Partial gestalts

Agnès Desolneux, Lionel Moisan and Jean-Michel Morel

CMLA, ENS Cachan, 61 av. du président Wilson, 94235 Cachan cedex, FRANCE.

DRAFT

Abstract

We expose a recently introduced method for computing geometric structures in a digital image, without any a priori information. According to a basic principle of perception due to Helmholtz, an observed geometric structure is perceptually "meaningful" if its number of occurences would be very small in a random situation : geometric structures are characterized as large deviations from randomness. This leads us to define and compute "partial gestalts" (in computer vision, features) like alignments, edges, clusters, groups of objects similar for some quality in an image by a parameter-free method. Maximal meaningful objects are defined, computed, and the results compared with the ones obtained by classical algorithms. A discussion ensues : are the partial gestalt (or feature) detectors enough to build up computer vision algorithms ? We show by experiments that this is rather an illusion : the "conflicts" between gestalt laws that were discussed in a phenomenological framework by the gestaltists indeed arise in well chosen images and lead to wrong (but explainable) detections. We show how no "good" feature detector can work if not applied simultaneously and in conflict with all other detectors. We are led to the conclusion that no correct image analysis can be obtained if the gestalt conflicts and the subsequent masking phenomena are not adressed.

I. What is a partial Gestalt ?

According to Gestalt theory, "grouping" is the main process in our visual perception (see [11], [24]). Whenever points (or previously formed visual objects) have one or several characteristics in common, they get grouped and form a new, larger visual object, a "Gestalt". Some of the main grouping characteristics are proximity (clustering), colour constancy (connectedness), "good continuation" (differentiability of boundaries), alignment (presence of straight lines or objects of a same kind aligned), parallelism (between lines, oriented objects, etc.), similarity of shape (between objects), common orientation (between points or oriented objects) convexity (of boundaries, of a group) and closedness (for a curve), constant width, ... 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 and so on. A simple drawn object like a square whose boundary has been drawn in black with a pencil on a white sheet perceived by connectedness (the boundary is a black line), constant width (of the stroke), convexity and closedness (of the black pencil stroke), parallelism (between opposite sides), orthogonality (between adjacent sides), finally equidistance (of both pairs of opposite sides).

Thus, we must distinguish between what we shall call *global* gestalt and *partial* gestalt.

The square is a global gestalt, but it is the result of a long list of concurring geometric qualities which we shall call *partial gestalts.* One can summarize the efforts of Computer Vision a a way to compute the (very diverse in nature) partial gestalts. To take an instance, the snakes method [12] attempts to capture the closed smooth curves, a combination of the "closedness" and "good continuation" gestalts. In the same way, have been proposed in Computer Vision : alignment detectors (e.g. Hough transforms), edge detectors, angle detectors, shape recognition methods (the similarity of shape gestalt), and texture segmenters, that is, a general way to group points according to common features which are, again, nothing but partial gestalts. The gestaltists have attempted between 1923 and 1975 to make a list of all partial gestalts relevant to human and animal vision. They also discovered the existence of conflicts between partial gestalts and the ability of the human vision to find the best solution to these conflicts. This solution results into the surprising phenomenon of *masking.* When the "best explanation" of a figure is given by one partial gestalt, other possible explanations are masked and the viewers are no more aware of their possibility. For instance, an hexagon is very close to a circle and a computer vision "circle detector" is very likely to detect it as a circle. This makes sense, but clearly the polygon explanation is more adequate and must be preferred. The circle explanation is then removed from awareness.

To summarize, Computer Vision, in a match with Gestalt Theory, seems to be stuck at the starting block. Indeed, not only partial gestalts (detectors) have not received a full treatment and agreement, but the main problems, namely the alluded two basic phenomena, *collaboration of partial gestalts* and *conflicts resulting into masking* are seldom adressed.

In this paper, we intend to show by simple formulas followed by experimental evidence that :

- there is a very simple computational principle which allows one to compute any partial gestalt (Section 2)
- this computational principle can be applied to a fairly wide series of examples of partial gestalts, namely alignments, clusters, boundaries, grouping by orientation or by size or by grey level (another recent example is the gestalt "vanishing points" [1]) ;

- the experiments yield evidence that in natural world images, partial gestalts often collaborate. [1] Thus, in most cases, a partial gestalt detector seems to extract correctly most of the scene elements, but this may well be an illusory success ;
- as a first evidence of the recursive character of gestalt laws, we push one of the experiments to prove that the partial gestalt recursive building up can be led up to the third level (gestalts built by three successive partial gestalt grouping principles) ;
- all partial gestalts are likely to lead to wrong scene interpretations because there may always be a more adapted partial gestalt which better explains the scene.

As a conclusion, we point out that all computer vision algorithms based on a pile of partial gestalts (feature detectors in the language of Computer Vision) are likely to yield wrong conclusions to some images. The main focus of research in Computer Vision should then be the synthesis of partial gestalts and basic research on the main gestaltic principles : masking, *Gliederung* or articulation whole-parts, [16] and the so called articulation without remainder (*articulazione senza resti*) [11]. We finally show some experiment suggesting that Minimal Description Length principles may be adequate to resolve the so called conflicts between gestalt laws.

II. Detecting partial gestalts

A. General detection principles

In this subsection, we review quickly anterior work where we proposed a general principle for computing any partial gestalt. This principle was applied to alignments and boundaries and will be extended in this paper to several new examples of computable partial gestalts.

Helmholtz Principle. In [3], we outlined a computational method to decide whether a given partial gestalt (computed by any segmentation or grouping method) is reliable or not. We treated the detection of alignments, as one of the most basic gestalts (see [24]). As we shall recall, our method gives *absolute thresholds*, depending only on the image size, permitting to decide when a peak in the Hough transform is significant or not.

A geometrically meaningful event is an event that, according to probabilistic estimates, should not happen in an image and therefore makes sense. This informal definition imme-

[1] Message to the referee : a reference is missing here to work made by Ingo Wundrich (Bochum) on the redundancy of gestalts. Will be added in the final version

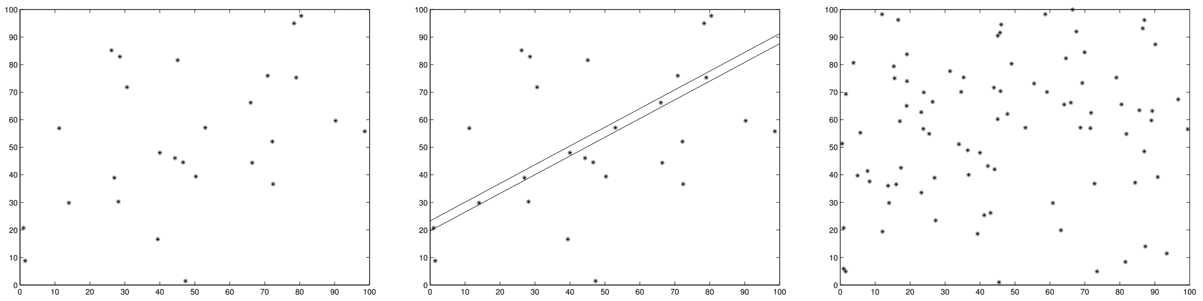

Fig. 1. An illustration of Helmholtz principle : non casual alignments are automatically detected by Helmholtz principle as a large deviation from randomness. Left, 20 uniformly randomly distributed dots, and 7 aligned added. Middle : this meaningful (and seeable) alignment is detected as a large deviation. Right : same aligment added to 80 random dots. The alignment is no more meaningful (and no more seeable). In order to be meaningful, it would need to contain at least 11 points.

diately raises an objection : if we do probabilistic estimates in an image, this means that we have an *a priori* model ([7]). We are therefore losing any generality in the approach, unless the probabilistic model could be proven to be "the right one" for the image under consideration. In fact, our proposition has been to do statistical estimates without any image model. Instead, we applied a general perception principle which we called Helmholtz principle. This principle yieds computational grouping thresholds associated with each gestalt quality. It can be stated in the following generic 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 and enough to group O_1, O_k ? In order to answer this question, we make the following mental experiment : we assume *a contrario* 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. We finally 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) [3] We say that an event of type "such configuration of points has such property" is ε-meaningful, if the expectation of the number of occurences of this event is less than ε under the uniform random assumption.

As an example of generic computation we can do with this definition, let us assume that the probability that a given object O_i has the considered quality is equal to p. Then, under the independence assumption, the probability that at least k objects out of the observed n have this quality is

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

i.e. the tail of the binomial distribution. In order to get an upper bound of the number of false alarms, i.e. the expectation of the geometric event happening by pure chance, we can simply multiply the above probability by the number of tests we perform on the image. Let us call N_T the number of tests. Then in most cases we shall consider in the next subsections, a considered event will be defined as ε-meaningful if

$$N_T B(p, n, k) \leqslant \varepsilon.$$

We call in the following the left hand member of this inequality on the the "number of false alarms" (NFA).

When $\varepsilon \leqslant 1$, we talk about meaningful events. This seems to contradict the necessary 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 [3] for a complete discussion of this definition.

The general method we have just outlined can be viewed as a systematization of Stewart's "MINPRAN" method [23]. The method was presented as a new paradigm, but was applied only to the 3D alignment problem. Now, Stewart actually adressed but did not solve two problems we have intended to overcome, in order to make the method fully general. One of the problems raised by Stewart was the generation of the set of samples, which generates in Stewart's method at least three user's parameters and the second one was the severe restriction about the independance of samples. We actually solved both difficulties simultaneously by introducing the number of samples as an implicit parameter of the method (computed from the image size and Shannon's principles) and by replacing in all calculations the "probability of hallucinating a wrong event" by the "expectation of the number of such hallucinations", namely what we call the false alarm rate NFA.

The method we develop here has probably been proposed several times in Computer Vision (e.g. in the early Lowe work [15]), but, to the best of our knowledge, not systematically developped.[2]

B. Meaningful alignments

Let us start by our first example, the detection of straight lines in an image. This was published elsewhere, but we explain it for two reasons : first for a sake of completeness and second because the developped formalism is in fact general and applicable to the other gestalt qualities we consider in the sequel.

Since images are blurry, noisy and aliased, we cannot hope for a strong accuracy in direction measurement at each pixel, and we shall, without need for many explanations, fix the accuracy of a measured gradient direction at a point equal to a factor $p\pi$ radians. This

[2] Let us quote David Lowe's program, whose mathematical consequences were partly developped in Stewart [23] : *"we need to determine the probability that each relation in the image could have arisen by accident, $P(a)$. Naturally, the smaller that this value is, the more likely the relation is to have a causal interpretation. If we had completely accurate image measurements, the probability of accidental occurence could become vanishingly small. For example, the probability of two image lines being exactly parallel by accident of viewpoint and position is zero. However, in real images there are many factors contributing to limit the accuracy of measurements. Even more important is the fact that we do not want to limit ourselves to perfect instances of each relation in the scene - we want to be able to use the information available from even approximate instances of a relation. Given an image relation that holds within some degree of accuracy, we wish to calculate the probability that it could have arisen by accident to within that level of accuracy. This can only be done in the context of some assumption regarding the surrounding distribution of objects, which serves as the null hypothesis against which we judge significance. One of the most general and obvious assumptions we can make is to assume that a background of independently positioned objects in three-space, which in turn implies independently positioned projections of the objects in the image. This null hypothesis has much to recommend it. (...) Given the assumption of independence in three-space position and orientation, it is easy to calculate the probability that a relation would have arisen to within a given degree of accuracy by accident. For example if two straight lines are parallel to within 5 degrees, we can calculate that the chance is only* $5/180 = 1/36$ *that the relation would have arisen by accident from two independent objects."*

means that a casual alignment of a direction with a prefixed one happens with probability p. In practice, $p = \frac{1}{16}$ is the best we can hope from digital images (and is even optimistic for aliased images). We consider the following event : "on a discrete segment of the image, joining two pixel centers, and with length l counted in points at Nyquist distance, at least k points have the same direction as the segment with precision p." The direction at each point is computed as the direction of the gradient rotated by $\frac{\pi}{2}$.

Definition : Consider a segment S of length l containing k aligned points. We call number of false alarms of S,

$$NFA(S) = N^4 \sum_{j=k}^{l} \binom{l}{j} p^j (1-p)^{l-j}.$$

We say that S is ε-meaningful if $NF(S) \leqslant \varepsilon$.

This notion of ε-meaningful segments has to be related to the classical "α-significance" in statistics, where α is simply $\frac{\varepsilon}{N^4}$. 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 overcomes a difficulty raised by Stewart [23]. This is not exactly the same situation as in failure detection, where the failures are somehow disjoint events. If on a straight line we have found a very meaningful segment S, then by enlarging slightly or reducing slightly S, we still find a meaningful segment. This means that meaningfulness cannot be a univoque criterion for detection, unless we can point out the "best meaningful" explanation of what is observed as meaningful. This is done by the following definition, which can be adapted as well to meaningful boundaries [4], meaningful edges [4], meaningful modes in a histogram [5] and clusters.

Definition 2: We say that an ε-meaningful geometric structure A is maximal meaningful if

- it does not contain a strictly more meaningful structure : $\forall B \subset A$, $NF(B) \geqslant NF(A)$.
- it is not contained in a more meaningful structure : $\forall B \supset A$, $B \neq A$, $NF(B) > NF(A)$.

It is proved in [5] that maximal structures cannot overlap, which is one of the main theoretical outcomes validating the above definitions.

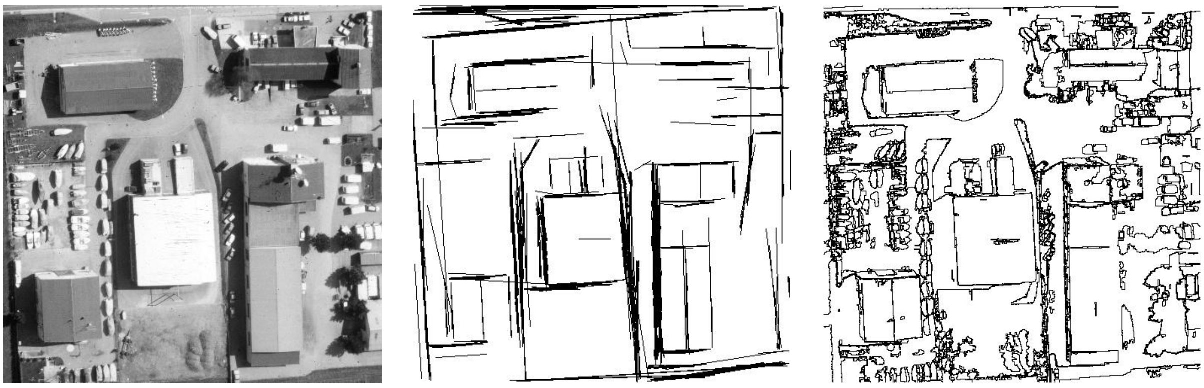

Fig. 2. Two partial gestalts, alignments and boundaries. Left : original aerial view (source : INRIA), middle : maximal meaningful alignments, right : maximal meaningful boundaries.

C. Edge and boundary detectors

We shall now review briefly our second example of partial gestalt, the boundaries : a classical example in Computer Vision ! The scope here is to point out the existence of a parameterless boundary detector deduced from the Helmholtz principle and again to compare it with the other definitions of partial gestalts. A much more detailed treatment is given in [4]. Let u be a discrete image of size $N \times N$. We consider the level lines at quantized levels $\lambda_1, ..., \lambda_k$. The quantization step q is chosen in such a way that level lines make a dense covering of the image. A level line can be computed as a Jordan curve contained in the boundary of a level set with level λ,

$$\chi_\lambda = \{x/u(x) \leqslant \lambda\} \quad \text{and} \quad \chi^\lambda = \{\mathrm{x/u(x)} \geqslant \lambda\}.$$

(See [2].) Notice that along a level line, the gradient of the image must be everywhere above zero. Otherwise the level line contains a critical point of the image and is highly dependent upon the image interpolation method. Thus, we consider in the following only level lines along which the gradient is not zero. The interpolation considered in all experiments below is the bilinear interpolation.

Let L be a level line of the image u. We denote by l its length counted in independent points. In the following, according to Helmholtz principle, we will detect structures against the null hypothesis that points at a geodesic distance (along the curve) larger than 2 are independent. More precisely, the gradient magnitudes at these points are independent random variables). Let x_1, x_2,...x_l denote the l considered points of L. For a point $x \in L$, we will denote by $c(x)$ the contrast at x. It is defined by

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

where ∇u is computed by a standard finite difference scheme on a 2×2 neighborhood [3]. For $\mu > 0$, we consider the event : for all $1 \leqslant i \leqslant l$, $c(x_i) \geqslant \mu$, i.e. each point of L has a contrast larger than μ. From now on, all computations are performed in the Helmholtz framework explained in the introduction : we make all computations as though the contrast observations at x_i were mutually independent. Since the l points are independent, the probability of this event is

$$\text{Prob}(c(x_1) \geqslant \mu) \cdot \text{Prob}(c(x_2) \geqslant \mu) \cdot ... \cdot \text{Prob}(c(x_l) \geqslant \mu) = H(\mu)^l, \tag{2}$$

where $H(\mu)$ is the probability for a point on any level line to have a contrast larger than μ. An important question here is the choice of $H(\mu)$. Shall we consider that $H(\mu)$ is given by an *a priori* probability distribution, or is it given by the image itself (i.e. by the histogram of gradient norm in the image ? In the case of alignments, we took by Helmholtz principle the orientation at each point of the image to be a random, uniformly distributed variable on $[0, 2\pi]$. Here, in the case of contrast, it does not seem sound at all to consider that the contrast is uniformly distributed. In fact, when we observe the histogram of the gradient norm of a natural image, we notice that most of the points have a “small” contrast (between 0 and 3), and that only a few points are highly contrasted. This is explained by the fact that a natural image contains many flat regions (the so called “blue sky effect”, [10]). In the following, we will consider that $H(\mu)$ is given by the image itself, which means that

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

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

Definition 3: (Number of false alarms) Let L be a level line with length l, counted in independent points. Let μ be the minimal contrast of the points x_1,..., x_l of L. The number of false alarms of this event is defined by

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

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

Notice that the number N_{ll} of level lines is provided by the image itself. We now define ε-meaningful level lines. The definition is analogous to the definition of ε-meaningful modes of a histogram or to the definition of alignments : the number of false alarms of the event is less than ε.

Definition 4 (ε-meaningful boundary) A level line L with length l and minimal contrast μ is ε-meaningful if

$$NFA(L) =\leqslant \varepsilon. \tag{5}$$

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

D. Histogram modes and groups

As we mentionned in the introduction, the main gestaltic grouping principle is this : points or objects having one or several features in common are being grouped because they have this feature in common. We shall consider here only grouping by a single feature and we shall see that this single-feature grouping already yields relevant results. We face here a general problem : assume k objects O_1, ... O_k, among a longer list O_1,..., O_n, have some quality Q in common. Assume that this quality is actually measured as a real number. Then our decision of whether the grouping of O_1,... , O_k is relevant must be based on the fact that the values $Q(O_1)$, ..., $Q(O_k)$ make a *meaningful mode* of the histogram of P. Thus, the single quality grouping is led back to the question of an automatic, parameterless, histogram mode detector. Of course, this mode detector depends upon the kind of feature under consideration. We shall consider two paradigmatic cases, namely the case of orientations, where the histogram can be assumed by Helmholtz principle to be flat, and the case of the objects sizes (areas) where the null assumption is that the size histogram is decreasing.

D.1 Meaningful groups of objects according to their orientation and to their grey level

In the sequel, we quantize the possible orientations and grey levels in the usual way and we assume that the M values of orientation (or grey level) are i.i.d. uniformly on $\{1, 2, ..., L\}$. Consider an interval $[a, b] \subset [1, L]$ and let $k(a, b)$ denote the number of objects with gestalt value in $[a, b]$. We define $p(a, b) = (b - a + 1)/L$ as the a priori probability that an object's quality $P(O)$ falls in $[a, b]$. With the same generic argument as in Section 1, we have

Definition 5: An interval $[a, b]$ is ε-meaningful if

$$NF([a, b]) = Ni \times B(p(a, b), M, k(a, b)) \leqslant \varepsilon,$$

where Ni is the number of considered intervals ($Ni \simeq L(L + 1)/2$). An interval $[a, b]$ is said maximal meaningful if it is meaningful and if it does not contain, or is not contained in, a more meaningful interval (see Definition 2

It can be proved in the same way as for alignments that maximal meaningful intervals do not intersect. Thus, we get an operational definition of meaningful modes as disjoint subintervals of $[1, L]$.

D.2 Size of objects

The preceding arguments are easily adapted to Helmholtz type assumptions on nonuniform histograms. A very generic way to group objects in an image is their similarity of size. This similarity lets groups perceptually pop out. Now, it would be a total nonsense to assume any uniform law on the objects sizes. There are several powerful arguments in favour of a statistical decreasing law for size. These arguments derive from perspective laws, or from the occlusion dead leaves model, or directly from statistical observations of natural images [8]. Our Helmholtz qualitative hypothesis is then : the prior distribution of the size of objects is **decreasing**.

Definition 6: An interval $[a, b]$ is ε-meaningful (for the decreasing assumption) if

$$NFA([a, b]) = Ni \cdot \max_{p \in \mathcal{D}} B(p(a, b), M, k(a, b)) \leqslant \varepsilon,$$

where $\mathcal{D}$ is the set of decreasing probability distributions on $\{1, 2, ..., L\}$, and $p(a, b) = \sum_{i=a}^{b} p_i$.

In the same way as in the flat histogram assumption, one can define maximal meaningful intervals and prove that maximal meaningful intervals do not intersect [6].

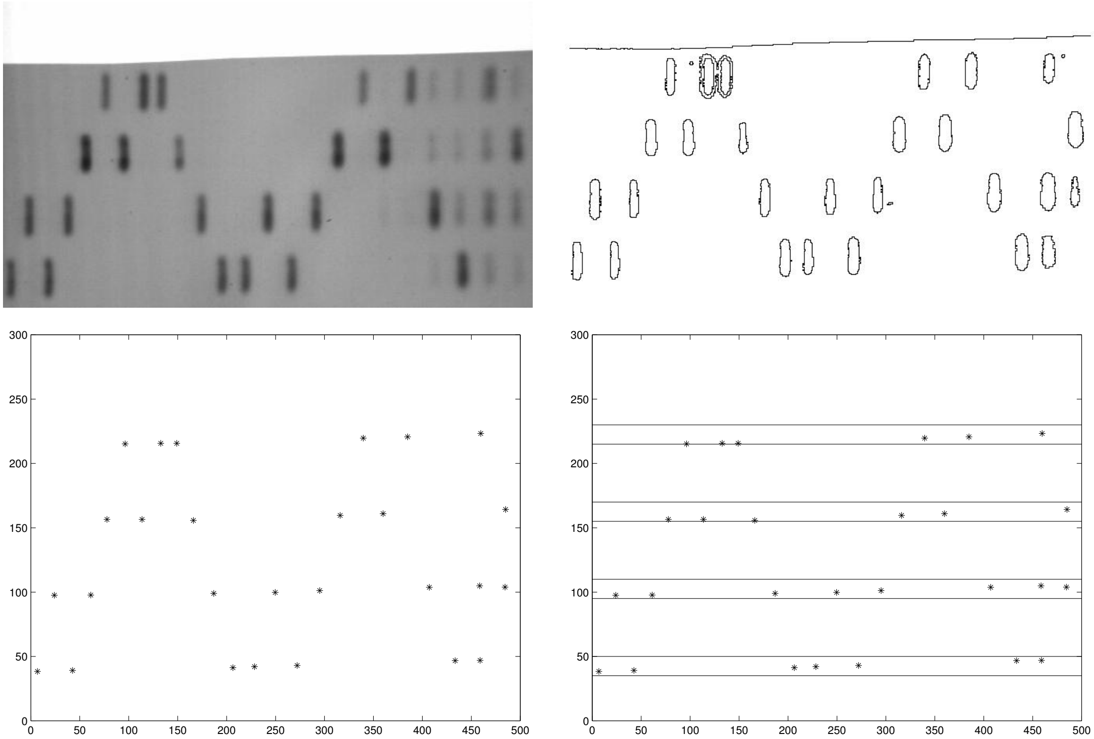

Fig. 3. Gestalt grouping principles at work for building an "order 3" gestalt (alignment of blobs of the same size). Top-left, original DNA image. Top-right, maximal meaningful boundaries. Bottom-left, barycenters of all meaningful regions whose area is inside the only maximal meaningful mode of the region areas histogram. Bottom-right : meaningful alignments of these points. *note to the referee : this experiment is not complete because we did not use thin enough widths for the strips. By taking in the algorithm into account thinner widths, one can detect some diagonal alignments in the same picture (actually also visible). This will be fixed in the final version of the present paper.*

E. Object alignments

The gestalt we now consider is not the same as the alignment gestalt considered at the begining of Section 2, where the aligned points had their own orientation. Here, we consider the case of objects whose barycenters are aligned. Assume that we observe M objects of a certain kind in an image. Our null hypothesis for the application of Helmholtz principle will be that the M barycenters (x_i, y_i) are i.i.d. uniformly on a domain Ω. A meaningful alignment of points must be a meaningful peak in the Hough Transform (see [13], [22] for a very similar approach). Now, the accuracy matter must be adressed. Points will be supposed to be aligned if they all fall into a strip thin enough, in sufficient number. Let S be a strip of width a. Let $p(S)$ denote the prior probability for a point to fall in S, and let $k(S)$ denote the number of points (among the M) which are in S. The following definition permits to compute all strips where a meaningful alignment is observed.

Definition 7: A strip S is ε-meaningful if

$$NF(S) = N_s \cdot B(p(S), M, k(S)) \leqslant \varepsilon,$$

where N_s is the number of considered strips. (One has $N_s \simeq 2\pi(R/a)^2$, where R is the diameter of Ω and a the minimal width of a strip.)

In practice, we sample all possible strip widths in a logarithmic scale (about 8 widths) and we sample accordingly the angles between tested strips in order to get a good covering of all directions. Thus, the number of strips N_s only depends on the size of the image and this yields a parameterless detection method.

F. Meaningful groups, or clusters

F.1 Model

The cluster example is the seminal one in Gestalt theory where it is called “proximity” gestalt ([11]). Assume that we see a set of dots on a white sheet and those dots happen to be grouped in one or several clusters, separated by desert regions. In order to characterize each cluster as an event with very low probability, we shall make all computations with the *a contrario* or *background* model that the dots have been uniformly distributed over the white sheet. This amounts to consider the dots as distributed over the sheet by a binomial process. We then call A the simply connected region, with area σ (the area of the sheet is normalized to 1), containing a given observed cluster of dots. Assume that we observe k points in A and $M - k$ outside. Then the ”cluster probability” of observing at least k points among the M inside A is given by

$$B(\sigma, M, k) = \sum_{i=k}^{M} \binom{M}{i} \sigma^i (1-\sigma)^{M-i}.$$

It is easily checked by large deviations estimates that if k/M exceeds A, this probability can become very small. Now, the event is not a generic event in that we have fixed a posteriori the domain A. The real a priori event we can define is "there is a simply connected domain A, with area σ, containing at least k points". Since the number of such domains A is a priori huge, we see that the expectation of such an event is by no means small. In the following, we shall consider a smaller set of domains $\mathcal{D}$ with cardinality N_D.

Definition 8: We say that a group of dots is ε-meaningful if $N_{\mathcal{D}}B(\sigma, M, k) \leqslant \varepsilon$.

In order to define $\mathcal{D}$ in a realistic way, we have to *sample* the set of simply connected domains by encoding their boundaries as "low resolution" Jordan curves. We consider a low resolution grid in the image, which for a sake of low complexity we take to be hexagonal, with mesh step m. The number of curves with length lm starting from a point and supported by the grid is bounded from above by 2^l. The overall number of low resolution curves with length lm is bounded by $N_m^2 2^l$, where $N_m = \frac{4}{3\sqrt{3}\,m^2}$ is the (approximate) number of mesh points lying on the sheet. Thus, we can consider several resolutions $m_1 < m_2 < \ldots < m_q$, for example in logarithmic scale, with m_1 larger than the pixel size and m_q lower than the image size, so that q is actually a small number. Our set of domains will be the set of all Jordan curves at all given resolutions, with discrete length measured in the corresponding mesh less than a fixed length L. Thus, the overall number of possible low resolution curves is bounded by $N^2 q 2^L$, where $N = N_{m_1}$. Notice that all numbers here are relatively small since the phenomenology excludes very intrincated cluster to be perceived. Thus, L is always taken to be smaller than, say, 30. We therefore define a meaningful cluster as a set of points contained in a low resolution curve defined as above, and such that $N^2 q 2^L B(\sigma, M, k) \leqslant \varepsilon$.

It can also happen that a cluster is not overcrowded, but only fairly isolated from the other dots. In such a case, we can find a low resolution curve surrounding the cluster and such that some dilate of the curve does not contain any point. Accordingly, we can modify the probability of the event : "the simply connected domain A has area σ, contains at least k points and is surrounded by an empty thick curve C with area σ'". In such a case the definition of meaningfulness for an isolated cluster becomes

$$N^2 qr 2^L \sum_{i=k}^{M} \binom{M}{i} \sigma^i (1-\sigma-\sigma')^{M-i} \leqslant \varepsilon,$$

where r is the number of allowed values for σ'.

F.2 Algorithm

Since this part has not been published elsewhere, and since the cluster detection algorithm is not obviously fast, we shall give some implementation details. Let P_i, $i = 1..M$ be the points observed. We assume that M is reasonably small, say $M \leqslant 1000$. We write $d(P_i, P_j)$ for the usual Euclidean distance between P_i and P_j.

1. Computation of the connectedness tree

Stage a. Sort the values $d(P_i, P_j), 1 \leqslant i < j \leqslant M$ to obtain the nondecreasing sequence

$$d_1 = d(P_{i_1}, P_{j_1}), d_2 = d(P_{i_2}, P_{j_2}), ...d_{M'} = d(P_{i_{M'}}, P_{j_{M'}})$$

with $M' = M(M-1)/2$. The complexity of this stage is $O(N^2 \log N)$ in the worst case (sort).

Stage b. Compute recursively the connectedness tree :

- at the begining, each point P_i is a tree (with a unique element).

- at step k, we look for the root A (resp. B) of the unique tree having P_{i_k} (resp. P_{j_k}) among its leaves. If $A \neq B$, we fuse A and B as the children of a new node, and store the value d_k at this node.

- iterate this process until all points have been agregated in a single tree.

The complexity of this stage is about $O(N^2 \log N)$ in the average, since at step k we need $O(\log k)$ operations to look for the root of the tree having a given P_i among its leaves.

2. Computation of the meaningfulness of each cluster

In the connectedness tree, each subtree associated to a root node A with value δ corresponds to a δ-cluster (also named A) made of the connected union of the disks with radius

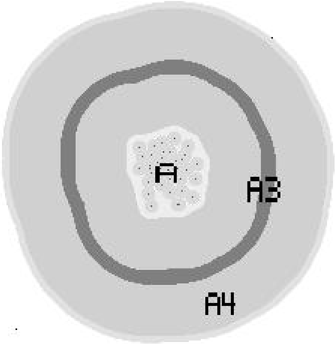

Fig. 4. The sets A_3 and A_4 associated to a cluster A.

$\delta/2$ centered on the points encountered in the subtree. We need to compute the meaningfulness of each cluster, for example by Meaningfulness(A) $= -\log_{10}\mathrm{NFA(A)}$, where $NFA(A)$ is the number of false alarms associated to A, itself given by

$$NFA(A) = N^2 qr 2^l \sum_{i=k}^{M} \binom{M}{i} \sigma^i (1-\sigma-\sigma')^{M-i}, \tag{6}$$

where k is the number of points defining A, l the discrete length of the (thick) low resolution curve enclosing A, σ' its area, and σ the area of A.

Now the point is to estimate l, σ and σ'. For each cluster A, we can compute ρ, the distance of A to the $\delta/2$-dilate of the remaining points. It is given by $\rho = \delta' - \delta$, where δ' is the value associated to the parent of A ($\delta' = +\infty$ if A is the root of the connectedness tree). If $\rho \neq 0$, we then compute, for $\alpha \in]0,1[$ fixed,

$$A_1 = D_\rho(A), \quad A_2 = A_1 - E_\rho(A_1), \quad A_3 = E_{\rho(1-\alpha)/2}(A_2), \quad A_4 = D_{\rho(1-2\alpha)/2}(A_3),$$

where E_r and D_r represent respectively the erosion and dilation operators associated to a disk with radius r (see Fig. II-F.2). We recall that $A = D_{\delta/2}(\cup_i\{P_i\})$, where the P_i's are the points encountered in the subtree defined by the node A.

The domain A_3 is a "thick low resolution curve" of width $\alpha\rho$, defined by the dilate of a low resolution curve C' lying on the hexagonal mesh. As we do not know where C' should precisely lie in A_3, only the A_4 domain will count as "empty domain", and not $D_{(1-\alpha)\rho/2}(A_3)$. Whe then define

$$l = C \cdot E^+\left[\frac{\mathrm{area(A_3)}}{\alpha^2\rho^2}\right], \quad s = \mathrm{area(A_4)}, \quad \sigma = \mathrm{area(A_2)},$$

where E^+ represents the upper integer part, and C is a constant such that for any continuous curve with length l_0, there exists a discrete curve with length less than Cl_0 supported by the unit step hexagonal mesh . We conjecture that $C \leqslant 3/2$, and use this value in practice.

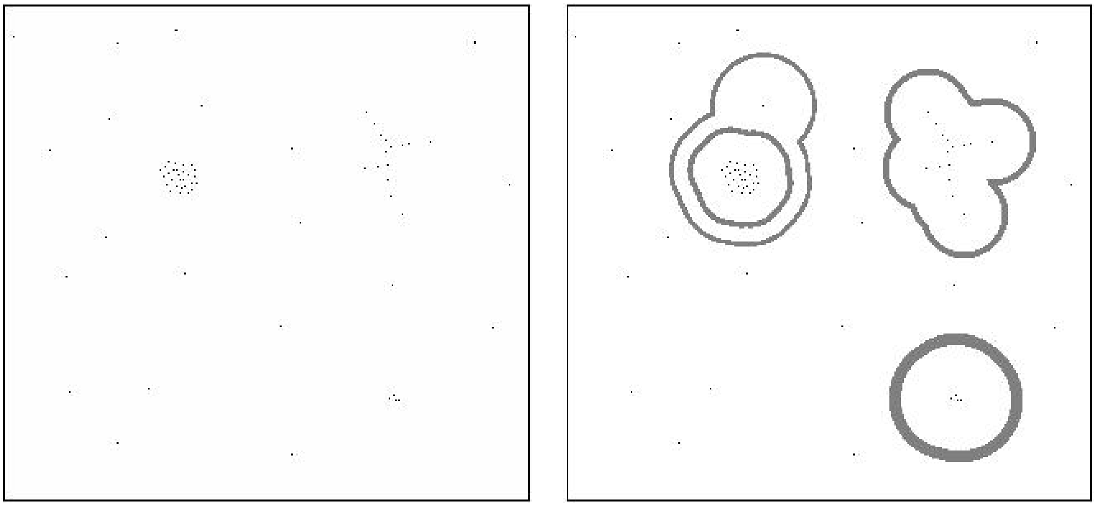

Fig. 5. Clusters of dots (left) and their automatic detection (right) : the thick (low resolution) curves indicate roughly the skeleton of the detected region which contains no dots. The cluster is meaningful when it contains enough points and is surrounded by a thick enough empty region.

The areas mentioned can be computed using a bitmapped image with a convenient size. This computation is done for some quantized values of α, provided that the associated discrete length l satisfies $l \leqslant L$. In theory, we cannot choose exactly ρ but we should take the nearest smaller value among the resolutions m_i. In practice, this does dot affect much the computations, since the number of resolutions chosen has very little effect on the NFA.

3. Maximal clusters.

Once we have computed the meaningfulness of each cluster, we can look for maximal meaningful clusters by selecting local maxima of the meaningfulness with respect to inclusion. Precisely, we shall say that a cluster A is maximal if for any child (resp. parent) B of A, one has $NFA(B) > NFA(A)$ (resp. $NFA(B) \geqslant NFA(A)$). As usual, we have the property that two maximal meaningful clusters are either equal or have no common point.

III. The limits of every partial gestalt detector

The preceding section argued in favour of a very simple principle, Helmholtz principle, applicable to the automatic and parameterless detection of any partial gestalt, in full agreement with our perception. In this section, we shall show by commenting briefly several experiments that "tout n'est pas rose" : there is a good deal of visual illusion in any apparently satisfactory result provided by a partial gestalt detector. We explained in the first section that partial gestalts often collaborate. Thus, in many cases, what has been detected by one partial gestalt will be corroborated by another one. For instance, boundaries and alignments in the experiment 2 are in good agreement. But what can be said about the experiment 6 ? In this cheetah image, we have applied an alignment detector. It works wonderfully on grass but we also see some alignment appearing somewhat unexpected in the fur. These alignments do exist : it happens that some lines are tangent to several of the convex dark spots on the fur. This generates a meaningful excess of aligned points on this line, the convex sets being smooth enough and having therefore on their boundary a long enough segment tangent to the line. Clearly, the right explanation for these "alignments" is the presence of a large number of convex spots. Thus, the presence in a image of a large number of smooth pieces of curves entails the possibility of meaningful alignments which are in no way the right description of what is being seen. We see that the convexity (or good continuation) gestalt should be searched simultaneously to the alignment gestalt. The detection of alignments which only are tangent lines to several smooth curves, should be inhibited when those smooth curves are detected by a "good continuation" principle. This can be stated in another way : no gestalt is just a positive quality : we see as alignment what indeed is aligned, but only under the condition that the alignment does not derive from several smooth curves... Let us mention that the good continuation principle has been extensively adressed in Computer Vision, first in [18], more recently in [21] and still more recently in [9].

The same argument applies to our next experimental example, in Figure 7. In that case, a dense cluster of points in present. Thus, it creates a meaningful amount of dots in many strips and the result is the detection of obviously wrong alignments. Again, the detection of a cluster must inhibit such alignment detections. We retrieve what the gestaltists called

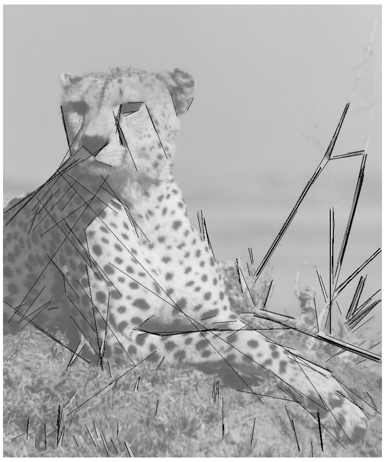

Fig. 6. Smooth convex sets or alignments ?

”conflicts of gestalts”.

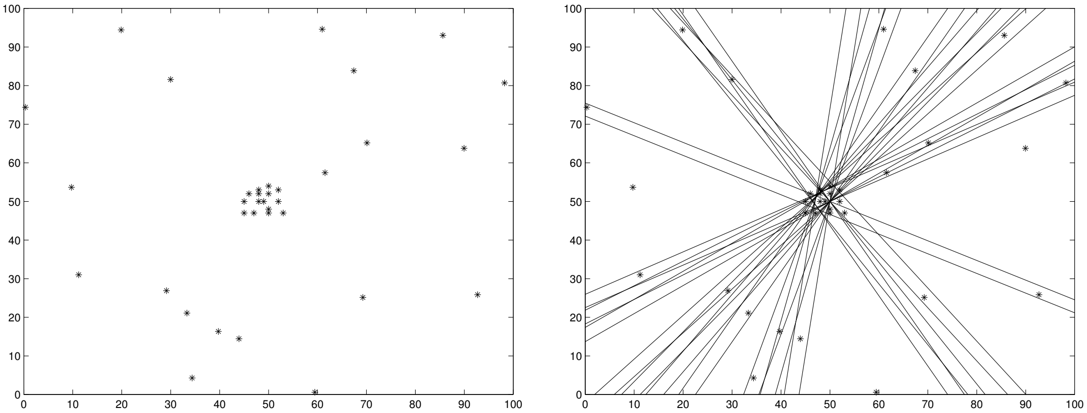

Fig. 7. One cluster, or several alignments ?

This same kind of gestalt conflict arises in Experiment 8. In this figure, a detector of arcs of circles has been applied (the detection definition is exactly the same as our definition of alignments). The main outcome of the experiment is this : since the MegaWave figure contains many smooth boundaries and several straight lines, lots of meaningful circular arcs are found. It may be discussed whether those circular arcs are present or not in the figure : clearly, any smooth curve is locally tangent to some circle. In the same way, two segments with an obtuse angle are tangent to several circular arcs. Now, the best explanation of the figure is not : “circular arcs”, but “smooth curves”, i.e. the “good continuation” gestalt. Thus, here again, we see that one partial gestalt must hide another. We also understand that we cannot hope any reliable explanation of any figure by summing up the results of one or several partial gestalts (or feature detectors). Only a global synthesis of all partial gestalts can give the correct result.

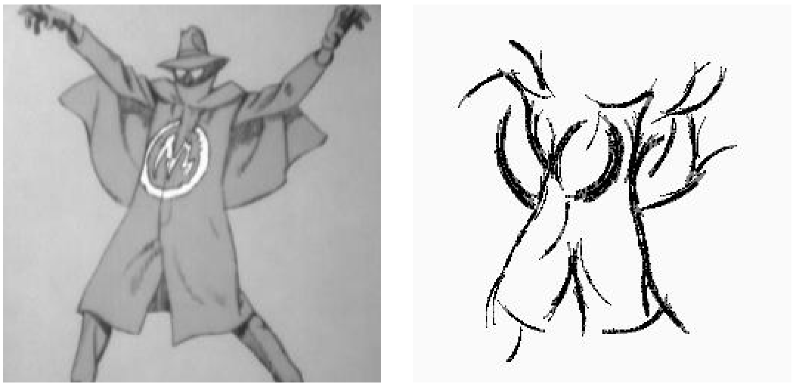

Fig. 8. Left : original “MegaWave” image. Right : an circular arc detector is applied to the image. Now, this image contains many smooth curves and obtuse angles to which meaningful circular arcs are tangent. This illustrates the necessity of the interaction of partial gestalts : the best explanation for the observed structures is ”good continuation” in the gestaltic sense, i.e. the presence of a smooth curve. Of course, this presence entails the presence of pieces of circles which are not the final explanation.

At this point, and in view of these experimental counterexamples, it may well be asked why partial gestalt detectors often work ”so well”. This is actually due to the redundancy of gestalt qualities in most natural images, as we explained in the first section with the example of a square. Indeed, most natural or synthetic objects are simultaneously conspicuous, smooth and have straight or convex parts, etc. Thus, in many cases, each partial gestalt detector will lead to the same group definition. Figure 9 illustrates the *collaboration of gestalt* phenomenon, which was not, in our opinion, given the right attention by the gestaltists. In the mentioned figure, we see several blobs of (very roughly) the same size, orientation and grey level. Using the histogram mode grouping as explained above, this very same group of blobs is correctly grouped (up to one outlier) as the unique maximal mode of the following histograms : orientation (of all blobs), size (area), and even mean grey level (inside each blob). Thus, the same group exists in agreement with three very different gestalts ! Each partial detector seems to be enough to perform the job... Now, we have seen this is an illusion which can be broken when partial gestalts do not collaborate.

We shall end this discussion by expressing some hope, and giving some arguments in

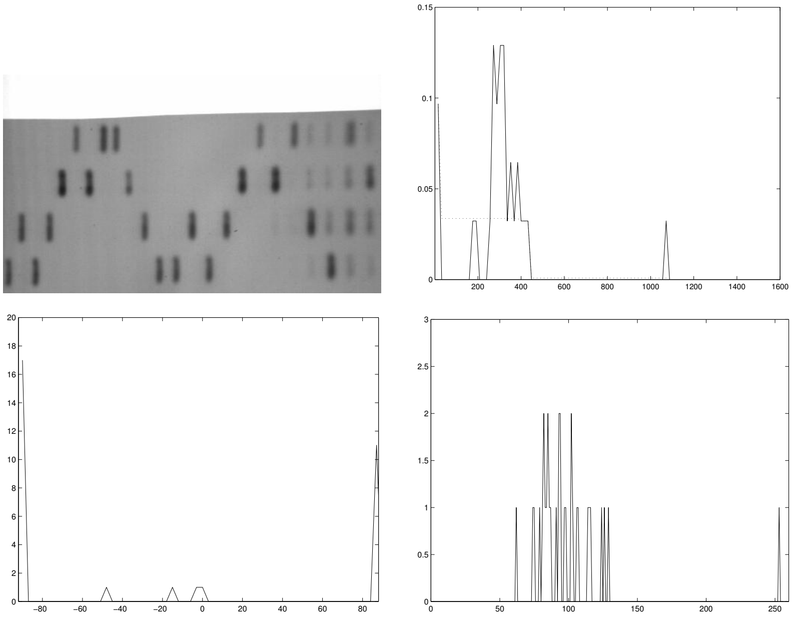

Fig. 9. Collaboration of gestalts : the objects tend to be grouped similarly by several different partial gestalts. Left : histogram of areas of the meaningful blobs. There is a unique maximal mode (256-416). The outliers are the double blob, the white background region and the three tiny blobs. Middle : histogram of orientations of the meaningful blobs (computed as the principal axis of each blob). There is a single maximal meaningful mode (interval). This mode is the interval 85-95. It contains 28 objects out of 32. The outliers are the white background region and three tiny spots. Right : histogram of the mean grey levels inside each block. There is a single maximal mode containing 30 objects out of 32, in the grey level interval 74-130. The outliers are the background white region and the darkest spot.

favour of this hope. First of all, gestaltists pointed out the relatively small number of relevant gestalt qualities for biological vision. Now, we have shown in this paper that many of them (and probably all) can be computed by the Helmholtz principle followed by a maximality argument. Second, the byzantine discussions of gestaltists about "conflicts of gestalts", so vividly explained in the books of Kanizsa, might well be solved by a few information theoretical principles. As a good example of it, let us mention how the dilemna alignment-versus-parallelism can be solved by an easy minimal description length principle (MDL) [20], [5]. Figure 10 shows the problem and its simple solution. On the middle, we see all detected alignments in the Brown image on the left. Clearly, those alignmnents make sense but many of them are slanted. The main reason is this : all straight edges are in fact blurry and therefore constitute a rectangular region where all points have roughly the same direction. Thus, since alignment detection is made up to some accuracy, the straight alignments are mixed up with slanted alignments. These slanted alignments are easily removed by the application of a MDL principle : we retain for each point only the most meaningful alignment to which it belongs. We then compute again the remaining maximal meaningful alignments and the result (right) shows that the conflict between parallelism and alignment has been solved. Clearly, information theoretical rules of this kind may be applied in a general framework and put order in the proliferation of "partial gestalts". Let us mention a good attempt of this kind in [14], where the author proposed a MDL reformulation of segmentation variational methods ([19])

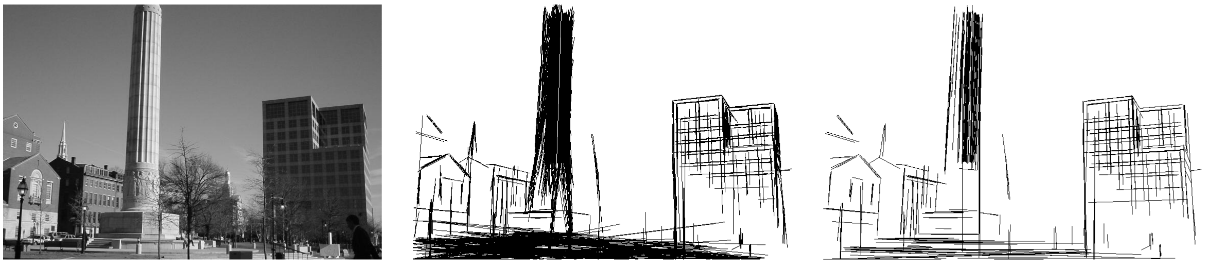

Fig. 10. Parallelism against alignment. Left, original Brown image. Middle : maximal meaningful alignments. Here, since many parallel alignments are present, secondary, parasite slanted alignments are also found. Right : Minimal description length of alignments, which eliminates the spurious alignments. This last method outlines a solution to conflicts between partial gestalts.

References

[1] A. Almansa, A. Desolneux and S. Vamech, Vanishing points are meaningful Gestalts. This same special issue (submitted), december (2001).

[2] Caselles V., B. Coll and J.-M. Morel, A Kanizsa Programme. *Progress in Nonlinear Differential Equations and their Applications* 25, pp. 35-55, 1996.

[3] A. Desolneux, L. Moisan et J.M. Morel, Meaningful Alignments, *International Journal of Computer Vision*, Vol.40(1), pp.7-23, october 2000.

[4] A. Desolneux, L. Moisan et J.M. Morel, Edge Detection by Helmholtz Principle, to appear in *Journal of Mathematical Imaging and Vision*, Vol. 14(3), pp. 271-284, may 2001.

[5] A. Desolneux, L. Moisan et J.M. Morel, Maximal Meaningful Events and Applications to Image Analysis, preprint, CMLA, `http://www.cmla.ens-cachan.fr/Cmla/Publications/2000` soumis en juillet 2000 à *Annals of Statistics*.

[6] A. Desolneux, L. Moisan et J.M. Morel, In preparation.

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

[8] Y. Gousseau, The size of objects in natural images PhD Dissertation, Université Paris-Dauphine, 2000.

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

[10] Huang, G. and D. Mumford, Statistics of Natural Images and Models. *Comp. Vision and Pattern Recognition*, 1999.

[11] Kanizsa, G. *La Grammaire du Voir.* Editions Diderot, arts et sciences, 1994.

[12] Kass, M., A. Witkin and D. Terzopoulos. Snakes: active contour models. *1st Int. Comp. Vision Conf.* IEEE 777, 1987.

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

[14] Leclerc, Y. Constructing Simple Stable Descriptions for Image Partitioning. *Int. J. of Comp. Vision* , 3, pp. 73-102, 1989.

[15] Lowe, D. *Perceptual Organization and Visual Recognition.* Kluwer Academic Publishers, 1985.

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

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

[18] U. Montanari, *On the Optimal Detection of Curves in Noisy Pictures*, CACM, 14, 1971, n 5, May, pp. 335–345,

[19] Mumford, D. and J. Shah. Boundary detection by minimizing functionals *IEEE Conf. on Comp. Vision and Pattern Recognition*, San Francisco, 1985.

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

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

[22] 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.

[23] C. V. Stewart "MINPRAN : a new robust estimator for Computer Vision", *I.E.E.E. Trans. on Pattern Analysis and Machine Intelligence* 17, pp. 925-938, 1995.

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