A Probabilistic Grouping Principle to Go from Pixels to Visual Structures

Agnès Desolneux

MAP5 (UMR CNRS 8145), University Paris Descartes,
45 rue des Saints-Pères, 75006 Paris, France

Abstract. We will describe here how the Helmholtz principle, which is a principle of visual perception, can be translated into a computational tool that can be used for many problems of discrete image analysis. The Helmholtz principle can be formulated as "we immediately perceive whatever has a low likelihood of resulting from accidental arrangement". To translate this principle into a computational tool, we will introduce a variable called NFA (Number of False Alarms) associated to any geometric event in an image. The NFA of an event is defined as the expectation of the number of occurrences of this event in a pure noise image of same size. Meaningful events will then be events with a very low NFA. We will see how this notion can be efficiently used in many detection problems (alignments, smooth curves, edges, etc.). The common framework of these detection problems is that they can all be translated into the question of knowing whether a given group of pixels is meaningful or not. This is a joint work with Lionel Moisan and Jean-Michel Morel.

Keywords: grouping laws, Gestalt theory, Helmholtz principle, rare events, alignments, edge detection, segmentation.

1 Introduction

When one looks at an image, one usually can see in it many geometric structures (straight segments, curves, homogeneous regions, etc.). But these objects are not really present in the image, they are only the result of our visual perception that is able to group pixels together according to some geometric criteria. Now, how can this grouping phenomenon be translated into a mathematical and computational principle in order to make a computer "see" geometric structures in an image? This can be achieved by formalizing and using the so-called *Helmholtz principle*, that we explain now.

1.1 Helmholtz Principle

The Helmholtz principle is a principle of visual perception that can be formulated two ways:

1. The first way is common sensical. It simply states that *"we do not perceive any structure in a uniform random image"*. In this form, the principle was first stated by Attneave in 1954 [1].

I. Debled-Rennesson et al. (Eds.): DGCI 2011, LNCS 6607, pp. 1–12, 2011.

2. In its stronger form, the Helmholtz principle states that whenever some large deviation from randomness occurs, a structure is perceived. In other words: *“we immediately perceive whatever has a low likelihood of resulting from accidental arrangement”*. It has been first stated under this form in Computer Vision by S.-C. Zhu [2] and D. Lowe [3].

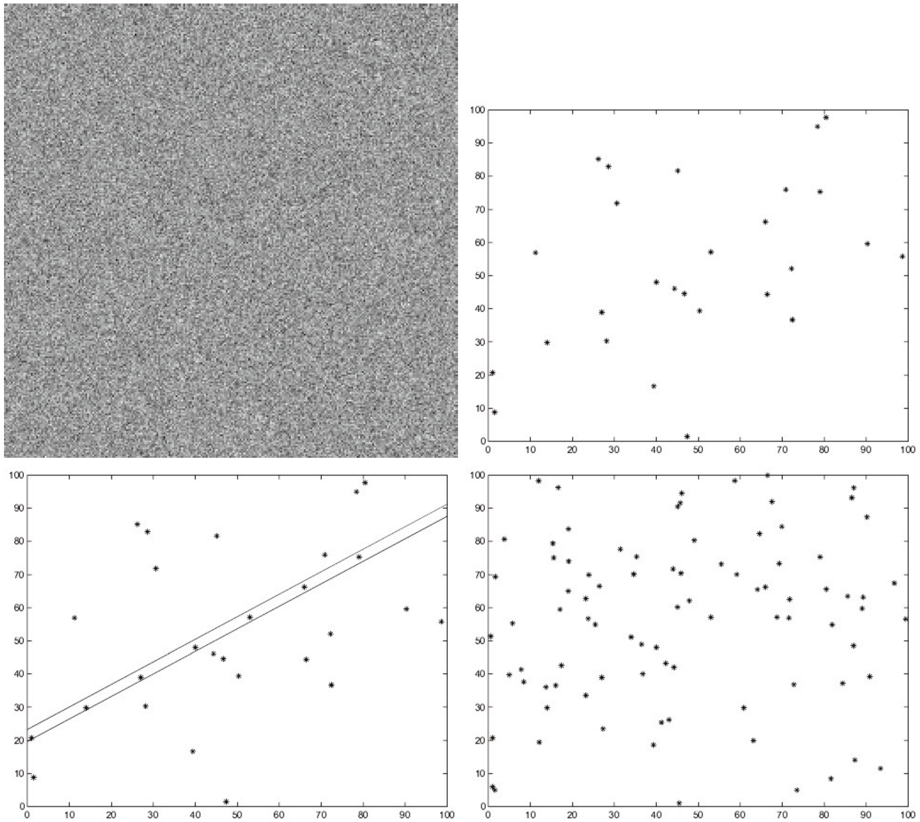


Fig. 1. Top left: an image of pure noise (pixels have independant identically distributed grey levels following the uniform distribution on $\{0, 1, \ldots, 255\}$). In this image, no visual structure is perceived. Top right: an image made of 27 points, such that 7 of them are aligned and the 20 others are randomly positioned. The alignment is immediatly perceived and will be detected by the *a contrario* method we will present (result on the bottom left figure). Bottom right: when there are 80 random points instead of 20, the fact to have 7 points aligned is not a rare event anymore, and we don’t perceive it, even if it is there.

Given a geometric structure, Helmholtz principle tells us that we immediately perceive it if it has a low probability of resulting from randomness. But what are the “interesting” structures for our visual perception? This question (among others) has been studied by the Gestalt School of Psychophysiology. We briefly recall part of their work in the following section.

1.2 Gestalt Theory

Before the development of Gestalt Theory, there were many psychophysiological experiments based on optic-geometric illusions (see for instance the left part of Figure 2). The aim of these illusions is to ask: “what is the reliability of our visual percpetion?” But Gestalt theory (developed by Wertheimer, Metzger, Kanizsa - see [4] for instance) does not continue on the same line. The question is not why we sometimes see a distorted line when it is straight; the question is why we do see a line at all. This perceived line is the result of a construction process. Gestalt theory starts with the assumption that there is a small list of active grouping laws in visual perception: vicinity, same attribute (like colour, shape, size or orientation), aligment, good continuation, symmetry, parallelism, convexity, closure, constant width, amodal completion, T-junctions, X-junctions, Y-junctions. Moreover, all grouping Gestalt laws are *recursive*: they can be applied first to atomic inputs and then in the same way to partial Gestalts already constituted. This is illustrated by the right part of Figure 2.

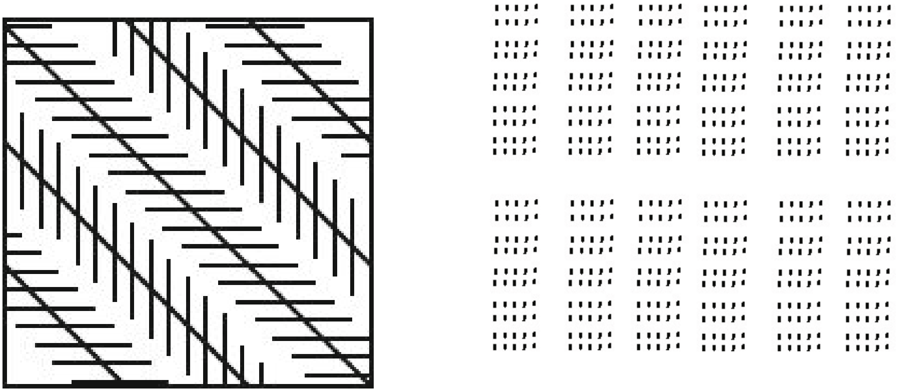


Fig. 2. Left: Zoellner’s Illusion (1860). Right: the same Gestalt grouping laws namely alignment, parallelism, constant width and proximity, are recursively applied not less than six times.

2 A Computational Tool: The Number of False Alarms

Having now a grouping principle (Helmholtz principle) and visually relevant structures in images (the ones that obey Gestalt grouping laws), we can combine this into a general framework which is the one of the so-called *a contrario* methodology. The general description of this methodology is the following. Given n geometric objects $O_1, \dots O_n$, let X_i be a variable describing an attribute (position, colour, orientation, size, etc...) of O_i. Then define a Null Hypothesis $\mathcal{H}_0$ (that is a noise model also called *a contrario* model): $X_1, \dots, X_n$ are independant identically distributed. Now, consider an observed geometric event E concerning k of the objects (for instance if X_i are spatial positions we observe $X_1, \dots, X_k$ are very close). The question is: can this observed geometric event E happen by chance? In other words, what is its likelihood under $\mathcal{H}_0$? To answer this, we define a computational tool called *Number of False Alarms* that is defined as the expected number of occurrences of the event E under $\mathcal{H}_0$:

$$\mathrm{NFA}(E) := \mathbb{E}_{\mathcal{H}_0}[\text{number of occurrences of the observed event}].$$

If the statistical test "NFA$(E) \leqslant \varepsilon$" (for ε fixed and small, meaning $\varepsilon \leqslant 1$) is positive then E is said to be an *ε-meaningful event.* When $\varepsilon = 1$, we simply talk about *meaningful event.*

We will see in the following through many different exemples that the NFA defined above is a universal variable adaptable to many detection problems. A detailed presentation of the whole *a contrario* methodology and its applications can be found in the book [5].

3 Examples

3.1 Alignments in an Image

The first example of the *a contrario* method we give here is the detection of alignments in a discrete image [6]. It corresponds to the grouping of pixels according to the parallelism of their orientation. Let us detail this.

Given a discrete image of size $N \times N$ pixels, at each pixel, we can compute an orientation ($\perp$ to the gradient). The noise model $\mathcal{H}_0$ is defined by: pixels at distance $\geqslant 2$ have i.i.d. orientations, uniformly distributed on $[0, 2\pi)$.

Definition 1 (Meaningful segment). *Let S be a sequence of l consecutive and aligned pixels such that k of them have their orientation aligned with the one of the segment at a given precision p (see Figure 3). We say that the segment S is ε-meaningful if:*

$$\mathrm{NFA}(S) = N^4 \times \mathcal{B}(l, k, p) = N^4 \sum_{j=k}^{l} \binom{l}{j} p^j (1-p)^{l-j} \leqslant \varepsilon.$$

Let us explain the formula for the NFA. The first term is the number of tests that we make: it is thus the total number of discrete straight segment in the image, that is $\simeq N^4$. The second term is the binomial tail: it is the probability that, under the noise model, at least k pixels among l pixels have their orientation aligned with the orientation of the segment according to the precision p.

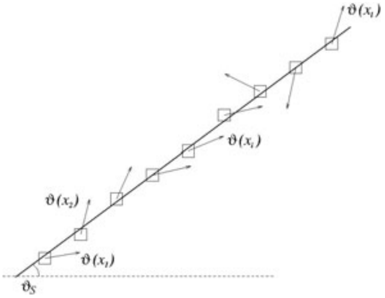


Fig. 3. A discrete straight segment made of l pixels taken at distance 2. At each pixel x_i, there is an orientation $\theta(x_i)$ and it is said to be aligned with the orientation θ_S of the segment according to precision p when $|\theta(x_i) - \theta_S| \leqslant p\pi$.

The main property is then that the expected number of ε-meaningful segments in a pure noise image of size $N \times N$ pixels is less than ε.

Now, when a segment is very meaningful (meaning that NFA(S) $<< \varepsilon$), then many segments it contains, or is contained in, are also meaningful. We thus need a kind of "selection rule", or "minimal representation rule" to keep only the "best representatives". Thanks to the NFA, we can compare segments, and we then decide to look only at segments which are local minima of the NFA, in the sense of the following definition.

Definition 2 (Maximal Meaningful segment). *A segment S is maximal meaningful if it is meaningful and if* $\forall S' \subset S$ *(resp.* $S' \supset S$*), then* $\mathrm{NFA}(S') \geqslant \mathrm{NFA}(S)$ *(resp.* $\mathrm{NFA}(S') > \mathrm{NFA}(S)$*).*

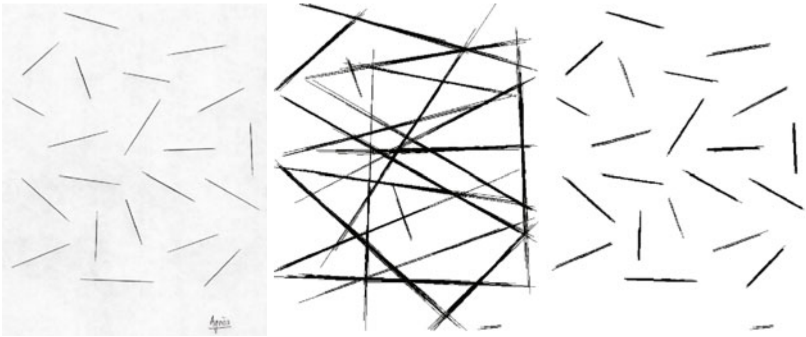


Fig. 4. From left to right: the original image, all meaningful segments, maximal meaningful segments

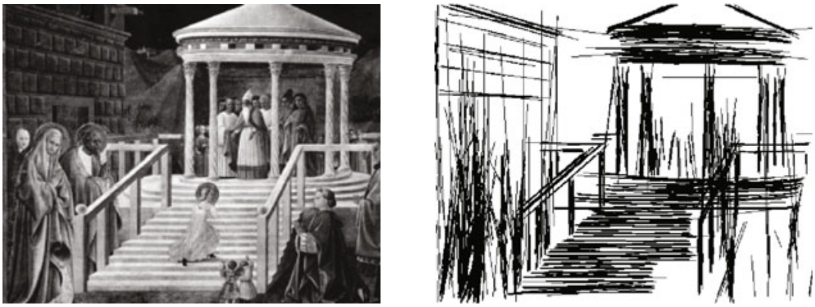


Fig. 5. An image (scan from a painting by Uccello), and its maximal meaningful alignments

3.2 Meaningful Boundaries

A second application of the *a contrario* method is the very general and very studied problem of "edge detection" in an image. The aim is to divide the image into "homogeneous" regions (this is the problem of Image Segmentation in Computer Vision). And the dual approach consists in finding the boundaries of these regions - as "highly" contrasted curves (this is the problem of edge detection).

As for the meaningful alignment, we first need to define the structures we look for and also the noise model. At each pixel x of an image u of size $N \times N$, we start by defining the contrast $c(x)$ by $c(x) = |\nabla u|(x)$. And then, the noise model is simply given by the empirical distribution of the contrast in the image with the additional independance hypothesis, which means that the probability that a pixel has a contrast larger than μ is given by

$$H(\mu) = \mathbb{P}\left(c(x) \geqslant \mu\right) = \frac{1}{N^2} \#\{y / |\nabla u|(y) \geqslant \mu\},$$

and pixels at distance larger than 2 are assumed (in the noise model) to have independant contrasts.

Finally, what are the candidate to be edge curves in the image ? It is not possible to look at all the curves in the image, and good candidates are the level lines of the image (defined as the boundaries of the level sets). Let thus N_{ll} be the number of level lines in the image. We then have the following definition.

Definition 3 (Meaningful boundaries). *Let $\mathcal{C}$ be a level line of the image, with length l and minimal contrast μ. We say that $\mathcal{C}$ is an ε-meaningful boundary if*

$$\mathrm{NFA}(\mathcal{C}) = N_{ll} \times H(\mu)^l \leqslant \varepsilon.$$

For the same reasons as the ones explained for meaningful alignments, we also need here a notion of maximality. We can then define maximal meaningful boundaries as local minima of the NFA for the relation of inclusion of level sets (see details in [7]). Also the same method can be applied to pieces of level lines (instead of whole level lines), and we then obtain *meaningful edges.*

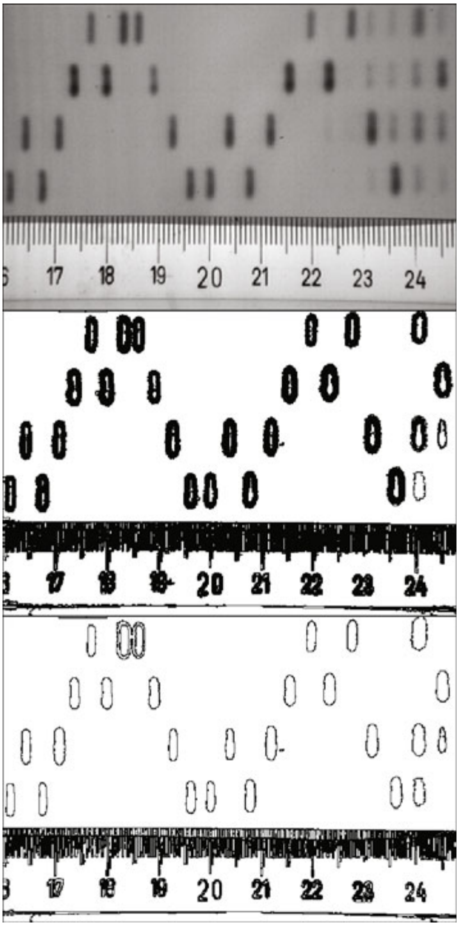


Fig. 6. Top: original image. Middle: all meaningful boundaries. Bottom: maximal meaningful boundaries.

3.3 Meaningful Good Continuations

Instead of looking at the contrast accross a level line, one can look at its regularity and see if it is more regular than "what we would expect in pure noise". This has been formalized by F. Cao in [8], and to keep the Gestalt theory term, such curves are called "good continuations". Thus, the aim is to look for meaningful "smooth" curves, without any contrast information.

Let $\Gamma = (p_0, \ldots, p_{l+1})$ be a discrete curve of length l, and let θ be its maximal discrete curvature (see also Figure 7):

$$\theta = \max_{1 \leqslant i \leqslant l} |\text{angle}(p_{i+1} - p_i, p_i - p_{i-1})|.$$

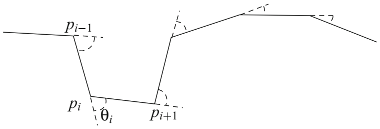


Fig. 7. A discrete curve is a sequence $(p_0, \ldots, p_{l+1})$ of points, and its maximal discrete curvature can be defined as $\theta = \max_{1 \leqslant i \leqslant l} |\theta_i|$

The noise model here is that the angles are i.i.d. uniform on $[0, 2\pi)$, *i.e.* the curve is a “random walk”. Let N_c be the number of considered curves (usually the number of pieces of level lines in the image).

Definition 4 (meaningful good-continuation). *We say that a discrete curve $\mathit{\Gamma}$ is an ε-meaningful good-continuation if*

$$\theta < \frac{\pi}{2} \quad \textit{and} \quad \mathrm{NFA}(\mathit{\Gamma}) = N_c \left(\frac{\theta}{\pi}\right)^l < \varepsilon.$$

Again, we have a definition of maximality: a meaningful good-continuation $\mathit{\Gamma}$ is maximal meaningful if: $\forall \mathit{\Gamma}' \subset \mathit{\Gamma}$, $\mathrm{NFA}(\mathit{\Gamma}') \geqslant \mathrm{NFA}(\mathit{\Gamma})$ and $\forall \mathit{\Gamma}' \supsetneq \mathit{\Gamma}$, $\mathrm{NFA}(\mathit{\Gamma}') > \mathrm{NFA}(\mathit{\Gamma})$. And interesting property is then that: if $\mathit{\Gamma}$ and $\mathit{\Gamma}'$ are two maximal meaningful good-continuations on the same level line, then $\mathit{\Gamma} \cap \mathit{\Gamma}' = \emptyset$.

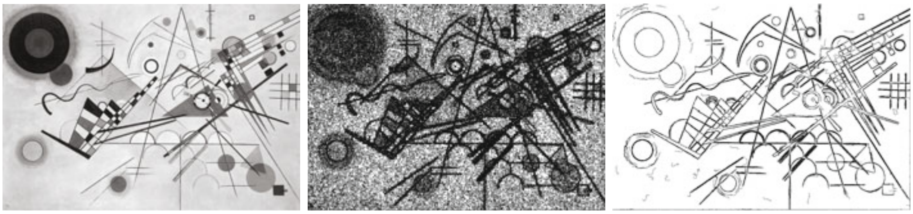


Fig. 8. From left to right: the original Image (painting by Kandinsky), all the level lines (with some quantization step for grey levels), maximal meaningful good-continuations

3.4 Similarity of a Scalar Attribute

Another application of the *a contrario* methodology is the grouping of objects according to any scalar attibute (like grey level, or orientation, etc.) More precisely, assume we have M “objects”, and each of them as an attribute $q \in \{1, 2, \ldots, L\}$. Let $a \leqslant b$ be two attibute values and let G be the group of objects (among the M objects) such that their scalar attribute q satisfies $a \leqslant q \leqslant b$. Then denote k the cardinality of G and define its NFA under the noise model that attibute values are i.i.d. uniform, by

$$\mathrm{NFA}(G) = \mathrm{NFA}([a, b]) = \frac{L(L+1)}{2} \cdot \mathcal{B}\left(M, k, \frac{b-a+1}{L}\right).$$

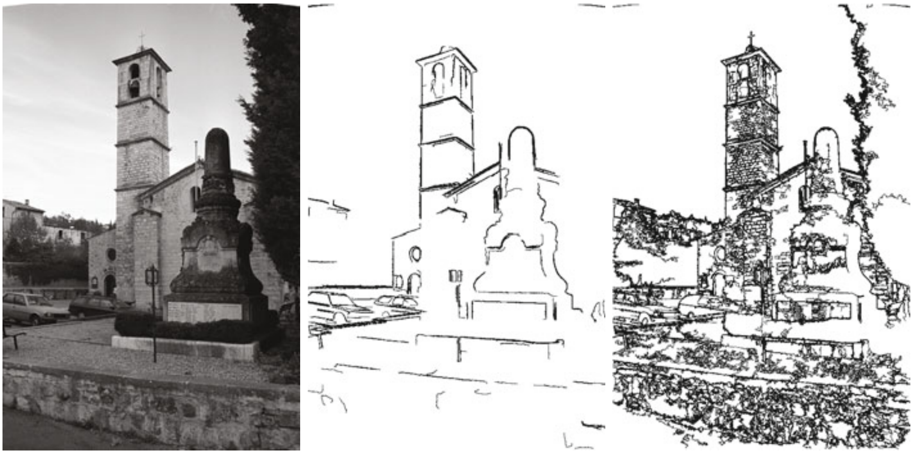
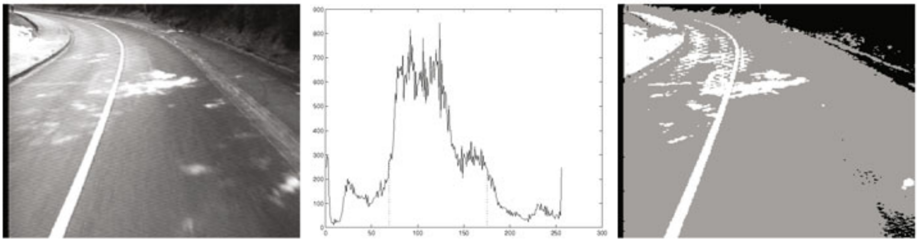


Fig. 9. From left to right: image (INRIA) of the church of Valbonne, maximal meaningful good-continuations, maximal meaningful boundaries

(In this formula, notice that $L(L+1)/2$ corresponds to the total number of possible groups that can made this way, and $\mathcal{B}$ denotes again, as in the case of meaningful alignments, the tail of the binomial distribution). The group G (or, in an equivalent way, the interval $[a, b]$) is said ε-meaningful when $\mathrm{NFA}(G) \leqslant \varepsilon$. And again, using the relation of inclusion of intervals, we can define maximal meaningful groups (intervals).

A first application is the study of an image grey level histogram. Indeed, looking for the maximal meaningful intervals of this histogram is a way to obtain an automatic grey level quantization of the image. See an illustration of this on Figure 10.



(a) (b) (c)

Fig. 10. Maximal meaningful intervals and optimal grey level quantization of a digital image. (a) original image. (b) histogram of grey-levels of the image with its single maximal meaningful interval $[69, 175]$ (between the dotted lines). (c) Quantized image: black points represent points in the original image with grey level in $[0, 68]$, grey points represent points with grey level in the maximal meaningful interval $[69, 175]$ and white points represent points with grey-level larger than 176.

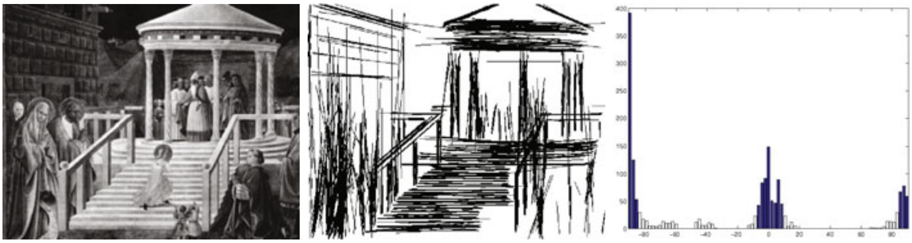


Fig. 11. Uccello's painting: maximal meaningful alignments and histogram of orientations. Two maximal meaningful intervals are found in this histogram corresponding respectively to the horizontal and vertical segments.

Instead of using a uniform assumption for the noise model distribution of attributes, we can more generally use any distribution or class of distributions. For instance, we can define meaningful groups according to an attribute that is assumed to have a decreasing distribution (like the length, the area, etc.) by setting for the Number of False Alarm of an interval $[a, b]$:

$$\mathrm{NFA}([a,b]) = \frac{L(L+1)}{2} \cdot \max_{p \in \mathcal{D}} \mathcal{B}\left(M, k, \sum_{i=a}^{b} p(i)\right)$$

where k is the number of objects having their attribute value in $[a, b]$ (i.e. it is the cardinality of G) and $\mathcal{D}$ is the set of decreasing distributions on $\{1, ..., L\}$.

An example of application is the recursivity of grouping as formulated by the Gestalt theory. See Figures 11 and 12 for some illustrations and comments of this.

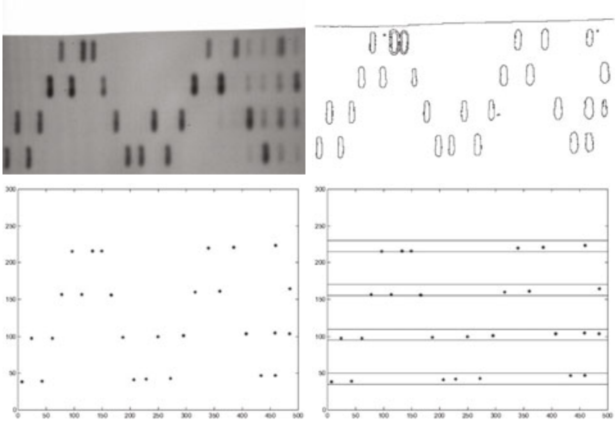


Fig. 12. Gestalt grouping principles at work for building an "order 3" Gestaltist description of the image (alignment of blobs of the same size). First row: original DNA image (left) and its maximal meaningful boundaries (right). Second row: left, barycenters of all meaningful regions whose area is inside the only maximal meaningful interval of the histogram of areas; right, meaningful alignments of these points.

4 Conclusion

Through many illustrations, we have seen that Helmholtz principle combined with Gestalt grouping laws can become, through the definition of a Number of False Alarms, a powerful computational tool. It can then be used in many applications in detection problems, but also in shape recognition (Musé et al. [9]); image matching (Rabin et al. [10]); epipolar geometry (Moisan and Stival [11]), motion detection and analysis (Veit et al. [12]); clustering (Cao et al. [13]); stereovision (Sabater et al. [14]); image denoising (by grain filters, Coupier et al. [15]); etc. The main advantage of the whole *a contrario* methodology is that, thanks to the framework of statistical testing, it provides a validation of found structures and also a way to set the different thesholds of a given problem in an automatic way. An interesting research direction is then the comparison of these predicted thresholds with the ones of our visual perception. Also from a mathematical point of view, the *a contrario* methodology raises many difficult questions of stochastic geometry.

Now, even if the *a contrario* method is very general and has many applications, some open questions still remain. A first question is: how to deal with the "overdetermination" of images (i.e. the fact that visual objects usually have several qualities at the same time)? This would require the definition of a generalized computational tool like the NFA that would be able to deal with several grouping criteria at the same time. A second open question is: how to deal with "conflicts" of qualities? Since there is no "universal hierarchy" of qualities, we cannot hope for any reliable explanation of a figure by summing up the results of one or several partial gestalts detector. Only a global synthesis, treating all conflicts of partial gestalts, can give the correct result. This would require the need of another framework than Helmholtz principle (like for instance the Minimum Description Length principle or Bayesian compositional approaches [16], [17]).

References

1. Attneave, F.: Some informational aspects of visual perception. Psychological Review 61, 183–193 (1954)
2. Zhu, S.C.: Embedding Gestalt Laws in Markov Random Fields. IEEE Trans. on Pattern Analysis and Machine Intelligence 21(11), 1170–1187 (1999)
3. Lowe, D.: Perceptual Organization and Visual Recognition. Kluwer Academic Publishers, Dordrecht (1985)
4. Kanizsa, G.: Grammatica del Vedere / La Grammaire du Voir. Il Mulino, Bologna / Éditions Diderot, arts et sciences (1980/1997)
5. Desolneux, A., Moisan, L., Morel, J.-M.: From Gestalt Theory to Image Analysis: A Probabilistic Approach. Springer, Heidelberg (2008)
6. Desolneux, A., Moisan, L., Morel, J.-M.: Meaningful Alignments. Int. Journal of Computer Vision 40(1), 7–23 (2000)
7. Desolneux, A., Moisan, L., Morel, J.-M.: Edge Detection by Helmholtz Principle. Journal of Mathematical Imaging and Vision 14(3), 271–284 (2001)
8. Cao, F.: Good continuation in digital images. In: Int. Conf. Computer Vision (ICCV), vol. 1, pp. 440–447 (2003)

9. Musé, P., Sur, F., Cao, F., Gousseau, Y.: Unsupervised thresholds for shape matching. In: Int. Conf. on Image Processing (ICIP 2003), vol. 2, pp. 647–650 (2003)
10. Rabin, J., Delon, J., Gousseau, Y.: A statistical approach to the matching of local features. SIAM Journal on Imaging Sciences 2(3), 931–958 (2009)
11. Moisan, L., Stival, B.: A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. International Journal of Computer Vision 57(3), 201–218 (2004)
12. Veit, T., Cao, F., Bouthemy, P.: An a contrario decision framework for region-based motion detection. International Journal of Computer Vision 68(2), 163–178 (2006)
13. Cao, F., Delon, J., Desolneux, A., Musé, P., Sur, F.: A unified framework for detecting groups and application to shape recognition. Journal of Mathematical Imaging and Vision 27(2), 91–119 (2007)
14. Sabater, N., Almansa, A., Morel, J.-M.: Rejecting wrong matches in stereovision. Technical report CMLA, ENS Cachan, No. 2008-28 (2008)
15. Coupier, D., Desolneux, A., Ycart, B.: Image denoising by statistical area thresholding. Journal of Mathematical Imaging and Vision 22(2-3), 183–197 (2005)
16. Bienenstock, E., Geman, S., Potter, D.: Compositionality, MDL Priors, and Object Recognition. In: Mozer, M.C., Jordan, M.I., Petsche, T. (eds.) Advances in Neural Information Processing Systems, vol. 9, pp. 838–844. MIT Press, Cambridge (1997)
17. Zhu, S.C., Mumford, D.: A Stochastic Grammar of Images. Foundations and Trends in Computer Graphics and Vision 2(4), 259–362 (2006)