSEEUW, P. J. and LEROY, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.
[30] RUDIN, L., OSHER, S. and FATEMI, E. (1992). Nonlinear total variation based noise removal algorithms. *Phys. D* **60** 259–268.
[31] SHA'ASHUA, A. and ULLMAN, S. (1988). Structural saliency: The detection of globally salient structures using a locally connected network. In *Proc. Second International Conference on Computer Vision* 321–327. IEEE, Washington.
[32] SHAKED, D., YARON, O. and KIRYATI, N. (1996). Deriving stopping rules for the probabilistic Hough transform by sequential analysis. *Computer Vision and Image Understanding* **63** 512–526.
[33] SLUD, E. (1978). Distribution inequalities for the binomial law. *Ann. Probab.* **5** 404–412.
[34] STEWART, C. V. (1995). MINPRAN: A new robust estimator for computer vision. *IEEE Trans. Pattern Anal. Machine Intelligence* **17** 925–938.
[35] WERTHEIMER, M. (1923). Untersuchungen zur Lehre der Gestalt II. *Psychologische Forschung* **4** 301–350.
[36] YUILLE, A. L., COUGHLAN, J. M., WU, Y.-N. and ZHU, S. C. (2001). Order parameters for detecting target curves in images: When does high-level knowledge help? *Internat. J. Comput. Vision* **41** 9–33.

CMLA, ENS CACHAN
61 AVENUE DU PRÉSIDENT WILSON
94235 CACHAN CEDEX
FRANCE
E-MAIL: desolneu@cmla.ens-cachan.fr