A probabilistic grouping principle to go from pixels to visual structures

Agnès Desolneux

CNRS, MAP5, Université Paris Descartes

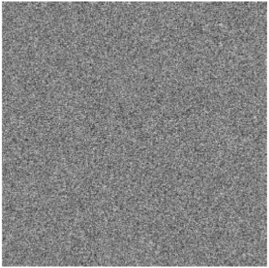
Conference DGCI 2011,
Nancy, 6-8 april 2011

Joint work with Lionel Moisan (Université Paris Descartes) and Jean-Michel Morel (Ecole Normale Supérieure de Cachan)

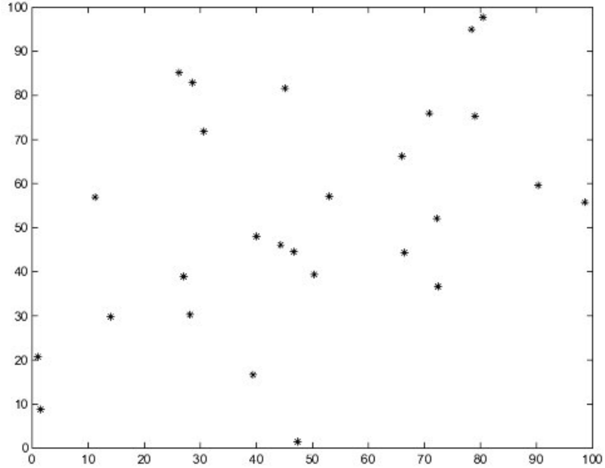
Helmholtz principle

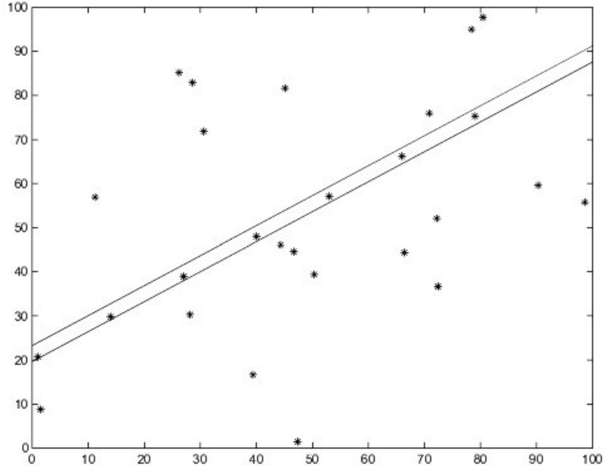
The Helmholtz principle can be formulated two ways :

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).

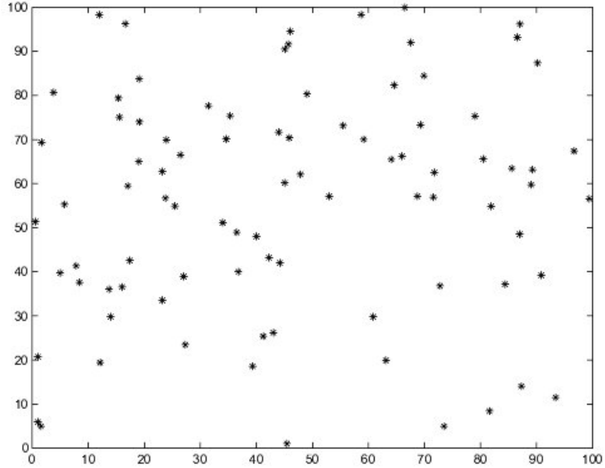


2. In its stronger form, the Helmholtz principle states that whenever some large deviation from randomness occurs, a structure is perceived. In other words : *“we immediately perceive whatever has a low likelihood of resulting from accidental arrangement”*.
   (Stated in Vision by S.-C. Zhu or D. Lowe)


100
90
80
70
60
50
40
30
20
10
0
0 10 20 30 40 50 60 70 80 90 100

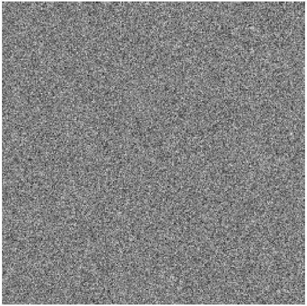


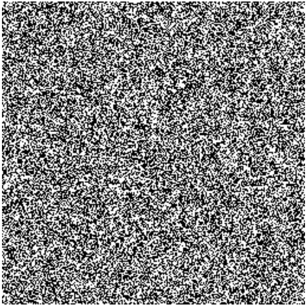
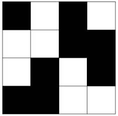
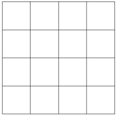
100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
80
90
100



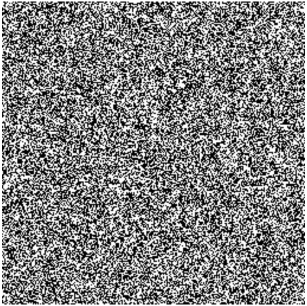
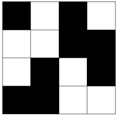
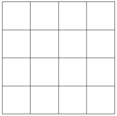
0
10
20
30
40
50
60
70
80
90
100

What's make an image an image and not noise ?



 contains   ?

- Not all possible structures are relevant for visual perception.

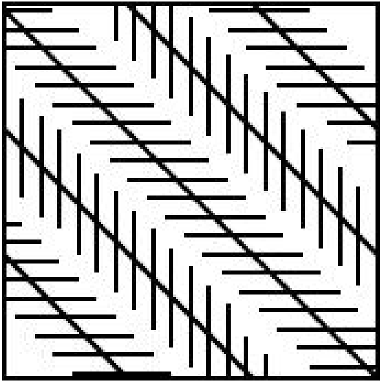
 contains   ?

- Not all possible structures are relevant for visual perception.
- The relevant structures for visual perception have been studied by the Gestalt School of Psychophysiology.

Before Gestalt Theory : optic-geometric illusions

The aim of these illusions is to ask : “what is the reliability of our visual perception ?”

An example : Zoellner’s Illusion (1860)



Gestalt Theory

- Gestalt theory does not continue on the same line. The question is not why we sometimes see non parallel line when they are parallel ; the question is why we do see a line at all. This perceived line is the result of a construction process whose laws it is the aim of Gestalt theory to establish.

Gestalt Theory

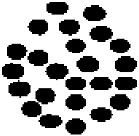
- Gestalt theory does not continue on the same line. The question is not why we sometimes see non parallel line when they are parallel ; the question is why we do see a line at all. This perceived line is the result of a construction process whose laws it is the aim of Gestalt theory to establish.
- Gestalt theory (Wertheimer, Metzger, Kanizsa) starts with the assumption of (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.

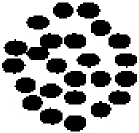
Gestalt Theory

- Gestalt theory does not continue on the same line. The question is not why we sometimes see non parallel line when they are parallel ; the question is why we do see a line at all. This perceived line is the result of a construction process whose laws it is the aim of Gestalt theory to establish.
- Gestalt theory (Wertheimer, Metzger, Kanizsa) starts with the assumption of (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.
- The above listed grouping laws belong, according to Kanizsa, to the so called primary process, opposed to a more cognitive secondary process.

Examples of Grouping Laws

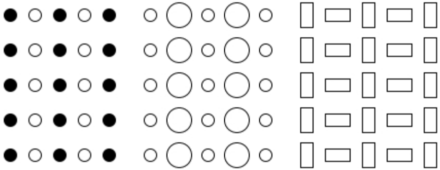
Vicinity





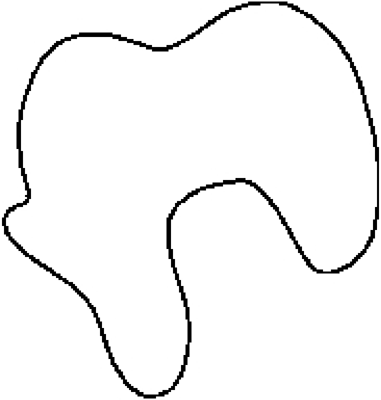
Examples of Grouping Laws

Same colour/size/orientation



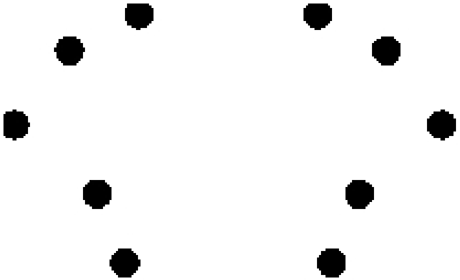
Examples of Grouping Laws

Closure



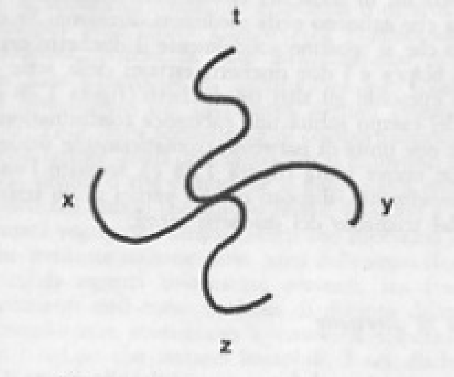
Examples of Grouping Laws

Symmetry



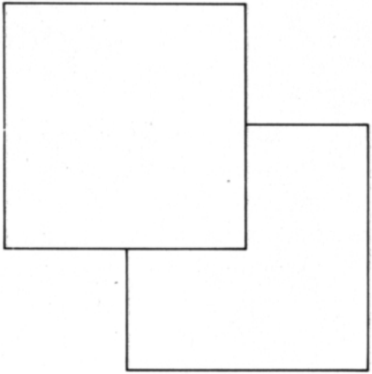
Examples of Grouping Laws

Good continuation



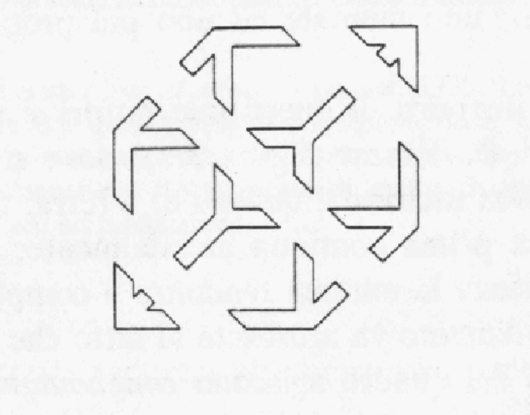
T- and X- junctions

T-Junctions



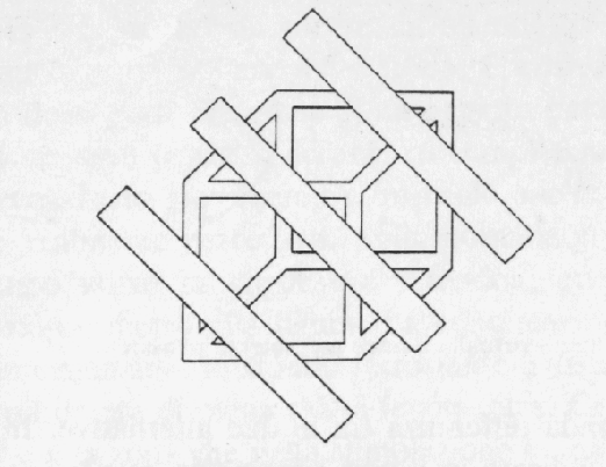
T- and X- junctions

T-Junctions



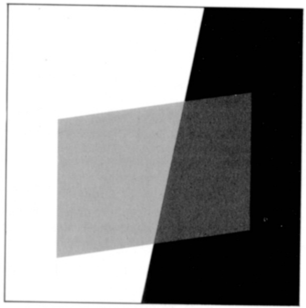
T- and X- junctions

T-Junctions

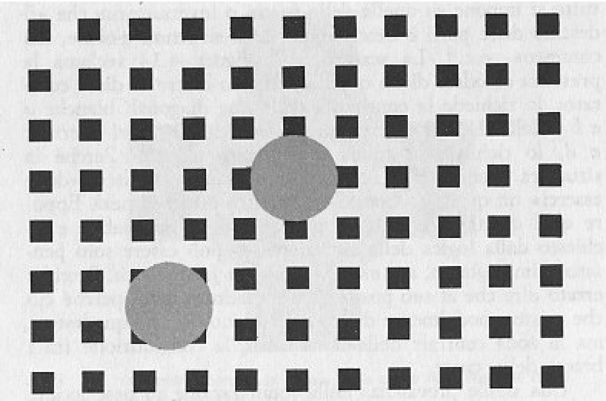


T- and X- junctions

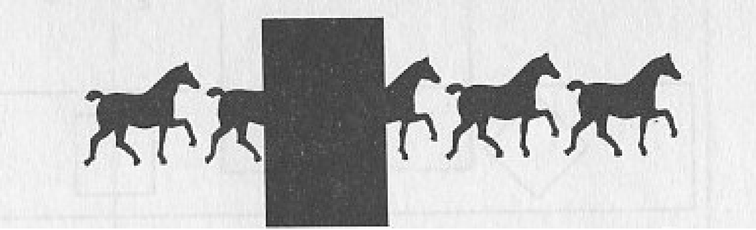
X-Junctions



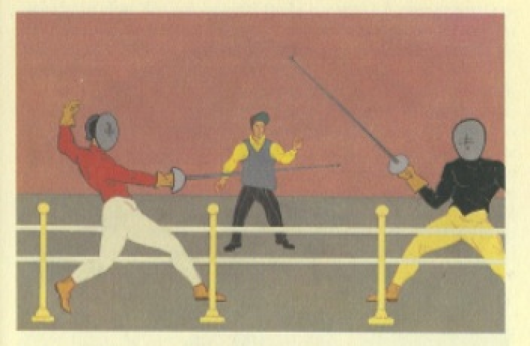
Amodal completion



Amodal completion



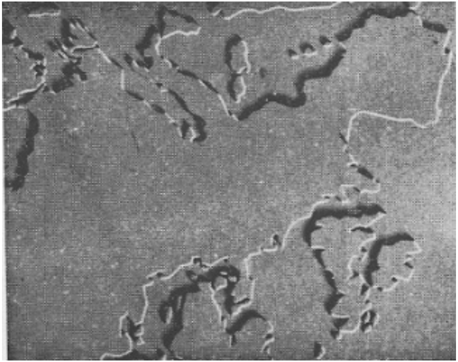
Amodal completion



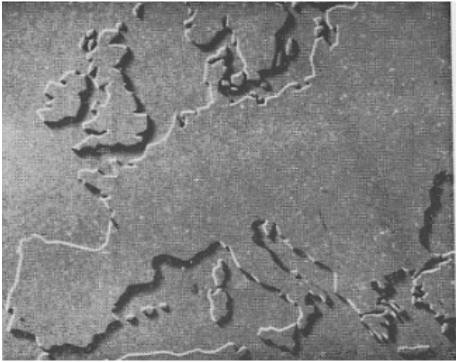
3.17 L'arbitro infilzato.

... and past experience ?

... and past experience ?

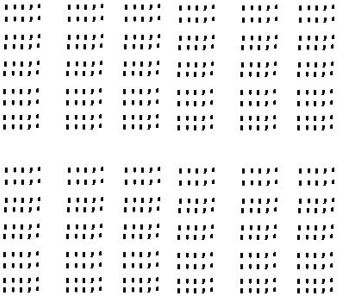


... and past experience ?



Recursivity

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.



The same partial Gestalt laws namely alignment, parallelism, constant width and proximity, are recursively applied not less than six times.

Combining Helmholtz principle and Gestalt grouping laws

General framework of *a contrario* methods :
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.

Null Hypothesis $\mathcal{H}_0$ (= noise model) : $X_1, \dots X_n$ are independant identically distributed.

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).

Can this observed geometric event E happen by chance ? (= what's its likelihood under $\mathcal{H}_0$?)

A computational tool : the Number of False Alarms

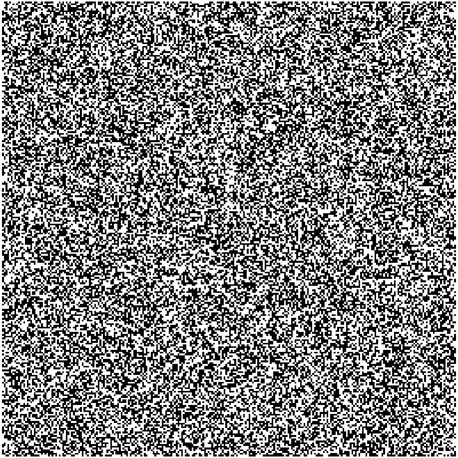
$$NFA(E) := \mathbb{E}_{\mathcal{H}_0}[\text{number of occurrences of the observed event}].$$

If the statistical test $NFA(E) \leqslant \varepsilon$ is positive then E is said to be an ε*-meaningful event*.

Rk : The *NFA* is a universal variable adaptable to many detection problems. Moreover, expectations are much easier to compute than probabilities because they add !

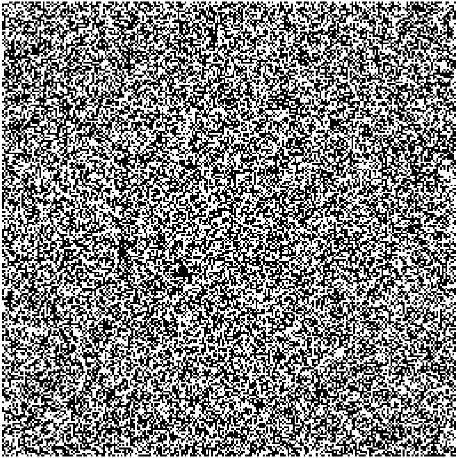
Example 1 : A small black square

$L = 256$, $p = 0.5$
$l_0 = 3$. $NFA = 125.9$.



Example 1 : A small black square

$L = 256$, $p = 0.5$
$l_0 = 5$. $NFA = 0.001$.



Alignements in an image

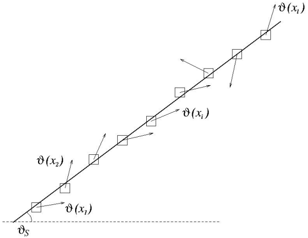
- Image of size $N \times N$ pixels.
- At each pixel, compute an orientation ($\perp$ to the gradient).

Noise model : Pixels at distance $\geqslant 2$ have i.i.d. orientations, uniformly distributed on $[0, 2\pi)$.

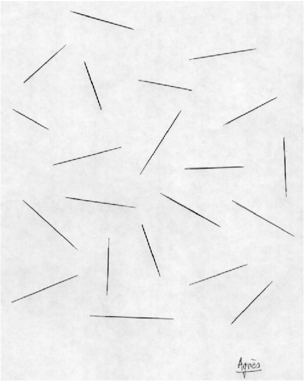
Definition (Meaningful segment)

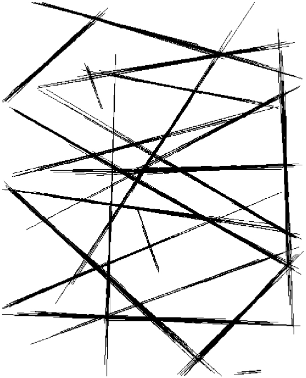
Let S be a sequence of l consecutive "aligned" pixels such that k of them have their orientation aligned with the one of the segment at a given precision p, which means $k = \sum_{i=1}^{l} \mathbb{1}_{\{|\theta(x_i) - \theta_S| \leqslant p\pi\}}$. We say that S is ε-meaningful if :

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



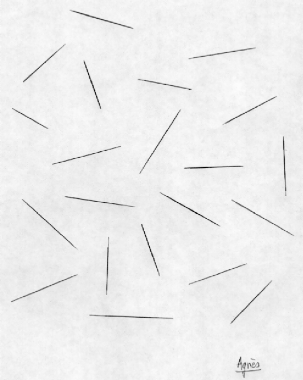
Meaningful alignments on an image

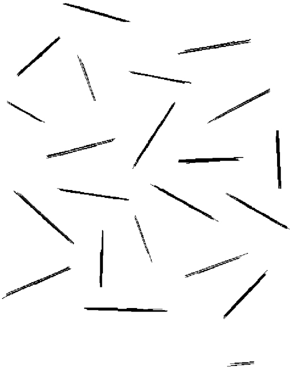




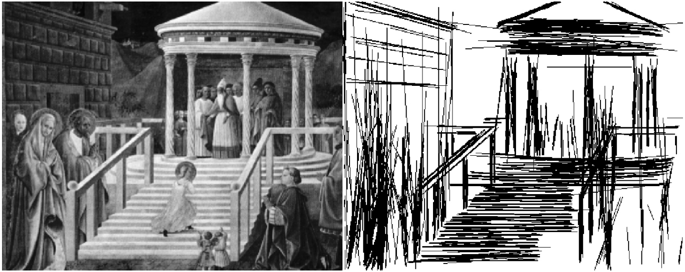
Definition (Maximal Meaningful segment)

A segment S is **maximal meaningful** *if it is meaningful and if $\forall\, S' \subset S$ (resp. $S' \supset S$), then $NFA(S') \geqslant NFA(S)$ (resp. $NFA(S') > NFA(S)$)).*





Maximal Meaningful alignments on an image



Color constancy and meaningful boundaries

Aim : Divide the image into “homogeneous” regions (= the problem of Image Segmentation in Computer Vision). Dual approach : find the boundaries of these regions - as “highly” contrasted curves (= the problem of edge detection in Computer Vision).

- At each pixel x of an image u of size $N \times N$, define the contrast by : $c(x) = |\nabla u|(x)$.
- **Noise model** : pixels have i.i.d. contrasts, distributed as the empirical distribution on the image, i.e.

$$\mathbb{P}_{\mathcal{H}_0}\left(c(x) \geqslant \mu\right) = H(\mu) = \frac{1}{N^2} \#\{y / |\nabla u|(y) \geqslant \mu\}.$$

- Let N_{ll} be the number of level lines in the image.

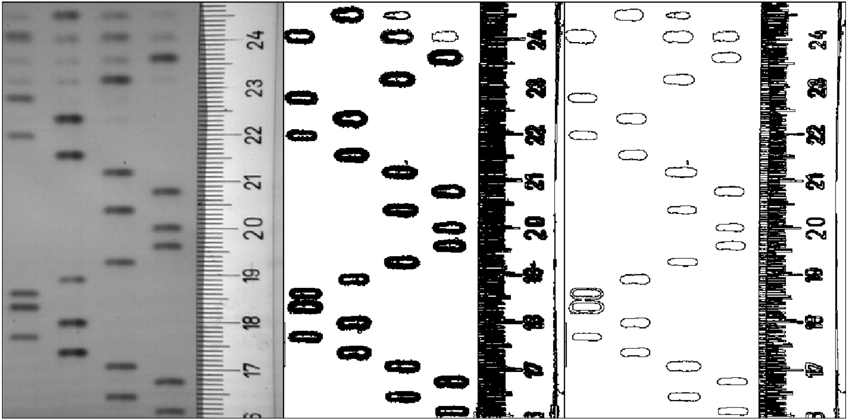
Definition (Meaningful boundaries)

*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** si*

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

Definition of **maximal meaningful boundaries** : local minima of the NFA for the relation of inclusion of level sets.

Do the same with pieces of level lines : obtain **meaningful edges**.



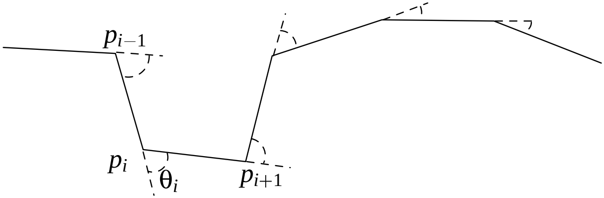
Meaningful good continuations

(From F.Cao. "Application of the Gestalt principles to the detection of good continuations and corners in image level lines", Computing and Visualization in Science, Vol 7 (1), 2004)

Aim : 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 :

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



Noise model : 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 (meaningful good-continuation)

We say that Γ *is an* ε*-meaningful good-continuation if*

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

Same definition of **maximality** : a meaningful good-continuation Γ is maximal meaningful if : $\forall \Gamma' \subset \Gamma$, $\textit{NFA}(\Gamma') \geqslant \textit{NFA}(\Gamma)$ and $\forall \Gamma' \supsetneq \Gamma$, $\textit{NFA}(\Gamma') > \textit{NFA}(\Gamma)$.
Property : if Γ and Γ' are two maximal meaningful good-continuations on the same level line, then $\Gamma \cap \Gamma' = \emptyset$.

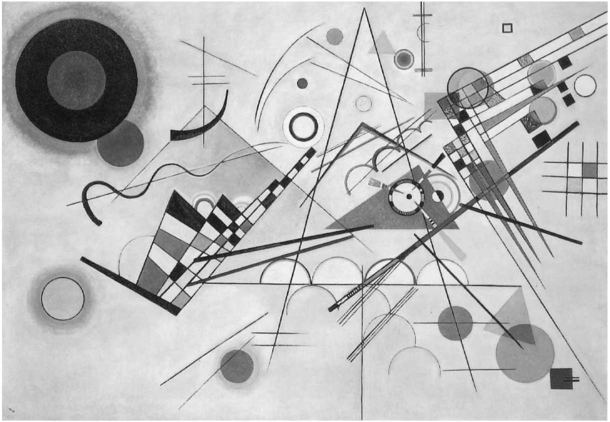


FIG.: Original Image : Kandinsky.

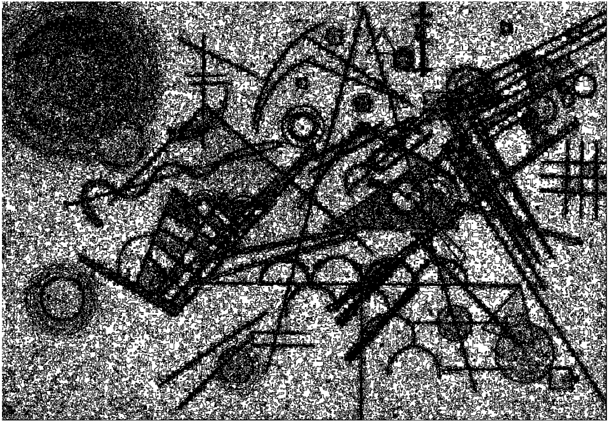


FIG.: All the level lines.

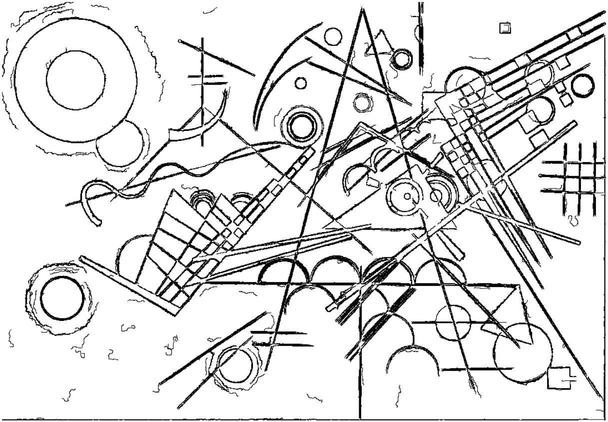


FIG.: Maximal meaningful good-continuations.

FIG.: Image (INRIA) of the church of Valbonne.

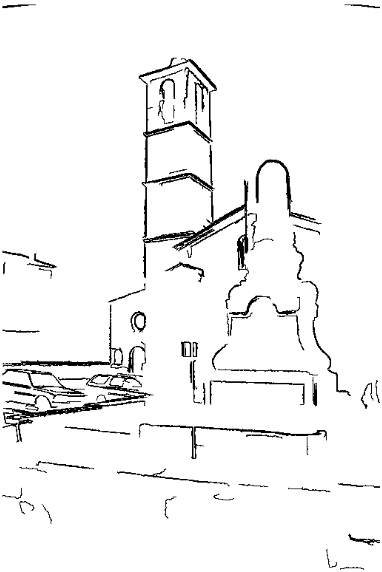


FIG.: Maximal meaningful good-continuations.

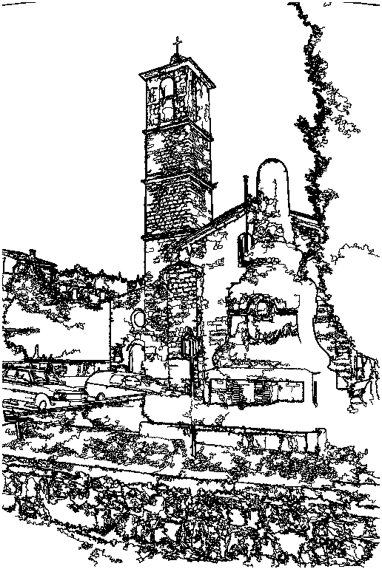


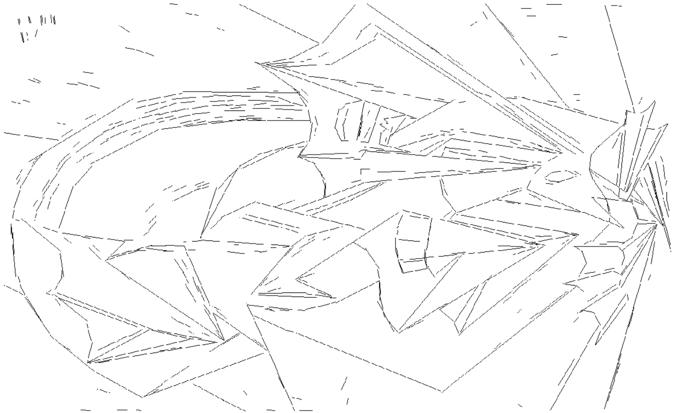
FIG.: Maximal meaningful boundaries.

Example on the same image

Original image

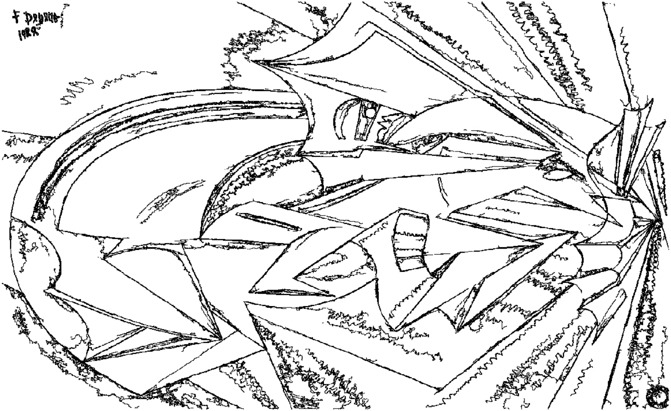
Example on the same image

Meaningful segments (Algo. LSD Grompone et. al)



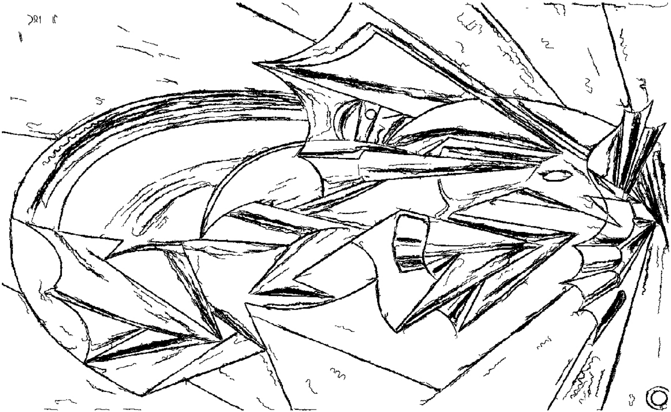
Example on the same image

Meaningful boundaries



Example on the same image

Meaningful good continuations



Similarity of a scalar attribute

For a uniform scalar attribute (gray level, orientation, etc.)
Assume we have M “objects”, and each of them has an attribute $q \in \{1, 2, \dots, L\}$. Let a group of k of them has their scalar attribute q such that $a \leqslant q \leqslant b$. Define its Number of False Alarm by

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

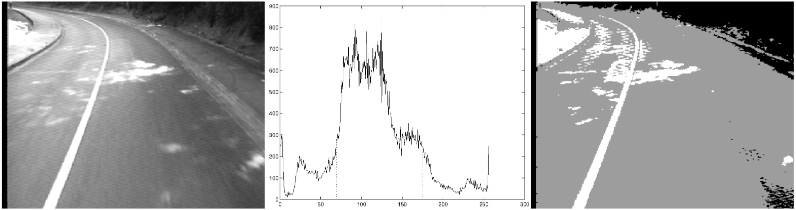
For a scalar attribute with decreasing distribution (area, length, etc.)
Define its Number of False Alarm by

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

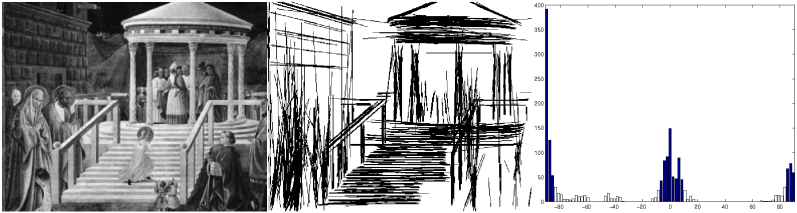
where $\mathcal{D}$: set of decreasing probability distributions on $\{1, ..., L\}$

A first example

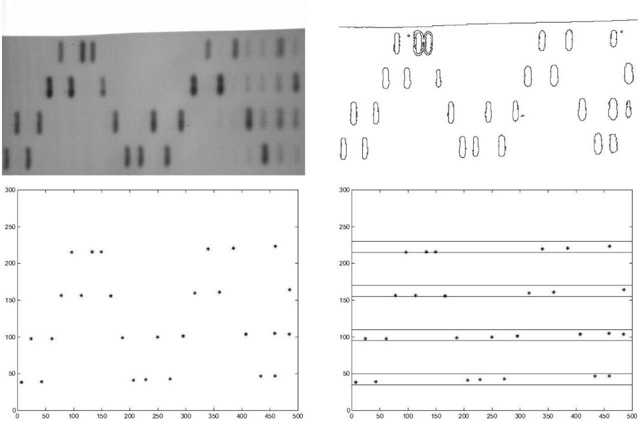
A first application is the study of an image grey-level histogram. Looking for the maximal meaningful intervals is a way to obtain an automatic gray level quantization.



A second example : recursivity



Uccello’s painting : maximal meaningful alignments and histogram of orientations. Two maximal meaningful modes are found corresponding respectively to the horizontal and vertical segments.



Gestalt grouping principles at work for building an “order 3” gestalt (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 mode of the histogram of areas ; right, meaningful alignments of these points.

Conclusion

- Helmholtz principle combined with Gestalt grouping laws can become a powerful computational tool.

Conclusion

- Helmholtz principle combined with Gestalt grouping laws can become a powerful computational tool.
- Many applications in detection problems, but also in shape recognition [Musé, Sur, Cao, Gousseau] ; image matching [Rabin, Delon, Gousseau] ; epipolar geometry [Moisan, Stival], motion detection and analysis [Cao, Veit, Bouthemy] ; clustering [Cao et al.] ; stereovision [Sabater et al.] ; image denoising (by grain filters) ; etc.

Conclusion

- Helmholtz principle combined with Gestalt grouping laws can become a powerful computational tool.
- Many applications in detection problems, but also in shape recognition [Musé, Sur, Cao, Gousseau] ; image matching [Rabin, Delon, Gousseau] ; epipolar geometry [Moisan, Stival], motion detection and analysis [Cao, Veit, Bouthemy] ; clustering [Cao et al.] ; stereovision [Sabater et al.] ; image denoising (by grain filters) ; etc.
- Automatic computation of thresholds

Conclusion

- Helmholtz principle combined with Gestalt grouping laws can become a powerful computational tool.
- Many applications in detection problems, but also in shape recognition [Musé, Sur, Cao, Gousseau] ; image matching [Rabin, Delon, Gousseau] ; epipolar geometry [Moisan, Stival], motion detection and analysis [Cao, Veit, Bouthemy] ; clustering [Cao et al.] ; stereovision [Sabater et al.] ; image denoising (by grain filters) ; etc.
- Automatic computation of thresholds
- For a mathematician : many interesting mathematical questions of stochastic geometry.

Conclusion

- Helmholtz principle combined with Gestalt grouping laws can become a powerful computational tool.
- Many applications in detection problems, but also in shape recognition [Musé, Sur, Cao, Gousseau] ; image matching [Rabin, Delon, Gousseau] ; epipolar geometry [Moisan, Stival], motion detection and analysis [Cao, Veit, Bouthemy] ; clustering [Cao et al.] ; stereovision [Sabater et al.] ; image denoising (by grain filters) ; etc.
- Automatic computation of thresholds
- For a mathematician : many interesting mathematical questions of stochastic geometry.
- Prediction of perception thresholds ?

Open questions

1. How to deal with the “over-determination” of images (i.e. the fact that visual objects usually have several qualities at the same time) ?

Open questions

1. How to deal with the "over-determination" of images (i.e. the fact that visual objects usually have several qualities at the same time) ?
2. How to define a computational tool (like the *NFA*) for more than one attribute ?

Open questions

1. How to deal with the “over-determination” of images (i.e. the fact that visual objects usually have several qualities at the same time) ?
2. How to define a computational tool (like the *NFA*) for more than one attribute ?
3. How to deal with “conflicts” of qualities ?

Examples of conflicts (1)

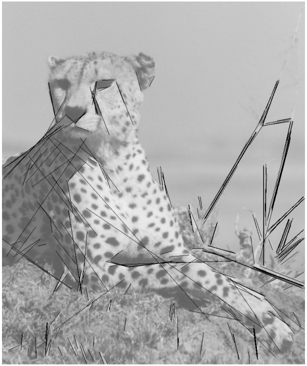


FIG.: Smooth convex sets or alignments ?

The casual alignments in the Cheetah fur are caused by the presence of many oval shapes.
Such alignments are perceptually masked and should be computationally masked !

Examples of conflicts (2)

A dense cluster of points 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 should inhibit such alignment detections.

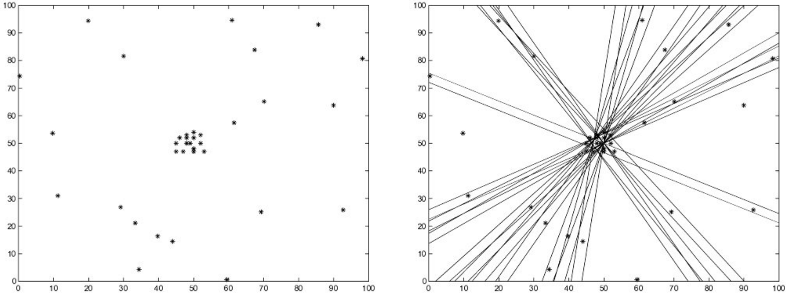
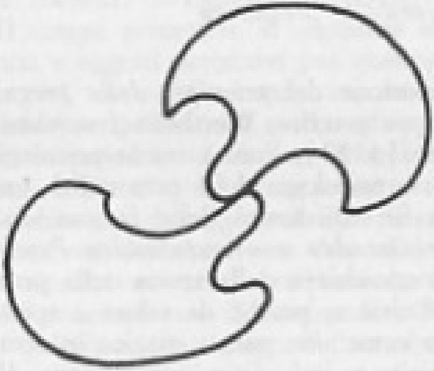


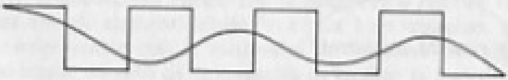
FIG.: One cluster or several alignments ?

Conflits between Gestalt grouping laws



Chiusura contro buona continuazione.

Conflits between Gestalt grouping laws



1.55 La continuità di direzione è più forte della chiusura.

How to solve conflicts ?

- There is no “universal hierarchy” of qualities.

How to solve conflicts ?

- There is no “universal hierarchy” of qualities.
- In any case, we cannot hope for any reliable explanation of a figure by summing up the results of one or several partial gestalts. Only a global synthesis, treating all conflicts of partial gestalts, can give the correct result.

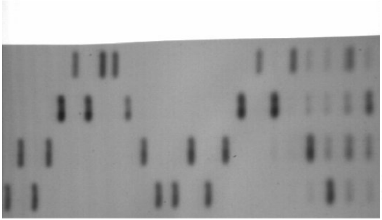
How to solve conflicts ?

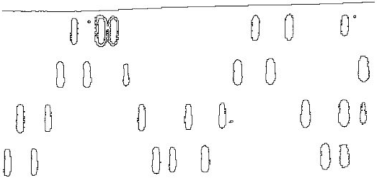
- There is no “universal hierarchy” of qualities.
- In any case, we cannot hope for any reliable explanation of a figure by summing up the results of one or several partial gestalts. Only a global synthesis, treating all conflicts of partial gestalts, can give the correct result.
- Need of another framework than Helmholtz principle : variational framework - MDL ?

How to solve conflicts ?

- There is no “universal hierarchy” of qualities.
- In any case, we cannot hope for any reliable explanation of a figure by summing up the results of one or several partial gestalts. Only a global synthesis, treating all conflicts of partial gestalts, can give the correct result.
- Need of another framework than Helmholtz principle : variational framework - MDL ?
- Partial answer : given by “Compositional approaches” (E. Bienenstock, S. Geman and D. Potter, S.C. Zhu and D. Mumford, Y. Amit and A. Trouvé). They use Probabilistic Grammars, and Binding functions to find the “global explanation” of a scene or an image.

Cooperation of attributes of objects





Thank you for your attention !