Maximal Meaningful Events and Applications to Image Analysis

Agnès Desolneux, Lionel Moisan, Jean-Michel Morel,
CMLA, ENS Cachan,
61 avenue du président Wilson,
94235 Cachan cedex - France

Abstract

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 Helmholtz principle. According to this principle, an observed geometric structure is perceptually "meaningful" if the expectation of its occurences (in other terms, 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 a parameter 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 two cases: alignments in a digital image and histogram modes. We show why these results make maximal meaningful structures computable and display a joint numerical application of both detection theories.

Contents

1 Introduction

In [7], we outlined a parameter free methodology in image processing which raised several mathematical questions which we address here. We shall also expose here a new application of the same methodology to the search of histogram modes. We think it necessary to summarize the arguments in favour of a parameter free methodology in image processing. We hope that this will enhance the interest for the mathematical framework. Most theories of image analysis tend to find in a given image geometric structures (regions, contours, lines, convex sets, junctions, etc.). These theories generally assume that the images contain such structures and then try to compute their best description. The variational framework is quite well adapted to such a viewpoint (for a complete review, see e.g. [25]). The general idea is to minimize a functional of the kind

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

where u_0 is the given image defined on a domain $\Omega \subset \mathbb{R}^2$, $F(u, u_0)$ is a fidelity term and $R(u)$ is a regularity term. F and R define an *a priori* model. Let us give two examples:

• The Mumford-Shah model (see [25], [26], [27], [28]), where the energy functional to be minimized is

$$E(u, K) = \lambda^2 \int_{\Omega - K} |\nabla u|^2 \, dx + \mu\lambda^2 \text{length}(K) + \int_{\Omega - K} (u - u_0)^2 \, dx,$$

where u is the estimated image, K its discontinuity set, and the result (u, K) is called a "segmentation" of u_0, i.e. a piecewise smooth function u with a set of contours K.

• The Bayesian model (see [12]): let us denote by $\vec{y} = (y_s)_{s \in S}$ the observation (the degraded image). The aim is to find the "real" image $\vec{x} = (x_s)_{s \in S}$ knowing that the degradation model is given by a conditional probability $\Pi(\vec{y}|\vec{x})$, and that the *a priori* law of $\vec{x}$ is given by a Gibbs distribution $\Pi(\vec{x}) = Z^{-1} \exp(-U(\vec{x}))$ (for binary images, the main example is the Ising model). We then have to find the M.A.P. (Maximum A Posteriori) of

$$\Pi(\vec{x}|\vec{y}) = \frac{\Pi(\vec{y}|\vec{x})\Pi(\vec{x})}{\Pi(\vec{y})} .$$

Assume that $\Pi(\vec{y}|\vec{x}) = C \exp(-V(\vec{x}, \vec{y}))$. For example, in the case of a Gaussian noise,

$$\Pi(\vec{y}|\vec{x}) = (\frac{1}{2\pi\sigma^2})^{\frac{|S|}{2}} \exp(-\frac{1}{2\sigma^2} \sum_{s \in S} (y_s - x_s)^2),$$

finding the MAP is equivalent to seeking for the minimum of the functional

$$V(\vec{x}, \vec{y}) + U(\vec{x}) .$$

A main drawback of all the variational methods is that they introduce normalization constants (λ, μ, ...) and the resulting segmentation depends a lot upon the value of these constants. Notice that in the second model, U contains several parameters and the resulting functional also depends upon a degradation model. The other point is that variational methods will always deliver a minimum for their functional. Now, they do not yield any criterion to decide whether an obtained segmentation is relevant or not. Of course, the probabilistic framework leading to variational methods should in principle give a way to estimate the parameters of the segmentation functional. In the deterministic framework, these parameters can sometimes be estimated as Lagrange multipliers when (e.g.) a noise model is at hand, like in the Rudin-Osher-Fatemi method (see [31]). It is nonetheless easy to check that, first, most variational methods propose a very rough and incomplete model for real world images, second, that their parameters are generally not correctly estimated anyway, yielding to supervised methods. Another possibility, which turns out to be a significant improvement of MAP methods, is the Minimal Description Length method (MDL) introduced by Rissanen [30] and first applied in image segmentation by Yvon Leclerc [20]. Actually, this last mentionned method, applied to detect regions and their boundaries in an image, permits to fix in an automatic way the weight parameters whose presence we criticized in the Mumford-Shah model. Now, the resulting segmentation model remains all the same unproved: the MDL principle does not prove the existence of regions: it only gives their best description, provided the image indeed is segmentable into constancy regions. This fact is easily explained: the MDL principle assumes that a model, or a class of models, is given and then computes the best choice of the model parameters, and of the model explaining the image. As far as perception theory is concerned, we request more, namely a proof that the model is the right one. Now, once detection of geometric structures in an image has been achieved, the resulting set of detected structures may be very redundant, and we may need the MDL principle as a further step, in order to give the "best explanation" of what has been previously detected. We shall briefly develop this point of view in the experimental section 9. Not all geometric detection method are variational ; let us mention as classical and complementary examples the Hough Transform (see [22]), the detection of globally salient structures by Sha'Ashua and Ullman (see [33]), the Extension Field of Guy and Medioni (see [14]) and the Parent and Zucker curve detector (see [29]). These methods have the same drawback as the variational models of segmentation described above. The main point is that they *a priori* suppose that what they want to find (lines, circles, curves...) is in the image. They may find too many or too little such structures in the image and do not yield an **existence proof** for the found structures. Let us describe the Hough transform. We assume that the image under analysis is made 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 the today technology, this threshold generally is given by a user or learned. The work of Kiryati, Eldar and Bruckstein [19] and of Shaked, Yaron and Kiryati [32] is, however, very close to what we develop here: these authors prove by large deviations estimates that lines in an image detected by 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 shall develop it here as a geometric definition tool. The Hough transform is nothing but a particular kind of "grouping".

According to Gestalt theory, "grouping" is the main process in our visual perception (see [17]). 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 colour constancy, "good continuation", alignment, parallelism, common orientation, convexity and closedness (for a curve), ... In addition, the grouping principle is recursive. For example, if points have been grouped into lines, then these lines may again be grouped according (e.g.) to parallelism.

Helmholtz Principle. In [7], we outlined a computational method to decide whether a given Gestalt (obtained by any segmentation or grouping method) is sure or not.We focussed on alignments, as one of the most basic Gestalt (see [40]). As we shall recall, our method gives *absolute thresholds*, that is, thresholds permitting to decide when a peak in the Hough transform is significant or not.

In this paper, we push the study to the end for the detection of alignments, but we will first give a general definition of what we will call "a meaningful event". Many of our statements apply to other Gestalt as well. In particular, we shall here prove that the mentionned definitions can be adapted to the really important problem of defining modes in a histogram without any *a priori* model. A meaningful event is an event that, according to probabilistic estimates, should not happen in an image and therefore is significant. In that sense, we can say that it is a "proven event". The above informal definition immediately raises an objection: if we do probabilistic estimates in an image, this means that we have an *a priori* model. We are therefore losing any generality in the approach, unless the probabilistic model could be proven to be "the right one" for any image. In fact, we do statistical estimates, but related not to a model of the images but to a general model of perception. We apply the so called Helmholtz principle. This principle attempts to describe when perception decides to group objects according to some quality (colour, alignment, etc.). It can be stated in the following way. Assume that objects O_1, O_2,...,O_n are present in an image. Assume that k of them, say O_1,...,O_k have a common feature, say, same colour, same orientation, etc. We are then facing the dilemna: is this common feature happening by chance or is it significant? In order 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, i.e. O_1, ...O_n. Notice that this quality may be spatial (like position, orientation); then we (mentally) assume that the observed position of objects in the image is a random realization of this uniform process. Then, we may 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, ...,O_k is very small, then the grouping of these object makes sense, is a Gestalt.

Definition 1 (ε-meaningful event) *[7] We say that an event of type "such configuration of points has such property" is ε-meaningful, if the expectation in a image of the number of occurences of this event is less than ε.*

When $\varepsilon \leqslant 1$, we talk about meaningful events. This seems to contradict our notion of a parameter-less theory. Now, it does not, since the ε-dependency of meaningfulness will be low (it will be in fact a $\log \varepsilon$-dependency). The probability that a meaningful event is observed by accident will be very small. In such a case, our perception is liable to see the event, no matter whether it is "true" or not. Our term ε-meaningful is related to the classical p-significance in statistics ; as we shall see further on, we must use expectations in our estimates and not probabilities. We refer to [7] for a complete discussion of this definition.

Let us now address briefly the other detection instance which we shall develop here. the detection of modes in a histogram, that is, meaningful intervals. This example is so much similar to the alignement detection, that we shall be able to accelerate a lot the discussion of meaningfulness and will give a mode detection algorithm. In histogram analysis, we can distinguish several classes of algorithms computing modes. First of all, a parametric model may be at hand, ensuring e.g. that the histogram is the an instance of k gaussian random variables whose average and variance have to be estimated from the histogram ([9], [36], [38]). Clearly, optimization algorithms can be defined for this problem and, if k is unknown, it may be found by using variants of the Minimal Description Length Principle. Then, many theories intend to threshold a histogram in an optimal way, that is, to divide the histogram into two modes according to some criterion. The most popular criterion is entropic (see [37], [1],[18],[4]): the authors try to find a threshold value m such that some entropy term of the bimodal histogram is maximal ; a generalization leads to find by entropic criteria multiple thresholds. This thresholding problem turns out to be very useful and relevant in image analysis, since it leads to the problem of optimal quantization of the grey levels. Here again, we can repeat the same criticism as for segmentation algorithms: the found thresholds are not proved to be relevant, and separating meaningful modes of the histogram. To take an instance, if the histogram is constant, the optimal threshold given by the mentionned methods is the median value. Now, a constant histogram is not bimodal. As in the alignment detection theory, we shall adopt the Helmholtz principle (we give up any a priori knowledge about the histogram model). Thus, we compute as though all samples were uniformly and independently distributed. Meaningful modes will be defined as counterexamples to this uniformity assumption and we define the actual modes as the maximal meaningful modes. We shall give a theorem proving that, in the large deviation framework, maximal meaningful intervals of the histogram are disjoint. We shall immediately apply the resulting algorithm to image analysis. Our goal is to show the reliability of the detection theory to give an account of the so called "visual pyramid", according to which geometric events (Gestalt) are grouped recursively at different scales of complexity. This hypothesis of Gestalt theory [23] shall be valid only if the detection of geometric events is robust enough to allow one to compute modes of properties of these events. We shall first compute all maximal meaningful alignments in several images, and then group them according to the mode of length, orientation they belong to.

Our plan is as follows. In Section 2, we explain our definition of meaningful alignments. Section 3 is devoted to the proof of first structure properties of the "number of false alarms". In Section 4, we prove asymptotic (as $l \to \infty$) and non-asymptotic estimates about the meaningfulness of the following observation : "k well-aligned points in a segment of length l". In Section 5, we point out some properties of meaningful segments. Section 6 introduces maximal meaningfulness as a mean to reduce the number of events and localize them. Section 6 also gives strong arguments in favour of our main conjecture : two maximal meaningful segments on the same line have an empty intersection, and shows that it is true in the large deviation framework. In Section 7, we briefly address the problem of the choice of the precision p. In Section 8, we develop a version of the theory adapted to the computation of modes of a histogram. We again prove that maximal meaningful intervals of a histogram do not meet and show that, in an intrinsic and parameter free way, one can define "modes" for every real valued histogram. In Section 9, we end with joint numerical experiments, identifying maximal alignments in a digital image and grouping them by parallelism.

2 Definition of meaningful segments

2.1 The discrete nature of applied geometry

Although mathematicians and even computer vision scientists sometimes allude to or presuppose the fact that an image has a potentially infinite resolution, it must be recalled here that all images of which we physically dispose are discrete events containing a finite amount of information. 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, e.g. no straight line, no circle, no convex set, etc. 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 (that is a regular grid of N^2 pixels). At each point x, or pixel, of the discrete grid, we have a grey level $u(x)$ which is quantized and therefore inaccurate. We may compute at each point the direction of the gradient, which is the simplest local contrast invariant information (local contrast invariance is a necessary requirement in image analysis and perception theory [40]). We compute a direction, which is the direction of the level line passing by the point calculated on a $q \times q$ pixels neighbourhood (generally $q = 2$). No previous

smoothing on the image will be performed and no restoration: such processes would loose the *a priori* independence of directions which is required for the detection method.
The computation of the gradient direction is based on an interpolation (we have $q = 2$). We define the direction at pixel (i, j) by rotating by $\frac{\pi}{2}$ the direction of the gradient of the order 2 interpolation at the center of the 2×2 window made of pixels (i, j), $(i+1, j)$, $(i, j+1)$ and $(i+1, j+1)$. We get

$$\mathrm{dir}(i,j) = \frac{1}{||\vec{D}||}\vec{D} \quad \text{where} \quad \vec{D} = \frac{1}{2}\begin{pmatrix} -[u(i,j+1) + u(i+1,j+1)] + [u(i,j) + u(i+1,j)] \\ [u(i+1,j) + u(i+1,j+1)] - [u(i,j) + u(i,j+1)] \end{pmatrix}.$$

Then we say that two points X and Y have the same direction with precision $\frac{1}{n}$ if

$$\mathrm{Angle}(\mathrm{dir}(X), \mathrm{dir}(Y)) \leqslant \frac{2\pi}{n}.$$

In agreement with psychopysics 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 $p = \frac{1}{n} < \frac{1}{2}$; 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 structureless image, when two pixels are such that their distance is more than 2, the directions computed at the two considered pixels should be independent random variables. By Helmholtz principle, every deviation from this randomness assumption will lead to the detection of a structure (Gestalt) in the image. Alignments provide a concrete way to understand Helmholtz principle. We know (by experience) that images have contours and therefore meaningful alignments. This is mainly due to the smoothness of contours of solid objects and the generation of geometric structure by most physical and biological laws.

From now on, we do the computation 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 of l independent pixels (it 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 to know what is the minimal number $k(l)$ of aligned points that we must observe on a length l segment 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 x_1, x_2, ... , 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 :

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

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

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

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 length l segment, 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 a $N \times N$ image. We define the total number of oriented segments in a $N \times N$ 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 (detection thresholds) *We call "detection thresholds" a family of positive values $w(l, \varepsilon, N)$, $1 \leqslant l \leqslant l_{max}$, such that*

$$\sum_{l=1}^{l_{max}} w(l, \varepsilon, N) m(l) \leqslant \varepsilon.$$

Definition 3 (ε-meaningful segment - general definition) *A length l segment is ε-meaningful in a $N \times N$ image if it contains at least $k(l)$ points having their direction aligned with the one of the segment, where $k(l)$ is given by*

$$k(l) = \min \{k \in \mathbb{N},\ \mathrm{P}[S_l \geqslant k] \leqslant w(l, \varepsilon, N)\}.$$

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

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

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

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

We compute here the expectation of R but not its law because it depends a lot upon 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

$$\mathrm{P}[S_l \geqslant k(l)] \leqslant w(l, \varepsilon, N), \quad \text{so that} \quad E(R) \leqslant \sum_{l=1}^{l_{max}} w(l, \varepsilon, N) m(l) \leqslant \varepsilon.$$

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 $w(l, \varepsilon, N)$. The difference which leads us to have a slightly different terminology is following: we are not in a position to assume that the segment detected as ε-meaningful are independent in anyway. Indeed, if (e.g.) a segment is meaningful it may be contained in many larger segments, which also are ε-meaningful. Thus, it will be convenient to compare the number of detected segments to the expectation of this number. This is not exactly the same situation as in failure detection, where the failures are somehow disjoint events. See remark (*) below.

The question of how to fix the detection thresholds is widely open. Our definition of ε-meaningful segment will be a restriction of the above general definition. Since there is a priori no reason to favour small or large segments, we choose a uniform family of detection thresholds:

$$\forall\, l \geqslant 1 \quad w(l, \varepsilon, N) = \frac{\varepsilon}{N^4}.$$

Our definition of ε-meaningful segment is then the following one.

Definition 4 (ε-meaningful segment) *A length l segment is ε-meaningful in a $N \times N$ image if it contains at least $k(l)$ points having their direction aligned with the one of the segment, where $k(l)$ is given by*

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

In the following, we write $P(k,l)$ for $\mathrm{P}\left[S_l \geqslant k\right]$.

Remark : We could have defined a ε-meaningful length l segment as a segment ε-meaningful only among the set of the length l segments. It would have been a segment with at least $k'(l)$ points having the "good" direction where $k'(l)$ is defined by $m(l) \cdot \mathrm{P}\left[S_l \geqslant k'(l)\right] \leqslant \varepsilon$. Notice that $m(l) \simeq N^3$ because there are approximately N^2 possible discrete straight lines in a $N \times N$ image and on each discrete line, about N choices for the starting point of the segment. But we did not keep this definition because when looking for alignments we cannot *a priori* know the length of the segment we look for. In the same way, we never consider events like : "a segment has exactly k aligned points", but rather "a segment has at least k aligned points", and k must be given, as we do, by a detectability criterion and not *a priori* fixed.

3 Number of false alarms

3.1 Definition

Definition 5 (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*

$$NF(k_0, l_0) = N^4 \cdot \mathrm{P}\left[S_{l_0} \geqslant k_0\right] = N^4 \cdot \sum_{k=k_0}^{l_0} \binom{l_0}{k} p^k (1-p)^{l_0 - k}.$$

Interpretation of this definition : the number $NF(k_0, l_0)$ 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 the one of the considered segment.

Remark : (*) (relative notion) Let A be a segment and $NF(k_0, l_0)$ its number of false alarms. Then A is ε-meaningful if and only if $NF(k_0, l_0) \leqslant \varepsilon$, but it is worth noticing that we could have compared $NF(k_0, l_0)$ not to ε but to the real number of segments with probability less than the one of A, observed in the image. For example, if we observe 100 segments of probability less than α, and if the expected value R of the number of segments of probability less than α was 10, we are able to say that this 100-segments event could happen with probability less than 1/10, since $10 = E(R) \geqslant 100 \cdot \mathrm{P}\left[R = 100\right]$. Now, each of these 100 segments only is 10-meaningful ! Of course, we cannot deduce in any way that each one of the segment is meaningful.

3.2 Properties of the number of false alarms

Proposition 1 *The number of false alarms $NF(k_0, l_0)$ has the following properties :*

1. *$NF(0, l_0) = N^4$, which proves that the event for a segment to have more than zero aligned points is never meaningful !*
2. *$NF(l_0, l_0) = N^4 \cdot p^{l_0}$, which shows that a segment such that all of its points have the "good" direction is ε-meaningful if its length is larger than $(-4 \ln N + \ln \varepsilon)/\ln p$.*
3. *$NF(k_0 + 1, l_0) < NF(k_0, l_0)$. This can be interpreted by saying that if two segments have the same length l_0, the "more meaningful" is the one which has the more "aligned" points.*

4. *$NF(k_0,l_0) < NF(k_0,l_0+1)$. This property can be illustrated by the following figure of a segment (where a $\bullet$ represents a misaligned point, and a $\rightarrow$ represents an aligned point) :*

$$\rightarrow\rightarrow\ \bullet\ \rightarrow\rightarrow\ \bullet\bullet\ \rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\ \bullet$$

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. *$NF(k_0+1,l_0+1) < NF(k_0,l_0)$. Again, we can illustrate this property :*

$$\rightarrow\rightarrow\ \bullet\ \rightarrow\rightarrow\ \bullet\bullet\ \rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow$$

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 a consequence of the definition and properties of the binomial distribution (see [10]).

If we consider a length l segment (made of l independent pixels), then the expectation of the number of points of the segment having the same direction as the one of the segment is simply the expectation of the random variable S_l, that is

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

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 $l_0 \geqslant 1$, containing at least k_0 points having the same direction as the one of A. If $NF(k_0,l_0) \leqslant p \cdot N^4$, (which is the case when A is meaningful because N is very large and thus, $pN^4 < 1$), then*

$$k_0 \geqslant pl_0 + (1-p).$$

This is a "sanity check" for the model.This proposition will be proved by Lemma 4, where we will extend the discrete function $P(k,l) = \mathrm{P}\left[S_l \geqslant k\right]$ to a continuous domain.

4 Thresholds and asymptotic estimations

In this section, we shall give precise asymptotic and non-asymptotic estimates of the thresholds $k(l)$, which roughly say that

$$k(l) \simeq pl + \sqrt{C \cdot l \cdot \ln \frac{N^4}{\varepsilon}},$$

where $2p(1-p) \leqslant C \leqslant \frac{1}{2}$. Some of these results are illustrated by Figure 1. 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 $k(l)$.

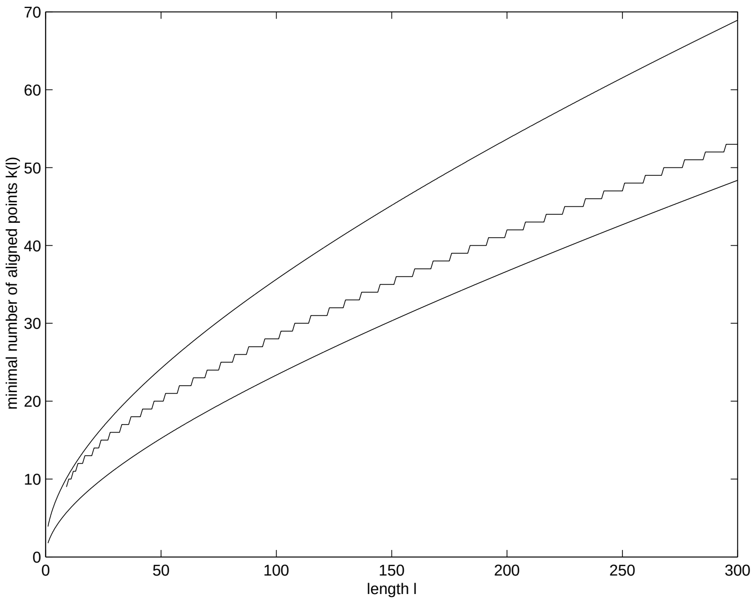


Figure 1: Estimates for the threshold of meaningfulness $k(l)$

The middle (stepcase) 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, for a direction precision of 1/16. The upper and lower curves represent estimates of this threshold obtained by Proposition 4 and Proposition 7.

4.1 Sufficient condition of meaningfulness

In this subsection, we will see how the theory of large deviations and other inequalities concerning the tail of the binomial distribution can provide us a sufficient condition of meaningfulness. The key point is the following result due to Hoeffding (see [15]).

Proposition 3 (Hoeffding's inequality) *Let k, l be positive integers with $k \leqslant l$, and p a real number such that $0 < p < 1$. Then if $r = k/l \geqslant p$, we have the inequalities*

$$P(k,l) \leqslant \exp\left(lr\ln\frac{p}{r} + l(1-r)\ln\frac{1-p}{1-r}\right) \leqslant \exp(-l(r-p)^2 h(p)) \leqslant \exp(-2l(r-p)^2),$$

where h is the function defined on $]0,1[$ by

$$h(p) = \frac{1}{1-2p}\ln\frac{1-p}{p} \qquad \text{for } 0 < p < \frac{1}{2},$$

$$h(p) = \frac{1}{2p(1-p)} \qquad \text{for } \frac{1}{2} \leqslant p < 1.$$

The function h defined above is continuous on $]0,1[$, decreasing on $]0,\frac{1}{2}]$ and increasing on $[\frac{1}{2},1[$. Its minimal value is 2. We plot this function on figure 2

Using the previous proposition, we deduce a sufficient condition for a segment to be meaningful. The size N of the image, and the precision $p < 1/2$ are fixed.

Proposition 4 (sufficient condition of ε-meaningfulness) *Let S be a length l segment, containing at least k aligned points. If*

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

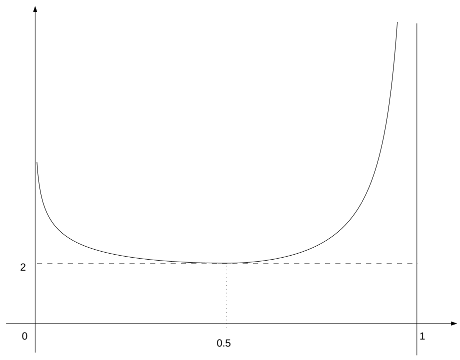


Figure 2: The graph of the function $p \mapsto h(p)$.

then S is ε-meaningful.

Proof : Let S be a length l segment, containing at least k aligned points, where k and l are such that

$$k \geqslant pl + \sqrt{\frac{4 \ln N - \ln \varepsilon}{h(p)}} \sqrt{l}.$$

If we denote $r = k/l$, then $r \geqslant p$ and

$$l(r-p)^2 \geqslant \frac{4 \ln N - \ln \varepsilon}{h(p)}.$$

By Proposition 3 we deduce that

$$P(k,l) \leqslant \exp(-l(r-p)^2 h(p)) \leqslant \exp(-4 \ln N + \ln \varepsilon) = \frac{\varepsilon}{N^4},$$

which means, by definition, that the segment S is ε-meaningful.

□

Corollary 1 *Let S be a length l segment, containing at least k aligned points. If*

$$k \geqslant pl + \sqrt{\frac{l}{2}(4 \ln N - \ln \varepsilon)},$$

then S is ε-meaningful.

Proof : This result is a simple consequence of Proposition 4 and of the fact that for p in $]0,1[$, $h(p) \geqslant 2$ (see Hoeffding [15]).

□

4.2 Necessary conditions for meaningfulness

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

$$p^l \leqslant \mathrm{P}\left[S_l \geqslant k(l)\right] \leqslant \frac{\varepsilon}{N^4},$$

so that

$$l \geqslant \frac{-4 \ln N + \ln \varepsilon}{\ln p}. \tag{1}$$

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 $l_{min} = 9$.

This necessary condition is only on l, so we now look for more precise conditions involving both k and l.

Lemma 1 *Let $0 < r < 1$ be a real number, and g_r the function defined on $]0, 1[$ by*

$$g_r(x) \;=\; r \ln x + (1-r)\ln(1-x) \;,$$

then g_r is concave and has its maximum at point $x = r$. Moreover if $0 < p \leqslant r$ then

$$2(r-p)^2 \;\leqslant\; g_r(r) - g_r(p) \;\leqslant\; \frac{(r-p)^2}{p(1-p)} \;.$$

Lemma 2 *If $N \geqslant 5$ and if $S = (k, l)$ is a ε-meaningful segment with $1 \leqslant k \leqslant l$, then if we denote $r = k/l$,*

$$g_r(r) - g_r(p) \;>\; \frac{3\ln N - \ln\varepsilon}{l}.$$

Proof : Let us assume first that $1 \leqslant k \leqslant l-1$.
Let $S = (k, l)$ be an ε- meaningful segment, then

$$\binom{l}{k} p^k (1-p)^{l-k} \leqslant P(k,l) \leqslant \frac{\varepsilon}{N^4}.$$

If n is an integer larger than 1, by the Stirling's formula refined to (see [10] for example), we have

$$n^n e^{-n}\sqrt{2\pi n} e^{1/(12n+1)} \leqslant n! \leqslant n^n e^{-n}\sqrt{2\pi n} e^{1/12n}.$$

We then deduce that

$$\binom{l}{k} \geqslant \frac{l^l\sqrt{2\pi l}}{k^k\sqrt{2\pi k}(l-k)^{(l-k)}\sqrt{2\pi(l-k)}} e^{\frac{1}{12l+1}-\frac{1}{12k}-\frac{1}{12(l-k)}}.$$

We assumed that $1 \leqslant k \leqslant l-1$ and so we get

$$\frac{1}{12k} + \frac{1}{12(l-k)} - \frac{1}{12l+1} \leqslant \frac{1}{12k} + \frac{1}{12(l-k)} \leqslant \frac{1}{12} + \frac{1}{12} = \frac{1}{6}.$$

On the other hand, we notice that

$$e^{l g_r(p)} = p^k(1-p)^{l-k} \quad \text{ and } \quad e^{l g_r(r)} = \left(\frac{k}{l}\right)^k \left(1-\frac{k}{l}\right)^{l-k}.$$

And we also have $\sqrt{k(l-k)} = l\sqrt{r(1-r)} \leqslant l/2$, and we then obtain

$$\frac{\varepsilon}{N^4} \geqslant P(k,l) \geqslant \binom{l}{k} p^k (1-p)^{l-k} \geqslant \frac{2}{\sqrt{2\pi l}} e^{-1/6} e^{l(g_r(p)-g_r(r))}.$$

And consequently

$$l(g_r(r) - g_r(p)) \geqslant 4\ln N - \ln\varepsilon - \frac{1}{2}\ln l + \ln\frac{2}{\sqrt{2\pi}} - \frac{1}{6}.$$

Since the size of the considered image is $N \times N$ and l is a length of a segment of the image, we have $l \leqslant \sqrt{2}N$. And so

$$l(g_r(r) - g_r(p)) \geqslant \left(4 - \frac{1}{2}\right) \ln N - \ln \varepsilon - \frac{1}{2} \ln \frac{\pi}{\sqrt{2}} - \frac{1}{6}.$$

This inequality permits to conclude since the hypothesis $N \geqslant 5$ ensures that

$$\frac{7}{2} \ln N - \frac{1}{2} \ln \frac{\pi}{\sqrt{2}} - \frac{1}{6} > 3 \ln N.$$

If $k = l$, then $r = 1$ and we simply have $g_1(1) - g_1(p) = -\ln p$. Now, since $S = (k, l)$ is ε-meaningful, we have

$$p^l \leqslant \frac{\varepsilon}{N^4},$$

and therefore

$$g_r(r) - g_r(p) = -\ln p \geqslant \frac{4 \ln N - \ln \varepsilon}{l} \geqslant \frac{3 \ln N - \ln \varepsilon}{l}.$$

□

4.3 Asymptotics for the meaningfulness threshold $k(l)$

In this section, we still consider that ε and p are fixed. We will work on asymptotic estimates of $P(k, l)$ when l is "large". We first recall a version of the central limit theorem in the particular case of the binomial distribution (see [10]).

Proposition 5 (De Moivre-Laplace limit theorem) *If α is a fixed positive number, then as l tends to $+\infty$,*

$$\mathrm{P}\left[S_l \geqslant pl + \alpha\sqrt{l \cdot p(1-p)}\right] \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{+\infty} e^{-x^2/2} dx.$$

Our aim is to get the asymptotic behaviour of the threshold $k(l)$ when l is large. The problem is that if l gets to infinity, we also have to consider that N tends to infinity (because, since l is the length of a segment in a $N \times N$ image, necessarily $l \leqslant \sqrt{2}N$). And so the α used in the De Moivre-Laplace theorem will depend on N. This is the reason why we use the following stronger version of the previous theorem (see [10]).

Proposition 6 (Feller) *If $\alpha(l) \to +\infty$ and $\alpha(l)^6/l \to 0$ as $l \to +\infty$, then*

$$\mathrm{P}\left[S_l \geqslant pl + \alpha(l)\sqrt{l \cdot p(1-p)}\right] \sim \frac{1}{\sqrt{2\pi}} \int_{\alpha(l)}^{+\infty} e^{-x^2/2} dx.$$

Proposition 7 (asymptotic behaviour of $k(l)$) *When $N \to +\infty$ and $l \to +\infty$ in such a way that $l/(\ln N)^3 \to +\infty$, one has*

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

Proof : We define, for $i \in \{0, 1\}$,

$$\alpha_i(l, N) = \frac{k(l) - i - pl}{\sqrt{lp(1-p)}}.$$

Lemmas 1 and 2 imply that

$$\alpha_0(l,N) \geqslant \sqrt{3\ln N},$$

so that $\alpha_i(l,N) \to +\infty$ as $l \to \infty$. Conversely, Corollary 1 implies that

$$k(l) \leqslant pl + \sqrt{\frac{l}{2}(4\ln N - \ln\varepsilon)} + 1,$$

from which we deduce that

$$\frac{\alpha_i^6(l,N)}{l} \leqslant C\frac{(4\ln N - \ln\varepsilon)^3}{l},$$

where C is a constant. Since ε is fixed and $l/(\ln N)^3 \to +\infty$, we get that $\alpha_i^6(l,N)/l \to 0$. Hence, we can apply Feller's Theorem to obtain

$$\forall i \in \{0,1\}, \quad \mathrm{P}\left[S_l \geqslant pl + \alpha_i(l,N)\sqrt{l\cdot p(1-p)}\right] \sim \frac{1}{\sqrt{2\pi}}\int_{\alpha_i(l,N)}^{+\infty} e^{-x^2/2}dx. \tag{2}$$

For $i=0$ (resp. for $i=1$), the left hand term of (2) is smaller (resp. larger) than ε/N^4. Besides, the right hand term is equivalent to

$$\frac{1}{\sqrt{2\pi}\alpha_i(l,N)}e^{-\alpha_i^2(l,N)/2}.$$

For $i=0$, we deduce that

$$\frac{1}{\sqrt{2\pi}}\frac{1}{\alpha_0(l,N)}e^{-\alpha_0^2(l,N)/2}(1+o(1)) \leqslant \frac{\varepsilon}{N^4},$$

which implies

$$O(1) + O(\ln(\alpha_0(l,N))) - \frac{\alpha_0^2(l,N)}{2} + o(1) \leqslant \ln\frac{\varepsilon}{N^4},$$

and finally

$$\alpha_0(l,N)^2 \geqslant 2\ln\frac{N^4}{\varepsilon} + O(\ln\ln N),$$

that is

$$k(l) \geqslant pl + \sqrt{2p(1-p)\cdot l\cdot\left(\ln\frac{N^4}{\varepsilon} + O(\ln\ln N)\right)}. \tag{3}$$

The case $i=1$ gives in a similar way

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

Finally (3) and (4) yield the estimation of $k(l)$ announced in Proposition 7.

□

4.4 Lower bound for the meaningfulness threshold $k(l)$

In this part, we refine the necessary condition of ε-meaningfulness obtained in Section 4.2 by using a comparison between the binomial and the gaussian laws given by the following

Proposition 8 (Slud 1977) *If $0 < p \leqslant 1/4$ and $pl \leqslant k \leqslant l$, then*

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

Proposition 9 (necessary condition of meaningfulness) *We assume that $0 < p \leqslant 1/4$ and N are fixed. If a segment $S = (k, l)$ is ε-meaningful then*

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

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

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

This proposition is a direct consequence of Slud's Theorem. The assumption $0 < p \leqslant 1/4$ is not a strong condition since it is equivalent to consider 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 3 *The map*

$$\tilde{P} : (k, l) \mapsto \frac{\displaystyle\int_0^p x^{k-1}(1-x)^{l-k} dx}{\displaystyle\int_0^1 x^{k-1}(1-x)^{l-k} dx} \tag{5}$$

is continuous on the domain $\{(k, l) \in \mathbb{R}^2, \quad 0 \leqslant k \leqslant l < +\infty\}$, decreasing with respect with k, increasing with respect with 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 $\tilde{P}(0, l) = 1$. Now, we prove that $\tilde{P}(k, l)$ is decreasing with respect with k. For that purpose, we introduce the map

$$A(k, l) = \frac{\displaystyle\int_0^p x^{k-1}(1-x)^{l-k} dx}{\displaystyle\int_p^1 x^{k-1}(1-x)^{l-k} dx}.$$

Since $1/\tilde{P} = 1 + 1/A$, we need to prove that A decreases with respect with k. We compute

$$\frac{1}{A}\frac{\partial A}{\partial k}(k, l) = \frac{\displaystyle\int_0^p x^{k-1}(1-x)^{l-k} \cdot \ln \frac{x}{1-x} dx}{\displaystyle\int_0^p x^{k-1}(1-x)^{l-k} dx} - \frac{\displaystyle\int_p^1 x^{k-1}(1-x)^{l-k} \cdot \ln \frac{x}{1-x} dx}{\displaystyle\int_p^1 x^{k-1}(1-x)^{l-k} dx},$$

and we apply the mean value theorem to obtain the existence of (α, β) such that

$$0 < \alpha < p < \beta < 1 \qquad \text{and} \qquad \frac{1}{A}\frac{\partial A}{\partial k}(k, l) = \ln \frac{\alpha}{1-\alpha} - \ln \frac{\beta}{1-\beta}.$$

The right hand term being negative, the proof is complete. The proof that P increases with respect with l is similar, the increasing map $x \mapsto \ln \frac{x}{1-x}$ being replaced by the decreasing map

$x \mapsto \ln(1-x)$. Finally, the fact that $\tilde{P}(k,l) = P(k,l)$ for integer values of k and l is a consequence of the relation $\tilde{P}(k+1,l+1) = p\tilde{P}(k,l) + (1-p)\tilde{P}(k+1,l)$ (see [10] for example). □

Remark : Properties (2) and (3) guarantee that $\tilde{P}$ is a "good" interpolate of P in the sense that the monotonicity of P in both variables k and l is extended to the continuous domain. Notice that a proof based on the same method (using that $x \mapsto \ln x$ is increasing) will establish that

$$\frac{\partial \tilde{P}}{\partial k} + \frac{\partial \tilde{P}}{\partial l} \leqslant 0,$$

which is the natural extension of the property $P(k+1,l+1) \leqslant P(k,l)$ previously established in Proposition 1.

From now on, we shall assume that $p < 1/2$. The following property is a good example of the interest of the continuous extension of P. This yields a proof of the announced Proposition 2.

Lemma 4 *If $l \geqslant 1$, then $p \leqslant \tilde{P}(p(l-1)+1,l) < \frac{1}{2}$.*

Proof : Using $A(k,l)$ as in Lemma 3 we see that it is sufficient to prove that if $k-1 = p(l-1)$, then

$$\frac{p}{1-p}\int_p^1 x^{k-1}(1-x)^{l-k}dx \leqslant \int_0^p x^{k-1}(1-x)^{l-k}dx < \int_p^1 x^{k-1}(1-x)^{l-k}dx. \tag{6}$$

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, $g'(x)$ writes

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

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

$$\int_0^p f(x)dx = \int_0^p f(p-x)dx < \int_0^p f(p+x)dx = \int_p^{2p} f(x)dx < \int_p^1 f(x)dx$$

and the right hand side of (6) is proved.

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

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

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

$$\int_0^p f(x)dx = \int_0^p f(p-x)dx \geqslant \int_0^p f(p+\frac{1-p}{p}x)dx = \frac{p}{1-p}\int_p^1 f(x)dx,$$

which proves the left hand side of (6). □

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 a $N \times N$ 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, i.e. compare

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

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

$$NF(S_\lambda) < NF(S).$$

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 a $N \times N$ image (with $N \geqslant 6$), then the function defined for $\lambda \geqslant 1$ by*

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

is decreasing.

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

Corollary 2 *Let $A=(k,l)$ and $B=(k',l')$ be two 1-meaningful segments of a $N \times N$ image (with $N \geqslant 6$) such that*

$$\frac{k'}{l'} \geqslant \frac{k}{l} \qquad \text{and} \qquad l' > l.$$

Then, B is more meaningful than A, that is $NF(B) < NF(A)$.

Proof : Indeed, we can take $\lambda = l'/l > 1$, so that $k' \geqslant \lambda k$. We then have, by Theorem 1,

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

and therefore $N^4 \tilde{P}(k',l') < N^4 \tilde{P}(k,l)$, i.e $NF(B) < NF(A)$. □

An interesting application of Corollary 2 is the concatenation of meaningful segments. Let $A=(k,l)$ and $B=(k',l')$ be two meaningful segments lying on the same line. Moreover we assume that A and B are consecutive, so that $A \cup B$ is simply a $(k+k', l+l')$ segment. Then, since

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

we deduce, thanks to the above corollary, that

$$NF(A \cup B) < \max(NF(A), NF(B)).$$

This shows the following corollary.

Corollary 3 *The concatenation of two meaningful segments is more meaningful than the least meaningful of both.*

The next lemma is useful to prove Theorem 1.

Lemma 5 *Define for $p < r \leqslant 1$, $B(r,l) = \tilde{P}(rl, l)$. Then, one has*

$$\frac{1}{B}\frac{\partial \ln B}{\partial l} < \frac{1}{l} - (g_r(r) - g_r(p)),$$

where g_r is the map defined in Lemma 1.

Proof : We first write the Beta integral in terms of the Gamma function (see [2]),

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

Thanks to (5), this yields

$$B(r,l) = \frac{\Gamma(l+1)}{\Gamma(rl)\,\Gamma((1-r)l+1)} \int_0^p x^{rl-1}(1-x)^{(1-r)l}\,dx. \tag{7}$$

We now use the expansion (see [2])

$$\frac{d\ln\Gamma(x)}{dx} \;=\; -\gamma - \frac{1}{x} + \sum_{n=1}^{+\infty}\left(\frac{1}{n} - \frac{1}{x+n}\right), \tag{8}$$

where γ is Euler's constant. Using (7) and (8), we obtain

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

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

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

Then

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

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 $0<r\leqslant 1$ we have $rl+x \leqslant l+x$ and $(1-r)l+x \leqslant l+x$, so that $f(x) \geqslant 0$ and

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

We deduce that for N integer larger than 1,

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

A simple integration gives

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

Finally

$$\sum_{n=1}^{+\infty}(\frac{r}{rl+n}+\frac{1-r}{(1-r)l+n}-\frac{1}{l+n})\leqslant -r\ln r-(1-r)\ln(1-r)\ ,$$

which yields

$$\frac{1}{B}\frac{\partial B}{\partial l}\leqslant\frac{1}{l}-r\ln r-(1-r)\ln(1-r)+r\ln p+(1-r)\ln(1-p)=\frac{1}{l}-g_r(r)+g_r(p).$$

□

Proof of Theorem 1 : Let us define $r=k/l$. Since S is 1-meaningful we have $r>p$ and also, thanks to Lemma 2,

$$g_r(r)-g_r(p)\geqslant\frac{3\ln N}{l}.$$

Let f be the function defined for $\lambda\geqslant 1$ 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 5, we get

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

which is negative thanks to the hypothesis $N\geqslant 6$. □

Remark : For the approximation of $\tilde{P}(k,l)$ given by the Gaussian Law

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

we immediatly have the result that $G(k',l')<G(k,l)$ when $k'/l'\geqslant k/l>p$ 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. $NF(S)<<1$). Then if we add some "spurious" points at the end of the segment we obtain another segment with probability higher than the one of S and having still a number of false alarms less than 1, which means that this new segment is still meaningful (see figure).

$$\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow\ \bullet\ \bullet\ \bullet\bullet$$

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

Definition 6 (Maximal segment) *A segment A is maximal if*

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

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 which are significant from the point of view of the construction of the whole object. For example, if one considers a square, the only significant segments of this square are the four sides, and not large parts of the sides. With our definition, long enough parts of a side may be meaningful segments, but only the whole side itself will be a maximal meaningful segment.

Proposition 10 (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 to characterize or compare meaningful segments. We now study the structure of maximal segments, and give some evidence that two distinct maximal segments on a 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*

$$\min\left(p, \tilde{P}(k, l), \tilde{P}(k + k' + k'', l + l' + l'')\right) < \max\left(\tilde{P}(k + k', l + l'), \tilde{P}(k + k'', l + l'')\right). \tag{9}$$

This conjecture can be deduced from a stronger (but simpler) conjecture : the concavity in a particular domain of the level lines of a natural continuous extension of P involving the incomplete Beta function. Let us state immediately some relevant consequences of Conjecture 1.

Corollary 4 (Union and Intersection) *If A and B are two segments on the same straight line, then, under Conjecture 1,*

$$\min\left(pN^4, NF(A \cap B), NF(A \cup B)\right) < \max\left(NF(A), NF(B)\right).$$

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 $|A \cup B| \leqslant 256$. For $p = 1/16$, we obtained

$$\min_{|A \cup B| \leqslant 256} \frac{\max\left((NF(A), NF(B)\right) - \min\left(pN^4, NF(A \cap B), NF(A \cup B)\right)}{\max\left((NF(A), NF(B)\right) + \min\left(pN^4, NF(A \cap B), NF(A \cup B)\right)} \simeq 0.000754697... > 0,$$

this minimum (independent of N) being obtained for $A = (23, 243)$, $B = (23, 243)$ and $A \cap B = (22, 230)$ (as before, the couple (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 : Suppose that one can find two maximal segments $(k+k', l+l')$ and $(k+k'', l+l'')$ that have a non-empty intersection (k,l) Then, according to Conjecture 1 we have

$$\min\Big(p, P(k,l), P(k+k'+k'', l+l'+l'')\Big) < \max\Big(P(k+k', l+l'), P(k+k'', l+l'')\Big)\,.$$

If the left hand 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 $(k+k'+k'', l+l'+l'')$. □

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 smaller than 256 are disjoint, which is enough for most practical applications.

6.3 A simpler conjecture

In this subsection, we state a simple geometric property entailing Conjecture 1.

Conjecture 2 *The map $(k,l) \mapsto \tilde{P}(k,l)$ defined in Lemma 3 has negative curvature on the domain $D_p = \{(k,l) \in \mathbb{R}^2,\ p(l-1)+1 \leqslant k \leqslant l\}$.*

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, i.e. satisfy

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

Remark : 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 binomail tail (correct for small deviations, that is $k \simeq pl + C\sqrt{l}$) and also for large deviations estimate.

Proposition 11 *The approximation of $P(k,l)$ given by the Gaussian law*

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

has negative curvature on the domain D_p.

Proof : The level lines $G(k,l) = \lambda$ of $G(k,l)$ can be written under the form

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

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

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

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

We shall investigate Conjecture 1 with several large deviations arguments. Cramér's theorem about large deviations (see [6], for example) applied to Bernoulli random variables yields to the following:

Proposition 12 (Cramér) *Let r be a real number such that $1 \geqslant r > p$, then*

$$\lim_{l\to+\infty} \frac{1}{l} \ln \mathrm{P}\left[S_l \geqslant rl\right] = -r \ln \frac{r}{p} - (1-r) \ln \frac{1-r}{1-p} = -g_r(r) + g_r(p).$$

Notice that Proposition 12 gives the asymptotic estimate of $\ln \mathrm{P}\left[S_l \geqslant rl\right]$ but not the asymptotic estimate of $\mathrm{P}\left[S_l \geqslant rl\right]$. Notice also that the limit given by Proposition 12 was the upper bound of $\ln \mathrm{P}\left[S_l \geqslant rl\right]$ given by Hoeffding's inequality (see Proposition 3).

Theorem 3 *The large deviations estimate of $\ln P(k,l)$ (see Proposition 12) given by*

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

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

Proof : The level lines of $H(k,l)$ are defined by

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

We fix λ and we just write $k(l,\lambda) = k(l)$. If we compute the first derivative of the above equation and then simplify we get:

$$k'(l) \ln k(l) - k'(l) \ln(pl) + (1 - k'(l)) \ln(l - k(l)) - (1 - k'(l)) \ln((1-p)l) = 0.$$

Now, again by differentiation, we get

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

It is equivalent to:

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

which shows that $H(k,l)$ has negative curvature on the domain $pl \leqslant k \leqslant l$.

□

6.4 Proof of Conjecture 1 under Conjecture 2

Lemma 6 (under Conjecture 2) *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 $x = 0$.

Proof : Call f this map, 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

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

thanks to Conjecture 2.

□

We now can prove Conjecture 1 under Conjecture 2.

Proof : Because the inequality we want to prove is symmetric in k' and k'', we can suppose that $k''/l'' \geqslant k'/l'$. If $k+k'-1 \leqslant p(l+l'-1)$, then $\tilde{P}(k+k', l+l') > p$ and we have finished. Thus, in the following we assume $k+k'-1 > p(l+l'-1)$. Let us define the map

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

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

$$k + x_0(k'+k'') = k + \frac{l'}{l'+l''}(k'+k'') \geqslant k + \frac{l'}{l'+l''}(k' + \frac{k'l''}{l'}) = k + k',$$

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

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

The set

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

is a connected segment that contains x_0 because

$$k + x_0(k'+k'') - 1 \geqslant k + k' - 1 > p(l+l'-1) = p\big(l + x_0(l'+l'') - 1\big).$$

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 thanks to Lemma 6, we conclude as announced that

$$f(x_0) > \min_{x\in S} f(x) = \min_{x \in \partial S} f(x) \geqslant \min\left(p, f(0), f(1)\right),$$

since if $x \in \partial S \cap]0,1[$, then $f(x) \geqslant p$ thanks to Lemma 4. □

Remark : This proof (and the proof of Lemma 6) only relies on the fact that there exists *some* smooth interpolation of the discrete $P(k,l)$ that has negative curvature on the domain D_p. There are good reasons to think that the $\tilde{P}(k,l)$ approximation satisfies this property, but it could be that another approximation also does, though we did not find any (for example, the piecewise bilinear interpolation of $P(k,l)$ is not appropriate).
On Figure 3, we give the geometric idea underlying the proof of Conjecture 1 under Conjecture 2.

6.5 Partial results about Conjecture 1

In this section, we shall give an asymptotic proof of Conjecture 2. 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 13 and Proposition 14.

Proposition 13 (precise large deviations estimate) *Let*

$$D(rl+1, l+1) = \frac{p(1-p)}{(r-p)\sqrt{2\pi l r(1-r)}} \ \exp\left[-l\left(r \ln\frac{r}{p} + (1-r)\ln\frac{1-r}{1-p}\right)\right]. \tag{10}$$

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

$$\frac{1 - \dfrac{4r}{(r-p)^2 l(1-p)}}{1 + \dfrac{1}{r(1-r)\sqrt{2\pi l r(1-r)}}} \leqslant \frac{\tilde{P}(rl+1, l+1)}{D(rl+1, l+1)} \leqslant \frac{1}{1 - \dfrac{2}{\sqrt{2\pi l r(1-r)}}}. \tag{11}$$

In particular, on 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 $]p,1[$.

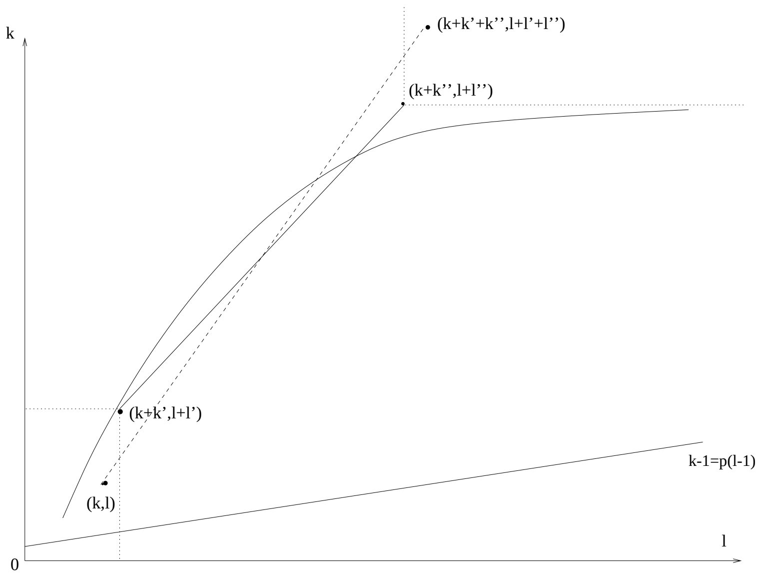


Figure 3: Geometric idea of the proof of Conjecture 1 under Conjecture 2.
We assume that $\tilde{P}(k+k'', l+l'') \leqslant \tilde{P}(k+k', l+l')$. We represent the concave level line of $\tilde{P}$ passing by $(k+k', l+l')$. The point $(k+k'', l+l'')$ is above this level line (indeed, $\frac{\partial \tilde{P}}{\partial k} < 0$). Since the segments $[(k+k', l+l'), (k+k'', l+l'')]$ and $[(k,l), (k+k'+k'', l+l'+l'')]$ have the same middle point, one sees that one of the points (k,l) and $(k+k'+k'', l+l'+l'')$ must lie above the concave level line.

Notice that the exponential term in (10) correspond to Hoeffding's inequality (see Theorem 3).

Proposition 14 *For any $\lambda \in [0,1]$ and $l > 0$, there exists a unique $k(l,\lambda)$ such that*

$$\tilde{P}(k(l,\lambda)+1, l+1) = \lambda. \tag{12}$$

Moreover, one has

$$\frac{\partial^2 k}{\partial l^2}(l, \tilde{P}(rl+1, l+1)) \quad \underset{l \to +\infty}{\sim} \quad -\frac{\left(r \ln \dfrac{r}{p} + (1-r) \ln \dfrac{1-r}{1-p}\right)^2}{l \cdot r(1-r) \cdot \left(\ln \dfrac{r(1-p)}{(1-r)p}\right)^3}. \tag{13}$$

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

We shall not prove these results here: the proof is given in [8] and for more precise results, see [24]. It is interesting to notice that (13) remains true when $k(l,\lambda)$ is defined not from $\tilde{P}$ but from its estimate D given by (10). In the same way, one can prove that

$$\frac{\partial k}{\partial l}(l, \tilde{P}(rl+1, l+1)) \quad \underset{l \to +\infty}{\longrightarrow} \quad \frac{\ln \dfrac{1-p}{1-r}}{\ln \dfrac{r(1-p)}{(1-r)p}}$$

is satisfied by both definitions of $k(l,\lambda)$. This proves that (10) 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 2) *There exists a continous map* $L :]p, 1[\to \mathbb{R}$ *such that* $(k,l) \mapsto \tilde{P}(k,l)$ *has negative curvature on the domain*

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

This result is illustrated on Figure 4.

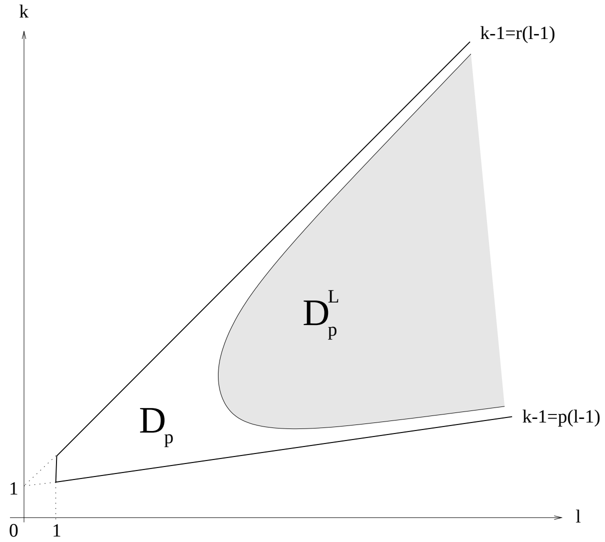


Figure 4: Conjecture 2 is proven on a subdomain D_p^L of D_p.

Proof : Define $k(l, \lambda)$ by (12). Thanks to Proposition 14, the function

$$r \quad \mapsto \quad \frac{\partial^2 k}{\partial l^2}\big(l, \tilde{P}(rl+1, l+1)\big) \; \cdot \; \frac{l \; \cdot \; r(1-r) \; \cdot \; \left(\ln \dfrac{r(1-p)}{(1-r)p}\right)^3}{\left(r \ln \dfrac{r}{p} + (1-r)\ln \dfrac{1-r}{1-p}\right)^2}$$

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

$$r \mapsto l(r) = \inf \Big\{ l_0 > 0, \; \forall l \geqslant l_0, \; \mathrm{curv}\, \tilde{P}(rl+1, l+1) < 0 \Big\}$$

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, on can take

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

where d_n is the unique linear function passing through the points $\big(a_{n-1}, \max_{t \in [a_{n-2}, a_n]} l(t)\big)$ and $\big(a_n, \max_{t \in [a_{n-1}, a_{n+1}]} l(t)\big)$, and $(a_n)_{n \in \mathbb{Z}}$ an increasing sequence such that $\lim_{n \to -\infty} a_n = p$ and $\lim_{n \to +\infty} a_n = 1$. □

7 About the precision p

In this subsection, 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 $p/10$ (for example)?
Knowing that there are k points with direction in $[-p\pi, +p\pi]$, we can assume (by Helmholtz principle) that the average number of points having their direction in $[-\frac{p}{10}\pi, +\frac{p}{10}\pi]$ is $k/10$. The aim now is to compare

$$\mathcal{B}(l, k, p) \qquad \text{and} \qquad \mathcal{B}(l, \frac{k}{10}, \frac{p}{10}),$$

where $\mathcal{B}(l, k, p) = \tilde{P}(k, l)$ for precision p (in the notation $P(k, l)$, we omitted the precision p because it was fixed).

Remark : A non-aligned point for precision p is also non-aligned for precision $p/10$.
Since we are interested in meaningful segments, we will only consider the case

$$\lambda = \frac{k}{l \times p} = \frac{k/10}{l \times p/10} > 1.$$

We then have to study the function $p \longmapsto \mathcal{B}(l, \lambda l p, p)$. Is it increasing, decreasing, ...?
If we consider the large deviations estimate given by

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

we can easily prove that the function $p \longmapsto \lambda p \log \lambda + (1 - \lambda p) \log \frac{1-\lambda p}{1-p}$ is increasing (for $\lambda > 1$). Consequently $p \longmapsto G(l, \lambda l p, p)$ decreases. Thus

$$G(l, k, p) < G(l, \frac{k}{10}, \frac{p}{10}).$$

This inequality has several consequences:

- If the observed alignement at precision $p/10$ 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, because when we observe a meaningful alignment at a very good precision (i.e. p very small), then the best explanation of this alignment is maybe at a larger precision.

Remark : A natural question is: is also $p \longmapsto \mathcal{B}(l, \lambda l p, p)$ decreasing ?

8 Modes of a histogram

When we observe a histogram (for example the histogram of grey-levels of an image), we usually observe "peaks" in the histogram. But peaks are not well-defined: their width and height can vary a lot. We will try here to define the notion of "meaningful peaks".

8.1 Meaningful intervals

We first consider a discrete histogram, that is a finite number M of points and a finite number L of values. This is for example the case of the grey-level histogram of a discrete image. We assume that the set of possible values is $\{1, ..., L\}$. Then for each discrete interval of values $[a, b]$, let $k(a, b)$ be the number of points (among the M) with value in $[a, b]$, and let $p(a, b) = (b - a + 1)/L$. It

represents the prior probability for a point to have value in $[a, b]$.
We are now interested in "peaks", or "modes" of the histogram, that is intervals $[a, b]$ which contain significantly more points than expected.

If we adopt the same definition as for alignments, the number of false alarms of an interval $[a, b]$ is:

$$NF([a,b]) = \frac{L(L+1)}{2}\mathcal{B}(M, k(a,b), p(a,b)),$$

where $\mathcal{B}(n,k,p) = \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j}$ denotes the tail of the binomial distribution of parameters n and p. Notice that the total number of possible intervals is $L(L+1)/2$.

An interval $[a, b]$ is ε-meaningful if $NF([a, b]) \leqslant \varepsilon$, that is

$$\mathcal{B}(M, k(a,b), p(a,b)) < \frac{2\varepsilon}{L(L+1)}.$$

When we consider an interval $[a, b]$, we want to know what minimal number $k_0(a, b)$ of points it has to contain in order to become a meaningful interval.

Notice that in the above definition, compared to definition of meaningful alignment, we use the binomial distribution in differents ways:

- For histograms: $\mathcal{B}(M, k(a,b), \frac{b-a+1}{L})$.
- For alignments: $\mathcal{B}(l, k, p)$.

In the first case M is fixed, the other arguments depend on the considered interval, including the probability $p(a, b) = (b - a + 1)/L$. In the second case, the precision p is fixed and the length l of the segment is a variable first argument of $\mathcal{B}$. Thus, our variables are used in quite different places of $\mathcal{B}$. Now, as we shall see, meaningfulness and maximal meaningfulness will receive a quite analogous treatment.

Proposition 15 *Let $[a, b]$ be a meaningful interval, then*

$$r(a,b) = \frac{k(a,b)}{M} > p(a,b)$$

and by Hoeffding's inequality we have

$$\mathcal{B}(M, k(a,b), p(a,b)) \leqslant e^{-M\left[r(a,b)\log\frac{r(a,b)}{p(a,b)} + (1-r(a,b))\log\frac{1-r(a,b)}{1-p(a,b)}\right]}.$$

Proof : This is a direct application of Proposition 2 and Hoeffding's inequality (Theorem 3). Notice that Proposition 2 provides some inequalities for the binomial distribution when $p \leqslant 1/2$. In order to have inequalities for $p > 1/2$, we use the following property:

$$\mathcal{B}(l, k, p) = \mathcal{B}(l, l-k+1, 1-p).$$

□

We will be interested in experiments on the histogram of grey-levels of a discrete $N \times N$ image. We consider an image of size $N = 256$ and with grey-level values in $\{0, 1, .., 255\}$. We fix $M = 256^2$ and $L = 256$, and we first give a table of detection thresholds. For each length l, such that $1 \leqslant l \leqslant L$, we compute the minimal number $k(l)$ of points (among the $N^2 = M$) that an interval of length l has to contain in order to become 1-meaningful. This means that $k(l)$ is defined as the smallest integer such that

$$\mathcal{B}(M, k(l), \frac{l}{L}) < \frac{2}{L(L+1)}.$$

We also compute the detection thresholds $k_d(l)$ given by the large deviations estimate of the binomial tail. This means that $k_d(l)$ is defined as the smallest integer above $M \times l/L$ such that

$$\frac{k_d(l)}{M}\log\frac{k_d(l)L}{Ml} + (1 - \frac{k_d(l)}{M})\log\frac{1 - k_d(l)/M}{1 - l/L} > \frac{1}{M}\log\frac{L(L+1)}{2}.$$

Thanks to Hoeffding's inequality, we have $k_d(l) \geqslant k(l) > Ml/L$.

On Figure 5, we plot $k(l)$, $k_d(l)$ (dotted curve) and $M \times l/L$ (dashed line) for l in $[1, 10]$. The maximal value of the relative error $l \mapsto (k_d(l) - k(l))/k(l)$ for $l \in [1, 256]$ is about 3%, attained for small values of l.

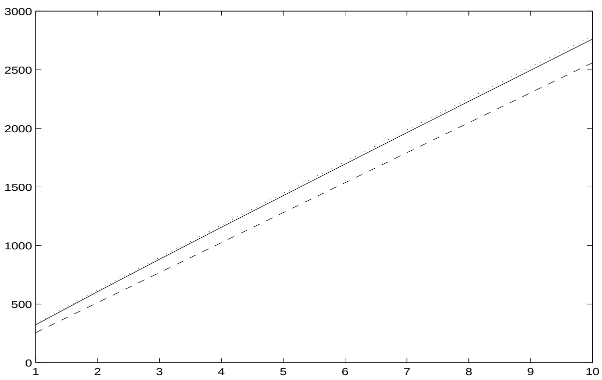


Figure 5: The detection thresholds $k(l)$ and $k_d(l)$ (dotted curve), and Ml/L (dashed line), for $1 \leqslant l \leqslant 10$.

These experiments justify the adoption of the large deviation estimate in order to define meaningful and maximal meaningful intervals [1].

Definition 7 (relative entropy) *We define the relative entropy of an interval* $[a, b]$ *(with respect to the prior distribution) by*

$$H([a,b]) = \begin{cases} 0 & \text{if } r(a,b) \leqslant p(a,b) \\ r(a,b)\log\frac{r(a,b)}{p(a,b)} + (1-r(a,b))\log\frac{1-r(a,b)}{1-p(a,b)} & \text{otherwise.} \end{cases}$$

In the case $r(a, b) > p(a, b)$, the relative entropy $H([a, b])$ is also called the Kullback-Leibler distance between the two Bernoulli distributions of respective parameter $r(a, b)$ and $p(a, b)$ (see [5]).

Remark :

This definition is related to coding and Information Theory (see also [5]). Let us explain in which sense. We consider the histogram of a set of M points distributed on a length L interval (called the reference interval). We fix an interval I of length $l \leqslant L$. Let k be the number of points, among the M, it contains. We want to encode a binarisation of the histogram defined in the following way: for each point we only keep the information of whether it belongs to the fixed interval I or not. Since the prior probability for a point to be in I is l/L, the prior expected bit-length needed to encode the histogram is

$$-k\log_2\frac{l}{L} - (M-k)\log_2(1-\frac{l}{L}).$$

On the other hand, the posterior probability for a point to be in I is k/M. Thus, the posterior expected bit-length needed to encode the histogram is

$$-k\log_2\frac{k}{M} - (M-k)\log_2(1-\frac{k}{M}).$$

This shows that the code gain is

$$-k\log_2\frac{l}{L} - (M-k)\log_2(1-\frac{l}{L}) - (-k\log_2\frac{k}{M} - (M-k)\log_2(1-\frac{k}{M})) = M(r\log_2\frac{r}{p} + (1-r)\log_2\frac{1-r}{1-p}),$$

[1] In pratice, $\mathcal{B}(M, k, p)$ is no more exactly computable for M exceeding 512×512.

where $r = k/M$ and $p = l/L$. Thus, our measure of "meaningfulness" of an interval is directly related to the gain between the prior and the posterior coding of the interval. The higher the gain is, the more meaningful the interval is.

Definition 8 (meaningful interval) *We say that an interval $[a, b]$ is ε-meaningful if its relative entropy $H([a, b])$ is such that*

$$H([a,b]) > \frac{1}{M} \log \frac{L(L+1)}{2\varepsilon}.$$

8.2 Maximal meaningful intervals

For the same reasons we had to introduce the notion of maximal meaningful alignment (see Section 6.1), we have here to define maximal meaningful intervals.

Definition 9 (maximal meaningful interval) *We say that an interval $I = [a, b]$ is maximal meaningful if it is meaningful and if*

$$\forall J \subset I \qquad H(J) \leqslant H(I),$$

$$\text{and} \qquad \forall J \supsetneq I \qquad H(J) < H(I).$$

The question is: can two maximal meaningful intervals have a non-empty intersection? We will see that the answer is no. But notice that we are not in the same case as for alignments, and so we cannot apply the same results. In the case of alignments, the probability p was a fixed number and the variables were the length l of the segment and the number k of aligned points on the considered segment. Now, in the case of histograms, the total number of points is a fixed number N and the variables are the prior probability $p(I)$ of interval I and the number $k(I)$ of points in I.

Theorem 5 *Let I_1 and I_2 be two meaningful intervals such that $I_1 \cap I_2 \neq \emptyset$, then*

$$\max(H(I_1 \cap I_2), H(I_1 \cup I_2)) \geqslant \min(H(I_1), H(I_2)),$$

and the inequality is strict when $I_1 \cap I_2 \neq I_1$ and $I_1 \cap I_2 \neq I_2$.

Proof : For an interval I, we denote by $r(I)$ the proportion of points it contains and $p(I)$ its relative length. Then the entropy of the interval is

$$H(I) = \begin{cases} 0 \quad \text{if} \quad r(I) \leqslant p(I) \\ F(r(I), p(I)) \quad \text{otherwise,} \end{cases}$$

where F is defined on $[0, 1] \times [0, 1]$ by

$$F(r,p) = r \log r + (1-r) \log(1-r) - r \log p - (1-r) \log(1-p).$$

For all $(r, p) \in [0, 1] \times [0, 1]$, $F(r, p)$ is positive and it is 0 if and only if $r = p$. Indeed $F(r, p) = g_r(r) - g_r(p)$, and we know by Lemma 1 that g_r attains its maximum at r.
We first prove that F is a convex function. The partial derivatives of F are:

$$\frac{\partial F}{\partial r} = \log \frac{r}{1-r} - \log \frac{p}{1-p} \qquad \text{and} \qquad \frac{\partial F}{\partial p} = \frac{p-r}{p(1-p)}.$$

$$\frac{\partial^2 F}{\partial r^2} = \frac{1}{r(1-r)}, \quad \frac{\partial^2 F}{\partial r \partial p} = \frac{-1}{p(1-p)} \quad \text{and} \quad \frac{\partial^2 F}{\partial p^2} = \frac{r}{p^2} + \frac{(1-r)}{(1-p)^2}.$$

Then, we get

$$\frac{\partial^2 F}{\partial r^2} > 0 \quad \text{and} \quad \det(D^2 F) = \frac{\partial^2 F}{\partial r^2} \times \frac{\partial^2 F}{\partial p^2} - \left(\frac{\partial^2 F}{\partial r \partial p}\right)^2 = \frac{(r-p)^2}{r(1-r)p^2(1-p)^2} \geqslant 0,$$

which shows that F is convex. Then the continuous function $\overline{H}(r,p)$ defined by $F(r,p)$ if $r \geqslant p$ and 0 otherwise, is also convex (the partial derivatives are continuous).

By hypothesis, we have $I_1 \cap I_2 \neq \emptyset$. We denote $I = I_1 \cap I_2$ and $J = I_1 \cup I_2$. Then

$$\left\{ \begin{array}{lll} r(I) + r(J) & = & r(I_1) + r(I_2) \\ p(I) + p(J) & = & p(I_1) + p(I_2) \end{array} \right. \tag{14}$$

and

$$\left\{ \begin{array}{ll} r(I) & \leqslant \min(r(I_1), r(I_2)) \leqslant \max(r(I_1), r(I_2)) \leqslant \quad r(J) \\ p(I) & \leqslant \min(p(I_1), p(I_2)) \leqslant \max(p(I_1), p(I_2)) \leqslant \quad p(J) \end{array} \right. \tag{15}$$

Now, we want to show that

$$\min(H(I_1), H(I_2)) \leqslant \max(H(I), H(J)),$$

and that the inequality is strict when $I_1 \cap I_2 \neq I_1$ and $I_1 \cap I_2 \neq I_2$.

In the plane $\mathbb{R}^2$, we consider the set R of points (r,p), such that $r(I) \leqslant r \leqslant r(J)$ and $p(I) \leqslant p \leqslant p(J)$. Then R is a rectangle and, by (15), it contains the points $X_1 = (r(I_1), p(I_1))$ and $X_2 = (r(I_2), p(I_2))$. Let A be the following set of points:

$$A = \{(r,p)/H(r,p) \leqslant max(H(I), H(J))\}.$$

A is a convex set because H is a convex function. Let $X = (r(I), p(I))$ and $Y = (r(J), p(J))$, then A contains the segment $[X, Y]$. Since $\frac{\partial F}{\partial r} \geqslant 0$ for $r \geqslant p$, the set A contains $R \cap \{r \geqslant p\} \cap \mathcal{P}_+$ where $\mathcal{P}_+$ is the half-plane above the line (X, Y) (see Figure 6).

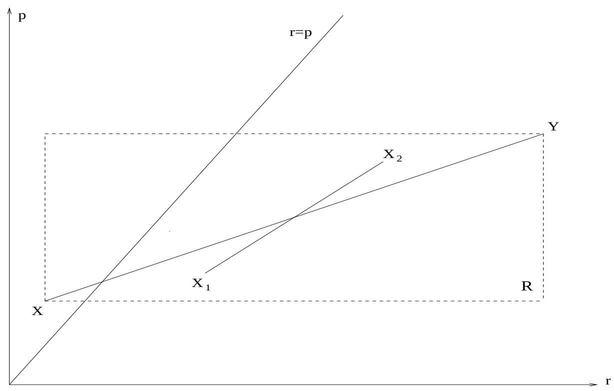


Figure 6:

Since I_1 and I_2 are meaningful, we get X_1 and X_2 in $R \cap \{r > p\}$. And then since the middle point of segment $[X_1, X_2]$ is also the middle point of segment $[X, Y]$ by (14), one of X_1 and X_2 is in $\mathcal{P}_+$. Consequently, X_1 or X_2 is in A, which shows that $\min(H(I_1), H(I_2)) \leqslant \max(H(I), H(J))$. If $I \neq I_1$ and $I \neq I_2$ then the inequality is strict, thanks to the fact that for $r > p$, $\frac{\partial F}{\partial r} > 0$, and we have the announced result.

□

Proposition 16 *Let I_1 and I_2 be two different maximal meaningful intervals, then*

$$I_1 \cap I_2 = \emptyset.$$

Proof : Assume that $I = I_1 \cap I_2 \neq \emptyset$.
If $I \neq I_1$ and $I \neq I_2$, then by Theorem 5, we have

$$\max(H(I_1 \cap I_2), H(I_1 \cup I_2)) > \min(H(I_1), H(I_2)),$$

which is a contradiction with the fact that I_1 and I_2 are maximal meaningful.
If for example $I = I_1 \cap I_2 = I_1$, then $I_1 \subset I_2$. Since by hypothesis I_1 and I_2 are maximal meaningful, we get by definition of maximality $H(I_1) \leqslant H(I_2)$ and $H(I_2) < H(I_1)$, which is again a contradiction.

□

Corollary 5 *Let I and J be two meaningful intervals such that*

$$H(I) = H(J) = \max_{K \subset [0,L]} H(K).$$

Then, either $I \subset J$ or $J \subset I$, or $I \cap J = \emptyset$.

Proof : By Theorem 5, if $I \cap J \neq \emptyset$ and if $I \subset J$ and $J \subset I$, we deduce that $H(I \cap J)$ or $H(I \cup J)$ exceeds $H(I) = H(J)$, which is a contradiction. □

8.3 Meaningful gaps and modes

In the previous section, we were interested in meaningful intervals, i.e. intervals which contain "more points" than the expected average in the sense that

$$\mathcal{B}(M, k(a,b), p(a,b)) < \frac{2}{L(L+1)}.$$

We are now interested in "gaps", i.e. intervals which contain "less points" than the expected average. Let us define this more precisely. Let $[a,b]$ be an interval with prior probability $p(a,b) = (b-a+1)/L$. Let k be an integer such that $0 \leqslant k \leqslant M$. Then the probability that the interval $[a,b]$ contains less than k points (among the total number M of points) is

$$\sum_{j=0}^{k} \binom{M}{j} p(a,b)^j (1-p(a,b))^{M-j} = \mathcal{B}(M, M-k, 1-p(a,b)) = 1 - \mathcal{B}(M, k+1, p(a,b)).$$

An interval $[a,b]$ containing $k(a,b)$ points is a meaningful gap if

$$\mathcal{B}(M, M-k(a,b), 1-p(a,b)) < \frac{2}{L(L+1)}.$$

Proposition 17 *An interval cannot be in the same time a meaningful interval and and a meaningful gap.*

Proof : Let $[a,b]$ be a meaningful gap, then thanks to Proposition 2, we have

$$M - k(a,b) > M \times (1 - p(a,b)),$$

i.e. $r(a,b) = k(a,b)/M < p(a,b)$. This shows that $[a,b]$ cannot be a meaningful interval. □

From now on, and by the same arguments as in subsection 8.1.1, we adopt the large deviation estimate.

Definition 10 (meaningful gap) *We say that an interval $[a,b]$ containing $k(a,b)$ points is a meaningful gap if and only if $r(a,b) = k(a,b)/M < p(a,b)$ and*

$$r(a,b) \log \frac{r(a,b)}{p(a,b)} + (1 - r(a,b)) \log \frac{1-r(a,b)}{1-p(a,b)} > \frac{1}{M} \log \frac{L(L+1)}{2}.$$

Definition 11 (meaningful mode) *We say that an interval is a meaningful mode if it is a meaningful interval and if it does not contain any meaningful gap.*

Definition 12 (maximal meaningful mode) *We say that an interval I is a maximal meaningful mode if it is a meaningful mode and if for all meaningful modes $J \subset I$, $H(J) \leqslant H(I)$ and for all meaningful modes $J \supsetneq I$, $H(J) < H(I)$.*

On Figure 8, we present some experimental results. Subfigure (a) is the original histogram. We have $L = 60$ and $M = 920$. We first compute maximal meaningful intervals (subfigure (b)). We find only one: the interval $[10, 22]$. The second "peak" $[40, 50]$ is not maximal meaningful because when we compute the number of false alarms, we find that

$$NF([10, 22]) < NF([10, 50]) < NF([40, 50]).$$

Next, we compute maximal meaningful modes (subfigure (c)) and we find the two modes $[10, 22]$ and $[40, 50]$.

8.4 Some properties

8.4.1 Mean value of an interval

Our aim here is to compare the relative entropy of two intervals which have the same mean value. The mean value of an interval $[a, b]$ is defined by $r(a, b)/p(a, b)$.

We are only interested in meaningful intervals, this means that we will consider intervals with mean value larger than 1.

Proposition 18 *Let I and J be two intervals with same mean value:*

$$\lambda = \frac{r(I)}{p(I)} = \frac{r(J)}{p(J)} > 1.$$

If $p(I) > p(J)$, then

$$H(I) > H(J),$$

which means that when the average is fixed, the more meaningful interval is the longer one.

Proof : Let $\lambda > 1$ be fixed. For p in $]0, 1[$ such that $r = \lambda p \leqslant 1$, we consider the function

$$g(p) = F(\lambda p, p) = \lambda p \log \lambda + (1 - \lambda p) \log \frac{1 - \lambda p}{1 - p}.$$

We want to show that g is increasing. We have

$$g'(p) = \lambda \left[\log \lambda - \log \frac{1 - \lambda p}{1 - p}\right] + \frac{1 - \lambda p}{1 - p} - \lambda = \lambda[\log \lambda - \log \alpha] - (\lambda - \alpha),$$

where $\alpha = (1 - \lambda p)/(1 - p)$. We have $\lambda > 1 > \alpha$, and there exists $c \in]\alpha, \lambda[$ such that $\log \lambda - \log \alpha = \frac{1}{c}(\lambda - \alpha)$, and then

$$g'(p) = \frac{\lambda}{c}(\lambda - \alpha) - (\lambda - \alpha) > 0.$$

□

The previous proposition has the following corollary which is a result about the concatenation of meaningful intervals.

Corollary 6 *Let $[a, b]$ and $[b + 1, c]$ be two consecutive intervals, then*

$$H(a, c) \geqslant \min[H(a, b), H(b + 1, c)],$$

which means that the interval $[a, c]$ is more meaningful than $[a, b]$ or $[b + 1, c]$.

Proof : Since $r(a,c) = r(a,b) + r(b+1,c)$ and $p(a,c) = p(a,b) + p(b+1,c)$, we get

$$\frac{r(a,c)}{p(a,c)} \geqslant \min \left[\frac{r(a,b)}{p(a,b)}, \frac{r(b+1,c)}{p(b+1,c)} \right],$$

and then the result is a direct consequence of the previous proposition. □

One possible application of this corollary is the fact that maximal meaningful intervals cannot be consecutive.

8.4.2 Structure of maximal meaningful intervals

Theorem 6 *Let h be a histogram defined on a finite set of values $\{1, ..., L\}$. If $[a,b]$ is a maximal meaningful interval such that $1 < a < b < L$, then*

$$h(a-1) < h(a) \quad \text{and} \quad h(b+1) < h(b),$$

$$h(a) > h(b+1) \quad \text{and} \quad h(b) > h(a-1).$$

On Figure 7, we show the structure of a maximal meaningful interval.

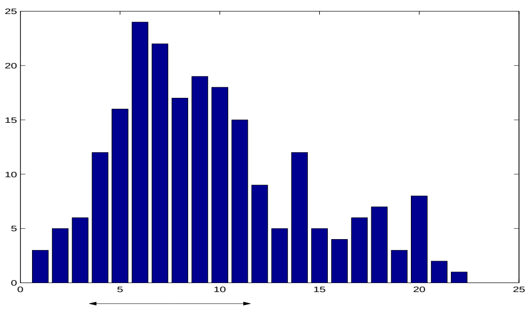


Figure 7: Maximal meaningful interval of a discrete histogram.

Proof : Let $M = \sum_{i=1}^{L} h(i)$ be the "total weight". For an interval $[i,j]$ we have

$$p(i,j) = \frac{j-i+1}{L} \quad \text{and} \quad r(i,j) = \frac{\sum_{x=i}^{j} h(x)}{M}.$$

The relative entropy $H([i,j]) = \overline{H}(r(i,j), p(i,j))$ of the interval $[i,j]$ is 0 if $r(i,j) < p(i,j)$ and $r(i,j)\log r(i,j) + (1-r(i,j))\log(1-r(i,j)) - r(i,j)\log p(i,j) + (1-r(i,j))\log(1-p(i,j))$ otherwise. We will use the fact that the function $(r,p) \mapsto \overline{H}(r,p)$ is convex (see the proof of theorem 5) and that $\frac{\partial \overline{H}}{\partial r} \geqslant 0$ for $r \geqslant p$.

Let $[a,b]$ be a maximal meaningful interval, we will prove that $h(a-1) < h(a)$ (the proof is exactly the same for the other inequalities).

Assume that $h(a-1) \geqslant h(a)$.

Since $[a,b]$ is meaningful, we have $r(a,b) > p(a,b)$. Using the strict convexity of $\overline{H}(r,p)$ for $r > p$, we have

$$H(a,b) < \max \left(\overline{H}(r(a,b) - \frac{h(a)}{M}, p(a,b) - \frac{1}{L}), \overline{H}(r(a,b) + \frac{h(a)}{M}, p(a,b) + \frac{1}{L}) \right).$$

Since $[a,b]$ is maximal, we have

$$H(a+1,b) = \overline{H}(r(a,b) - \frac{h(a)}{M}, p(a,b) - \frac{1}{L}) \leqslant H(a,b).$$

Thus,

$$H(a,b) < \overline{H}(r(a,b) + \frac{h(a)}{M}, p(a,b) + \frac{1}{L}).$$

This shows that $r(a,b) + h(a)/M > p(a,b) + 1/L$. Using the fact that $\frac{\partial \overline{H}}{\partial r} \geqslant 0$ for $r \geqslant p$, we get

$$H(a-1,b) = \overline{H}(r(a,b) + \frac{h(a-1)}{M}, p(a,b) + \frac{1}{L}) \geqslant \overline{H}(r(a,b) + \frac{h(a)}{M}, p(a,b) + \frac{1}{L}).$$

Thus,

$$H(a-1,b) > H(a,b),$$

which is a contradiction with the maximality of $[a,b]$.

□

8.4.3 The reference interval

We address here the problem of the choice of the reference interval. Assume for example that we observe the histogram of grey-levels of an image, knowing a priori that grey-levels have value in $[0,255]$. Now, suppose that the resulting histogram has for example support in $[50,100]$. If we want to detect meaningful and maximal meaningful intervals and modes, which reference interval shall we consider? Shall we work on $[0,255]$ or on the support of the histogram? In order to answer this question, we first have to know what happens when the length of the reference interval becomes very large compared to the fixed length of the support of the histogram.

Let h be a discrete histogram a priori defined on a finite set of values $\{1,...,L\}$. We assume that the support of the histogram is $[1,n]$, i.e. $h(1) > 0$, $h(n) > 0$ and $h(x) = 0$ for $x > n$. For a discrete interval $[a,b] \subset [1,n]$, we will denote $H_L([a,b])$ its relative entropy when the reference interval is $[1,L]$ and $H([a,b])$ its relative entropy when the reference interval is $[1,n]$, i.e. the support of the histogram.

Proposition 19 *Let h be a discrete histogram with support $[1,n]$. Let L be the length of the reference interval. Then there exists L_0 such that*

$$\forall L \geqslant L_0, \quad \forall [a,b] \neq [1,n] \qquad H_L([a,b]) < H_L([1,n]).$$

This means that when the length of the reference interval is large enough, then the support of a discrete histogram is maximal meaningful (and it is the only one).

Proof : For a discrete interval $[a,b] \subset [1,n]$, we denote $p(a,b)$ its relative length and $r(a,b)$ its relative weight when the reference interval is the support $[1,n]$. We also denote $p_L(a,b)$ its relative length and $_r(a,b)$ its relative weight when the reference interval is the support $[1,L]$. We then have

$$p_L(a,b) = \frac{n}{L}p(a,b) \qquad \text{and} \qquad r_L(a,b) = r(a,b).$$

Thus,

$$H_L([a,b]) = H([a,b]) + r(a,b)\log\frac{L}{n} + (1-r(a,b))\log\frac{1-p(a,b)}{1-np(a,b)/L}. \tag{16}$$

In particular, we have $H_L([1,n]) = \log(L/n)$ and the last term of (16) being negative (because $L \leqslant n$), we get

$$H_L([a,b]) \leqslant H([a,b]) + r(a,b)\log\frac{L}{n}.$$

If $[a,b] \neq [1,n]$, then $1 - r(a,b) > 0$. Consequently, there exists a constant C such that

$$\forall [a,b] \neq [1,n] \qquad \frac{H([a,b])}{1-r(a,b)} < C.$$

It shows that for all L such that $\log(L/n) > C$, then $H_L([a,b]) < H_L([1,n])$ for all $[a,b] \neq [1,n]$.

□

9 Applications and experimental results

In this subsection, we will present some joint applications of meaningful alignments in an image and of modes of a histogram.
In all the following experiments, the direction at a pixel in an image is computed on a 2×2 neighborhood with the method described in section 2.1 ($q = 2$) and the precision 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 $\pi/200$. And then on each line, we compute the meaningful segments. For each segment, let l be its length counted in independant 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 the precision p), and finally we compute $P(k,l)$: if it is less than $\frac{\varepsilon}{N^4}$, we say that the segment is ε-meaningful. Notice that $P(k,l)$ can be simply tabulated at the begining 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 (e.g.) two points can be considered as aligned with, say the 0 direction if their angles with this direction are up to ± 22.5 degrees ! It is clear that these bounds are very rough, but in agreement with the more pessimistic estimates for the vision accuracy in psychophysics and the 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 the one of the segment.

For each image, in a first step, we find the maximal meaningful alignments of the image. We obtain a finite set of segments. Each one of these segments has an orientation (valued in $[0, 2\pi[$ because segments are oriented). The precision of the direction of the segment is related to its length: if l denotes the length of the segment, the precision of its direction is $1/l$.
The second step is to get the discrete histogram of the orientations of the detected alignments. The interval $[0, 2\pi[$ is decomposed into $n = 2\pi l_{min}$ bins, where l_{min} is the minimal length of the detected segments. Thus, the size of a bin is $1/l_{min}$.
The third step is to look for maximal meaningful modes of the histogram of orientations. Notice that the framework is a little different. Let us explain this: a histogram of orientations is defined on the "circular" interval $[0, 2\pi[$. Thus, when we look for meaningful intervals $[a, b]$, we do not only consider intervals with $0 \leqslant a \leqslant b < 2\pi$, but also intervals such that $0 \leqslant b \leqslant a < 2\pi$. We define an interval $[a, b]$ such that $0 \leqslant b \leqslant a < 2\pi$ as the union $[a, 2\pi[\cup [0, b]$.

Image 1 : Pencil strokes (see Figure 9). 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 (image (a)); the scanner's apparent blurring kernel is about two pixels wide and some aliasing is perceptible, making the lines somewhat blurry and dashed. Two pairs of pencil strokes are aligned on purpose. We display in the first experiment all ε-meaningful segments for $\varepsilon = 10^{-3}$ (image (b)). 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 this above ; clearly, the pencil strokes boundaries are very meaningful, thus generating larger meaningful segments which contain them.

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 blurring effect of the scanner's optical convolution. 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 along a meaningful segment be distributed on both sides of the lines 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 alignements. 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 no way an elementary phenomenon in Gestalt.

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

On Figure 10 we first present the histogram of the length of the obtained maximal meaningful segments (Figure (a)). We compute the maximal meaningful modes of this histogram and we find the interval $[22, 51]$. On Figure (b), we present the histogram of the orientation modulo π of the obtained maximal meaningful segments. We measure the orientation in degrees: the interval $[-90, 90]$ degrees is divided into $2\pi l_{min} = 112$ bins. We then compute maximal meaningful modes of this histogram. We find five intervals: $[85, 93]$, $[-44, -39]$, $[-5, 18]$, $[28, 31]$, $[39, 45]$. Finally, on figure 11, for each one of the five maximal meaningful modes of the histogram, we show the segments which have their orientation in the mode.

Image 2: Uccello's painting (see Figure 12). This image (a) is a result of the scan of an 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. In Figure (b), we display all maximal ε-meaningful segments with $\varepsilon = 10^{-6}$. Notice how maximal segments are detected on the staircase in spite of the occlusion by the going up child. All remarks made in Image 1 apply here (parallelisms due to the blur, etc...). On the last figure (c), we compute the histogram of the orientations modulo 2π of the obtained maximal meaningful segments. We measure the orientation in degrees, and the interval $[-180, 180]$ degrees is divided into $2\pi l_{min} = 138$ bins. We compute the maximal meaningful modes of the histogram and we find five intervals: $[175, -175]$, $[-92, -87]$, $[-4, 9]$, $[87, 95]$ and $[156, 162]$. The interval mode $[156, 162]$ corresponds to the left side of the roof of the temple. The four others modes correspond to the oriented vertical and horizontal lines. For each mode, we show on Figure 13 the segments which have their orientation in the mode.

Image 3: Building in Cachan (Figure 14 (a)). The size of this image is 901×701. On Figure (b), we display all maximal ε-meaningful segments for $\varepsilon = 10^{-6}$. Notice that we find a lot of diagonal alignments. The explanation of this phenomenon is the fact that when we have many long and parallel alignments (for example at the top of the building), we also detect slanted (with angle less than the precision p) alignments. In Figure (c) we only display a minimal length description of the same segments. This means that, once detected, the alignements must be given their best explanation. One point of the image may belong to many maximal meaningful alignments. We say that a point x is maximal for a segment S if x belongs to S, the direction at point x is aligned (up to precision p) with the direction of the segment S and if S is the most meaningful (smallest number of False Alarms) segment containing x and aligned with the direction at x. Finally, on Figure (c), we only display the maximal meaningful segments of (b) having the property that they are still meaningful when we only count as aligned the number of maximal points they contain.

Aknowledgements:
Work supported by Office of Naval Research under grant N00014-97-1-0839. We thank Richard Lau for personal interest and encouragement.
We thank Jean Bretagnolle, Nicolas Vayatis, Frédéric Guichard, Isabelle Gaudron-Trouvé and Guillermo Sapiro for valuable suggestions and comments.

References

[1] A.S. Abutaled, "Automatic thresholding of gray-level pictures using two-dimensional entropy", *Computer Vision, Graphics and Image Processing*, 47, 22-32 (1989).

[2] J. Anastassiadis, *Définition des fonctions eulériennes par des équations fonctionnelles*, ed. Gauthier-Villars 1965.

[3] H. Bateman, *Higher Transcendental Functions*, vol 1.

[4] C.-I. Chang, K. Chen, J.Wang and M. Althouse , "A relative entropy-based approach to image thresholding", *Pattern Recognition*, vol.27, No.9, pp 1275-1289, 1994.

[5] T.M.Cover and J.A.Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, 1991

[6] A.Dembo and O Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett Publishers 1993.

[7] A. Desolneux,L. Moisan and J.-M. Morel, "Meaningful Alignments", Proceedings of SCTV'99, electronic publication at http://www.cis.ohio-state.edu/szhu/SCTV99.html, to appear in *IJCV*

[8] A. Desolneux,L. Moisan and J.-M. Morel, "Meaningful Alignments", preprint of CMLA, http://www.cmla.ens-cachan.fr/Cmla/Publications/1999/

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

[10] W. Feller, *An introduction to probability theory and its applications*, vol. I, 3rd edition, Wiley 1968.

[11] M. Fréchet, *Les probabilités associées à un système d'événements compatibles et dépendants*, 1ère et 2ème parties, Actualités Scientifiques et Industrielles, Hermann 1940.

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

[13] A. Grigis and J. Sjostrand, "Microlocal analysis for Differential Operators : An introduction", *London Mathematical Society, Lecture Note Series 196*, Cambridge University Press, 1994.

[14] G. Guy and G. Medioni, "Inferring global perceptual contours from local features", *IEEE Trans. Pattern Anal. Machine Intell.*, 1992.

[15] W. Hoeffding, "Probability inequalities for sum of bounded random variables", *Journal of the American Statistical Association*, No.58, p 13-30.

[16] L. Hörmander, *The Analysis of Linear Partial Differential Operators I*, Grundlehren, Springer, 256 (1983).

[17] G. Kanizsa, *La Grammaire du Voir*, Editions Diderot, arts et sciences.

[18] J.N. Kapur, P.K. Sahoo and A.K.C Wong, "A new method for gray-level picture thresholding using the entropy of the histogram", *Computer Vision, Graphics and Image Processing*, 29, 273-285 (1985).

[19] N.Kiryati, Y.Eldar and A.M.Bruckstein, "A Probabilistic Hough Transform", *Pattern Recognition*, vol.24, No.4, pp 303-316, 1991.

[20] Y.Leclerc, "Constructing Simple Stable Descriptions for Image Partitioning", *International Journal of Computer Vision* , 3, 73-102 ,1989 .

[21] D. Lowe, *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers.

[22] H. Maitre, "Un panorama de la transformation de Hough", *Traitement du signal*, vol.2 No.4, 1985.

[23] Metzger, *Gesetze des Sehens*, Waldemar Kramer, 1975.

[24] L.Moisan, "Asymptotic Estimates and Inequalities for the Tail of the Binomial Distribution", in preparation.

[25] J.-M. Morel and S. Solimini, *Variational Methods In Image Segmentation*, Birkhäuser.

[26] D.Mumford and J.Shah, "Boundary detection by minimizing functionals", *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, 1985.

[27] D.Mumford and J.Shah, "Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems", *Communications on Pure and Applied Mathematics*, vol.XLII, No.4, 1989.

[28] N.Nitzberg, D.Mumford and T.Shiota, "Filtering, Segmentation and Depth", *Lecture Notes in Computer Science 662*, Springer-Verlag, 1993.

[29] P. Parent and S.W. Zucker, "Trace inference, curvature consistency and curve detection", *IEEE Trans. Pattern Anal. Machine Intell.* vol.2, No.8, 1989.

[30] J.Rissanen "A universal prior for integers and estimation by Minimum Description Length" , *Annals of Statistics*, 11 (2), 1983 .

[31] L. Rudin, S. Osher and E. Fatemi, "Nonlinear total variation based noise removal algorithms", *Physica D 60*, No.1-4, pp. 259-268, 1992.

[32] D.Shaked, O.Yaron and N.Kiryati, "Deriving Stopping Rules for the Probabilistic Hough Transform by Sequential Analysis", *Computer Vision and Image Understanding*, vol.63, No.3, pp. 512-526, 1996.

[33] A. Sha'Ashua and S. Ullman, "Structural saliency : the detection of globally salient structures using a locally connected network", *Proceedings of the 2d Int. Conf. on Computer Vision*, pp. 321-327, 1988.

[34] E. Slud, "Distribution inequalities for the binomial law", *Annals of Probability*, vol. 5, pp. 404-412.

[35] J. Stojanov, I. Mirazchiiski, Z. Ignatov and M. Tanushev, *Exercise Manual in Probability Theory*, Kluwer Academic 1989.

[36] D.-M. Tsai and Y.-H. Chen, "A fast histogram-clustering approach for multi-level thresholding", *Pattern Recognition Letters*, 13 (1992) , 245-252.

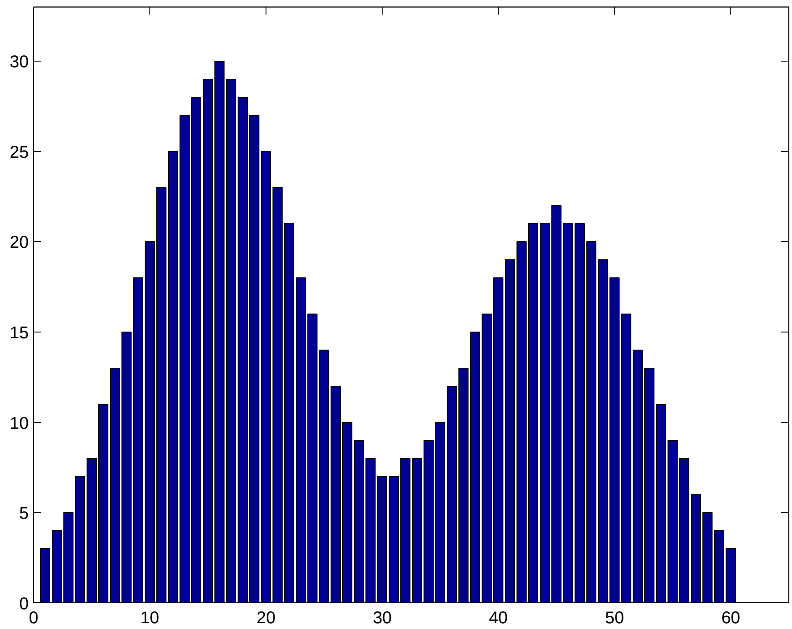
[37] T.Pun, "Entropic thresholding, a new approach", *Computer Graphics and Image Processing*, 16, 210-239 (1981).

[38] J. Puzicha, T. Hofmann and J.M. Buhmann, "Histogram Clustering for Unsupervised Image Segmentation", *CVPR99* , II, 602-608 ,1999.

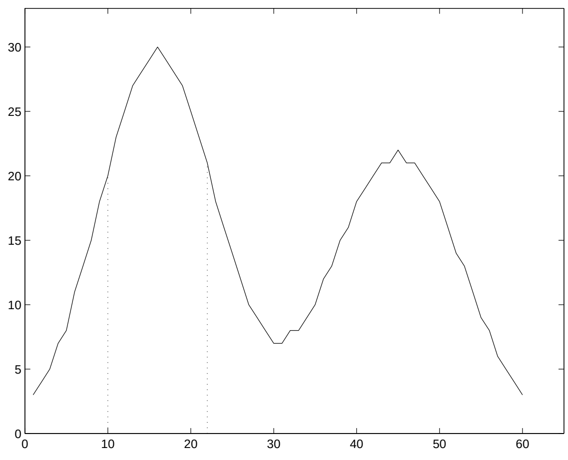
[39] R.Struble, *Nonlinear Differential Equations*, McGraw-Hill, 1962.

[40] M. Wertheimer, "Untersuchungen zur Lehre der Gestalt", II, *Psychologische Forschung*, vol. 4, pp. 301-350, 1923.

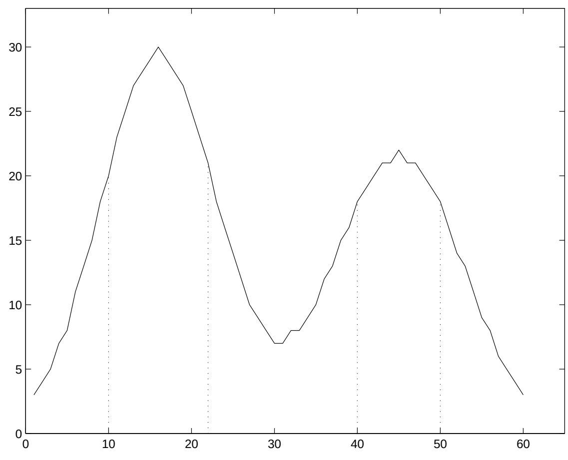
[41] A.P.Witkin and J.P.Tenenbaum, "On the role of structure in vision", Beck, Hope and Rosenfeld, Eds. (New York: Academic Press, 1983), pp. 481-543.



(a) The original histogram

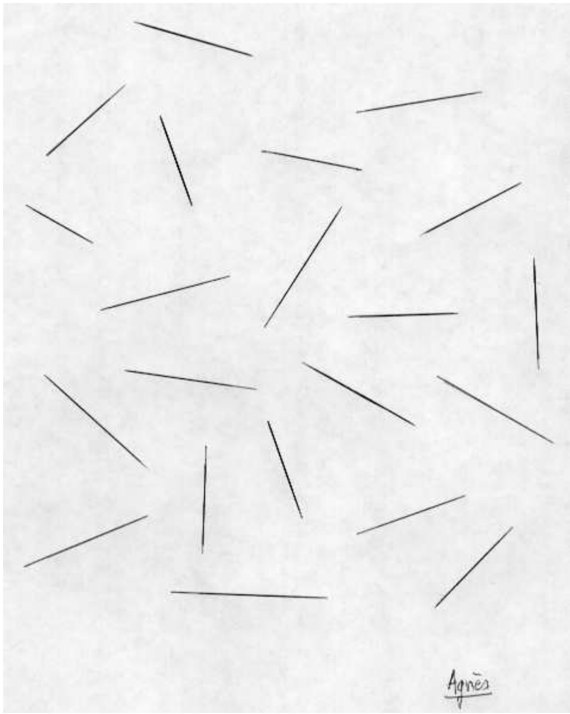


(b) Maximal meaningful intervals

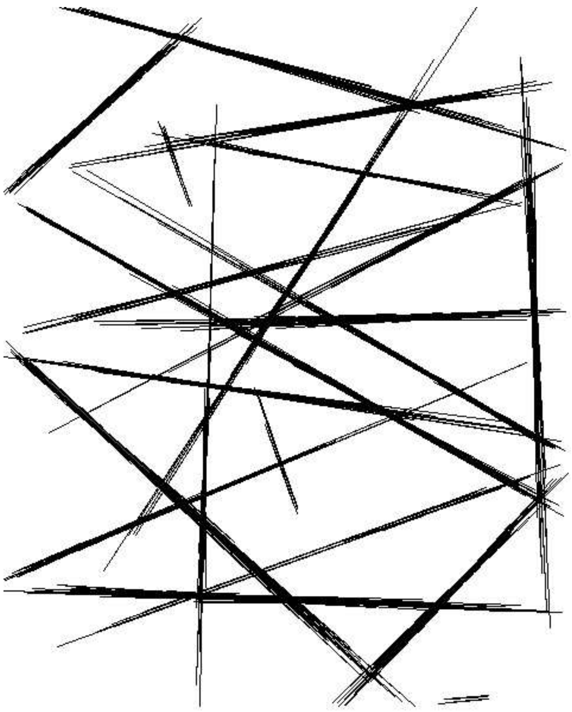


(c) Maximal meaningful modes

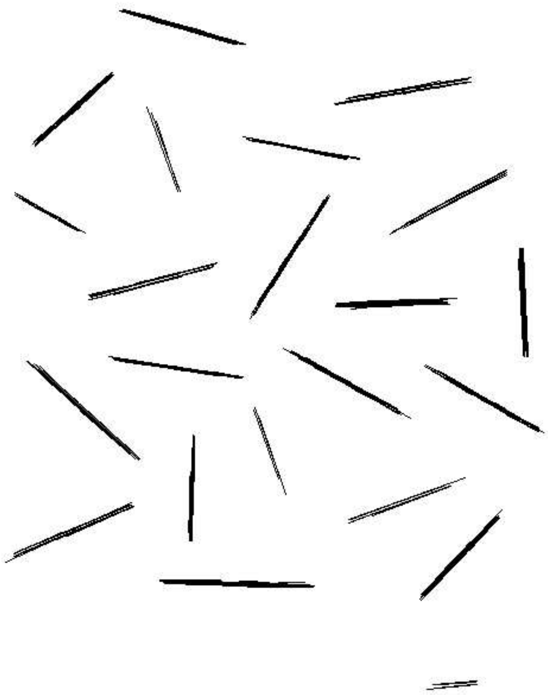
Figure 8: Comparison between maximal meaningful intervals and maximal meaningful modes.



(a) The original "Pencil strokes" image

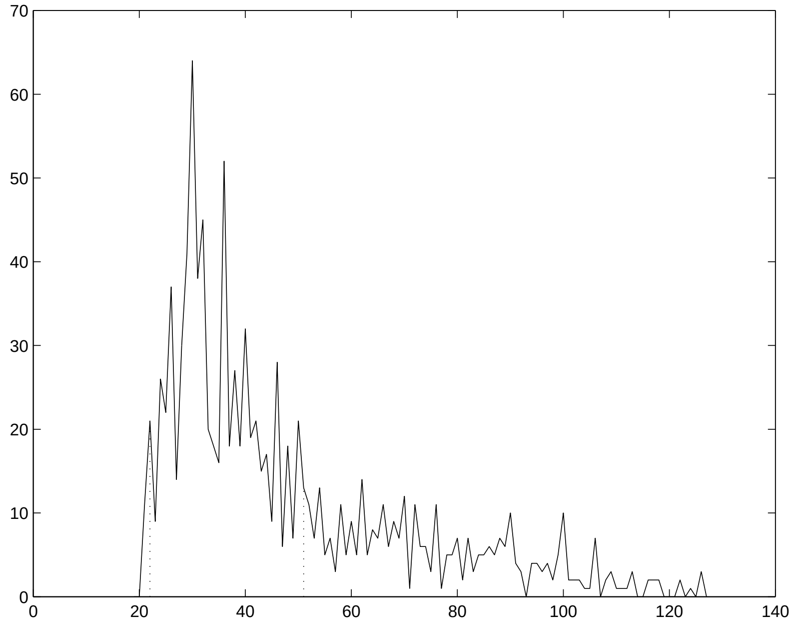


(b) ε-meaningful alignments with $\varepsilon = 10^{-3}$

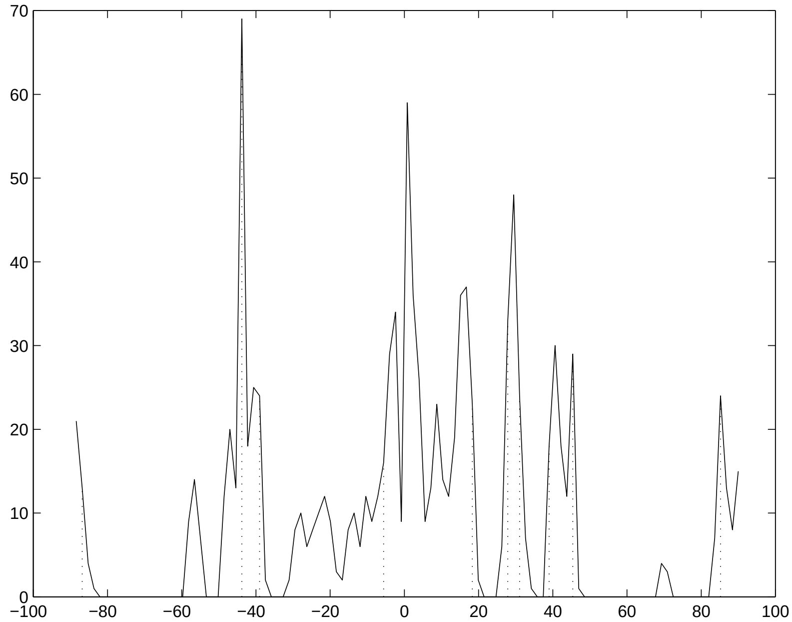


(c) maximal meaningful alignments

Figure 9: Pencil strokes image: meaningful and maximal meaningful alignments.

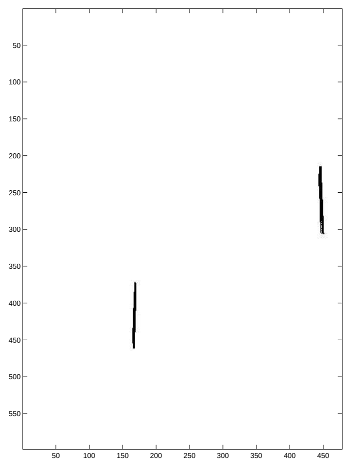


(a) The histogram of the length of the maximal meaningful segments: one mode: $(22, 51)$

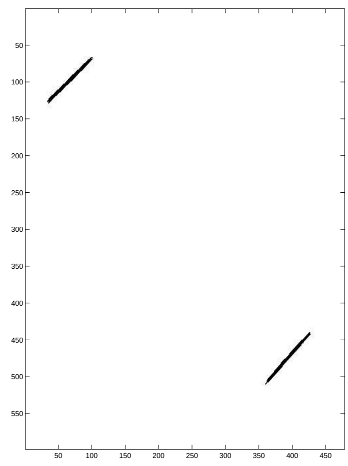


(b) The histogram of the orientation modulo π of the maximal meaningful segments. We measure the orientation in degrees. We find five modes: $(85, 93)$, $(-44, -39)$, $(-5, 18)$, $(28, 31)$, $(39, 45)$.

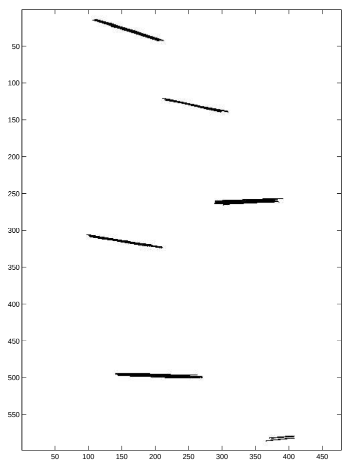
Figure 10: Histogram of length and histogram of orientation with maximal meaningful modes.



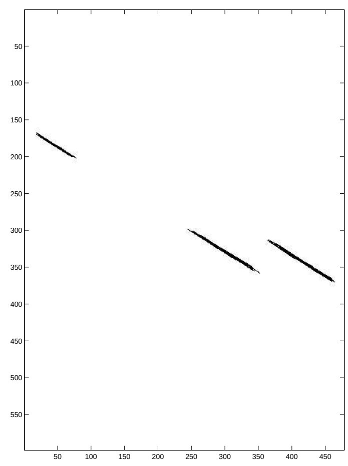
(a) Segments with orientation in the mode $(85, 93)$



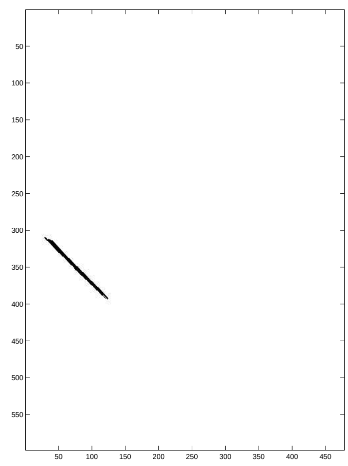
(b) Segments with orientation in the mode $(-44, -39)$



(c) Segments with orientation in the mode $(-5, 18)$



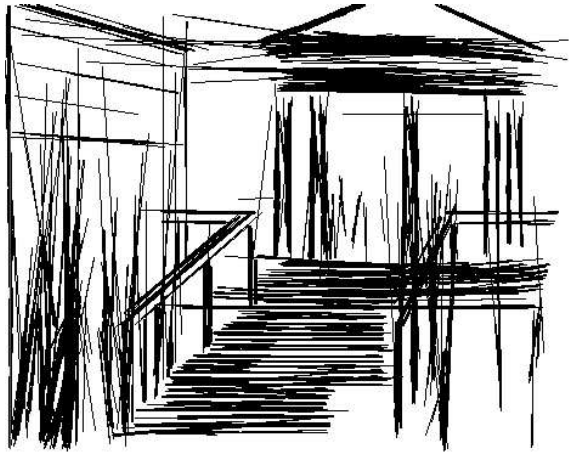
(d) Segments with orientation in the mode $(28, 31)$



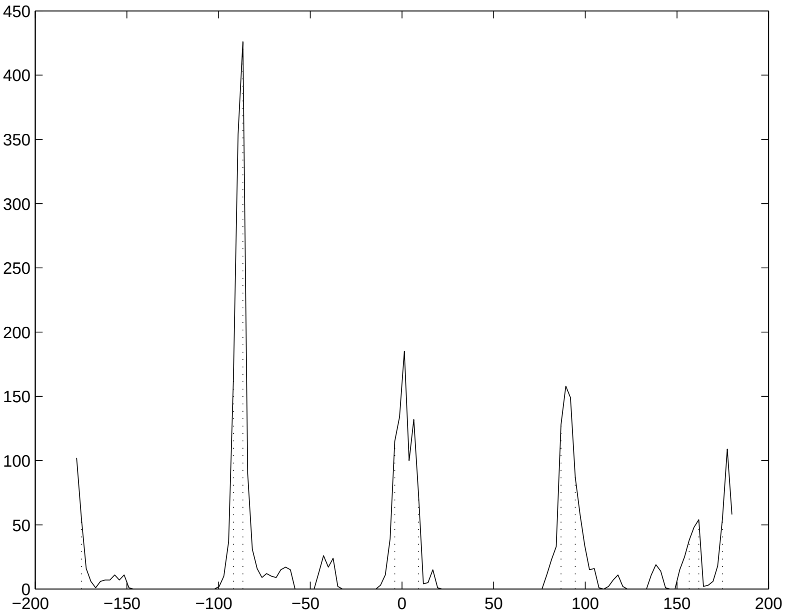
(e) Segments with orientation in the mode $(39, 45)$

Figure 11: Grouping of segments according to common orientation.

(a) The original image: Uccello's painting

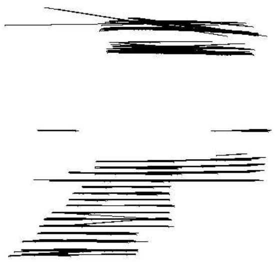


(b) Maximal ε-meaningful segments for $\varepsilon = 10^{-6}$.



(c) Histogram of orientations modulo 2π of the maximal meaningful segments: five maximal meaningful modes.

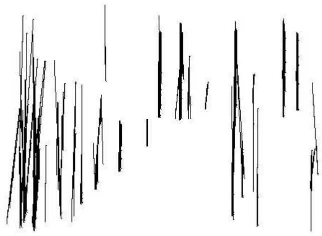
Figure 12: Uccello's painting: maximal meaningful alignments and histogram of orientations



(a) Segments with orientation in the mode $(-4, 9)$

(b) Segments with orientation in the mode $(175, 185)$

(c) Segments with orientation in the mode $(-92, -87)$

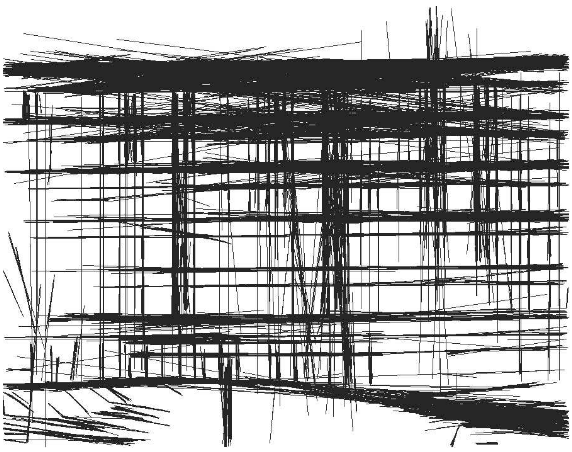


(d) Segments with orientation in the mode $(87, 95)$

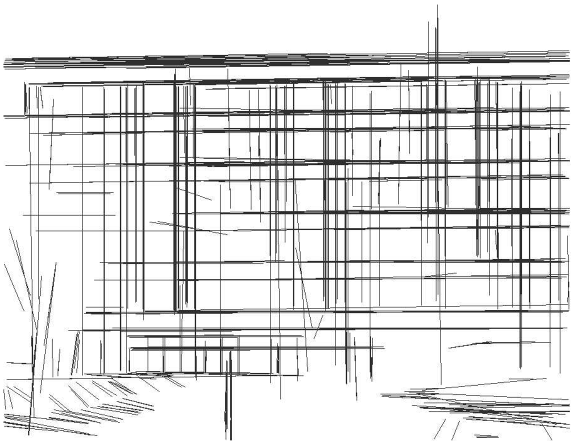
(e) Segments with orientation in the mode $(156, 162)$

Figure 13: Grouping of segments according to common orientation.

(a) The original "Cachan building" image



(b) Maximal ε-meaningful alignments with $\varepsilon = 10^{-6}$.



(c) Minimal description of the set of maximal meaningful segments

Figure 14: Building image: maximal meaningful segments and their minimal description.