COMPARING HUMAN AND MACHINE DETECTION THRESHOLDS

An *a contrario* model for non-accidentalness.

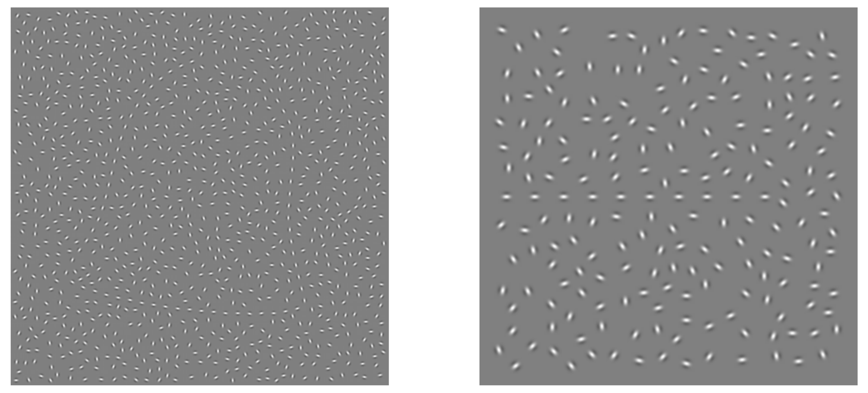
Samy Blusseau [1] , José Lezama [1,2], Rafael Grompone von Gioi [1,2], Jean-Michel Morel [1], Gregory Randall [2]

1: CMLA, ENS Cachan, France   2: IIE, Universidad de la República, Uruguay

The mathematical theory of *a contrario* detection formalizes the non-accidentalness principle [2] and attempts to predict ideal perception thresholds. Thus, it is natural to reconsider from a computational perspective, classic and new psychophysical experiments evaluating the human perception performance. To this aim, we chose the psychophysical experiments by Wagemans et al. [3] where subjects are presented with Gabor-rendered outlines of real world objects. In these experiments, orientation jitter was added to the elements with the aim of determining its effect on human object detection performance. Using the *a contrario* theory, the human detection thresholds can be compared rationally to the algorithmic ones. To allow a broader experimentation, we built an online web facility where users can perform object detection experiments, and compare their detection curves to the ones predicted analytically by the computational model.

Background

Contours detection

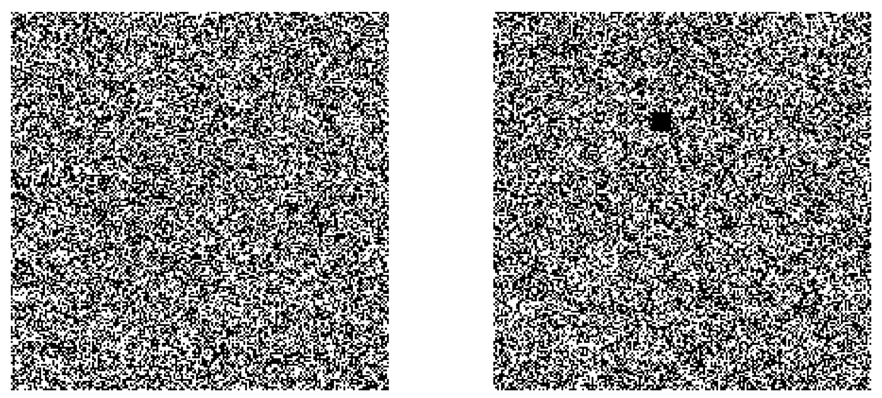


Contour of a bell [3]   Straight contour (our experiment)

From Wagemans et al's experiment [3], we kept the Gabor-rendering of shapes and their masking by adding orientation jitter on contours. In this first attempt to predict detection thresholds with *a contrario* theory, we focused on straight contours for their simplicity.

A contrario model

The non-accidentalness principle states that, among a set of potential structures, only the configurations that would rarely appear by chance are perceptually relevant. The "*a contrario*" model translates such a principle in a mathematical language, as follows: a configuration is perceptually meaningful when its expectation in noise is less than 1. This means that in average, only one false dectection would be made in a noise image.
We define an upper bound of this expectation of an event in noise, and call it "Number of False Alarms", or NFA.



Noise   An unlikely event in noise

- $N = 200^2 = 10000$ pixels, each of colour black or white with probability $p = \frac{1}{2}$
- The number $\mathcal{N}_{squares}$ of squares of all possible sizes fitting in the image, is approximately $N^{3/2}$
- Given a $n = q \times q$ pixels square, the probability to have at least k pixels of same color within this square is

$$\mathbb{P}_{k,n} = \mathcal{B}(k, n, p) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i} \tag{1}$$

and its NFA, an overestimation of the expected number of such events in the image, is defined as :

$$NFA = \mathcal{N}_{squares} \times \mathbb{P}_{k,n} \tag{2}$$

The expected number of $n = 10 \times 10 = 100$ black pixels squares, such as the one in the above right hand image, is upper bounded by its NFA, whose value is: **NFA** $= 200^3 \times 0,5^{100} << \mathbf{1}$.
This NFA becomes (much) greater than 1 for $n = 4 \times 4 = 16$. Indeed, colour squares smaller than 4×4 pixels do occur by chance and are not conspicuous.

Human detection

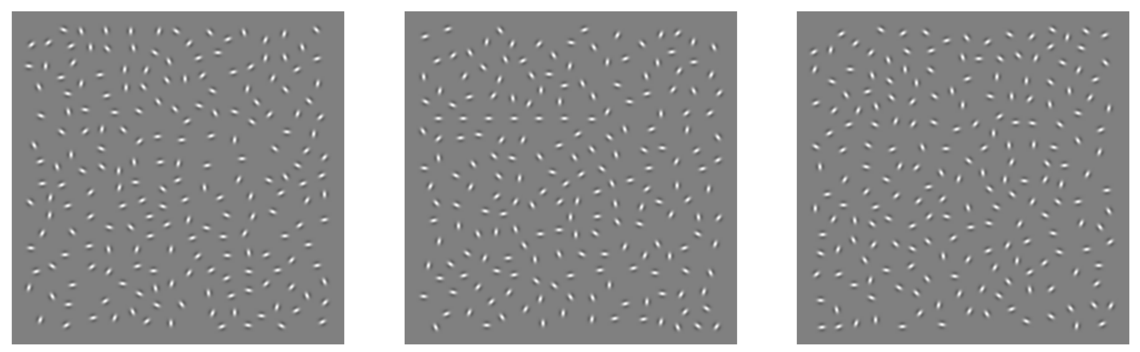
Protocole

This experiment is accessible **on the web** at `http://bit.ly/aligned_gabors`.
During a session of the experiment the subject sees **35 images**. More precisely :

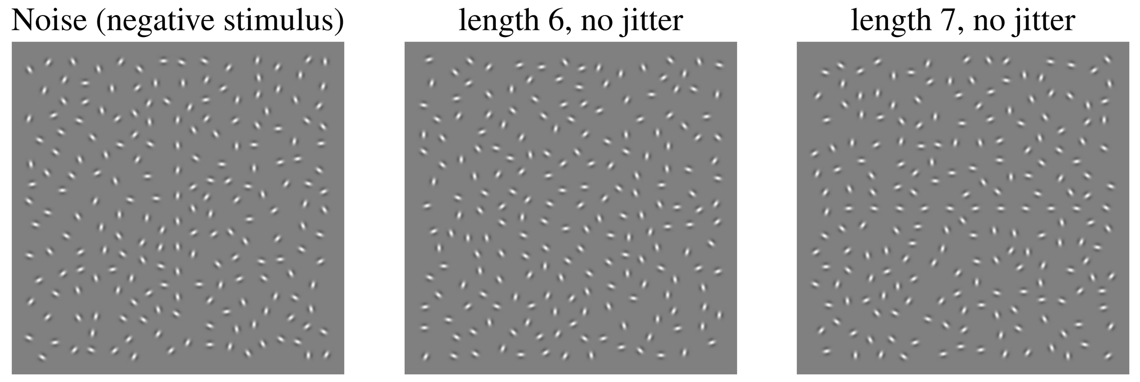
- **5 training stimuli** (the first 5 images)
- **30 images** are **randomly sampled** from the database according to the following probabilities: 25 % for negative stimuli (all elements have random orientations), 75 % for positive stimuli (some elements have constrained orientation).
- A **Yes/No question** for each stimulus: the subject has to answer whether he sees or not a straight line ; his **response time is measured** but no time limitation is imposed.

Stimuli Database

The database is large enough to avoid repetitions (more than 14 000 images), and was generated with GERT (v1.1) [1]. Each image contains $N = 200$ Gabor elements, not too close from each other.



Noise (negative stimulus)   length 6, no jitter   length 7, no jitter



length 8, jitter 15° length 9, jitter 22.5° length 10, jitter 22.5°

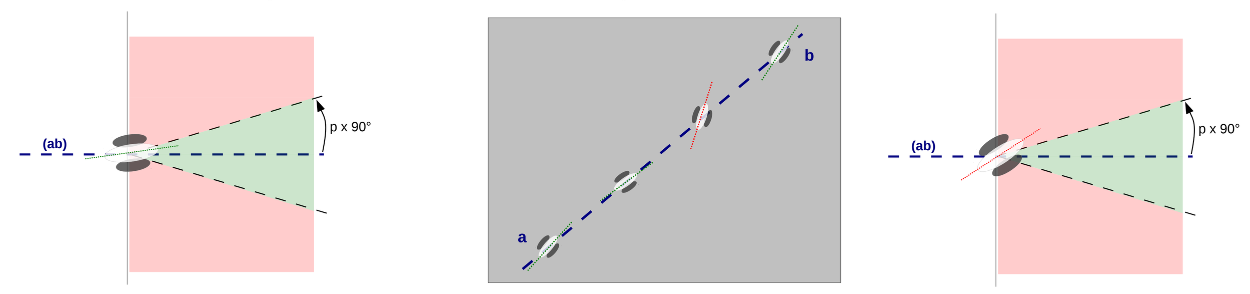
The positive stimuli (containing a straight line) vary according to :

- the straight line's length : from 3 to 10 aligned elements
- noise levels : the added orientation jitter belongs to an interval $[-\theta, \theta]$ where $\theta \in \{0^\circ, 15^\circ, 22.5^\circ, 30^\circ, 45^\circ, 60^\circ, 67.5^\circ, 75^\circ, 90^\circ\}$
- the position of the segment's center : 25 positions covering the image's area
- the slope of the segment, defined by the angle $\alpha \in \{-60^\circ, -45^\circ, -30^\circ, 0^\circ, 30^\circ, 45^\circ, 60^\circ, 90^\circ\}$ with the horizontal axis.

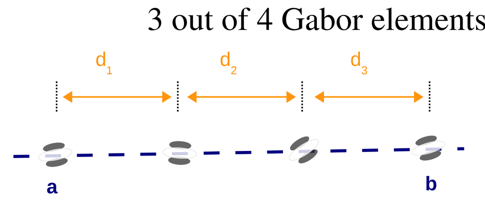
Machine detection

Grouping laws

Orientation similarity and width constancy



3 out of 4 Gabor elements sharing orientation **(ab)** with precision $p \in [0, 1]$.



Width constancy: $d_1 \approx d_2 \approx d_3 \approx \frac{d_{ab}}{3}$

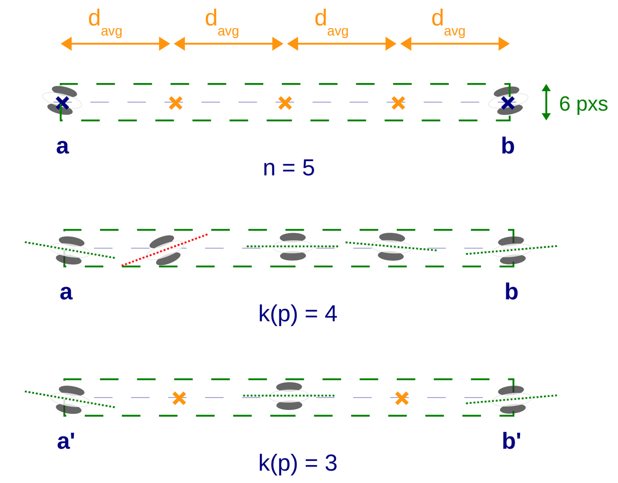
References

[1] M. Demeyer and B. Machilsen. The construction of perceptual grouping displays using GERT. *Behavior Research Methods, online first.*, pages 1–8, 2011.

[2] Agnès Dsolneux, Jean-Michel Morel, and Lionel Moisan. *From Gestalt Theory to Image Analysis, a Probabilistic Approach*. 2008.

[3] Johann Wagemans, T. Van Looy, and G. E. Nygard. The influence of orientation jitter and motion on contour saliency and object identification. *Vision Research*, 49:2475–2484, 2009.

Algorithm



Given a pair $\{a, b\}$ of Gabor elements, we define n as the expected number of elements in the stripe of length ab and width 6 pixels, knowing that the average distance between two neighbours is d_{avg} ; thus $n \approx \frac{d_{ab}}{d_{avg}} + 1$.

Then, for a precision $p \in [0, 1]$, $k(p)$ is the actual number of elements that are in the green stripe and whose orientation is parallel to (ab) with precision p.
On the left hand illustrations, the one in the middle shows a "full" stripe in which one element is not parallel to (ab) with precision p ; in the third one, only 3 elements are in the stripe, all with same orientation under precision p.

The binomial tail $\mathcal{B}(k(p), n, p) = \sum_{i=k(p)}^{n} \binom{n}{i} p^i (1-p)^{n-i}$ can be computed for each pair and any precision p. In the algorithm, each pair is tested with 5 precisions : $p_1 = \frac{1}{3}$, $p_2 = \frac{1}{4}$, $p_3 = \frac{1}{6}$, $p_4 = \frac{1}{8}$, $p_5 = \frac{1}{10}$.

For an image containing N Gabor elements, the total number of tests is

$$\mathcal{N}_{tests} = number\ of\ pairs\ \times\ number\ of\ tested\ precisions = \frac{N(N-1)}{2} \times 5$$

and for a given pair $\{a, b\}$, the significance of the corresponding straight line is given by its NFA

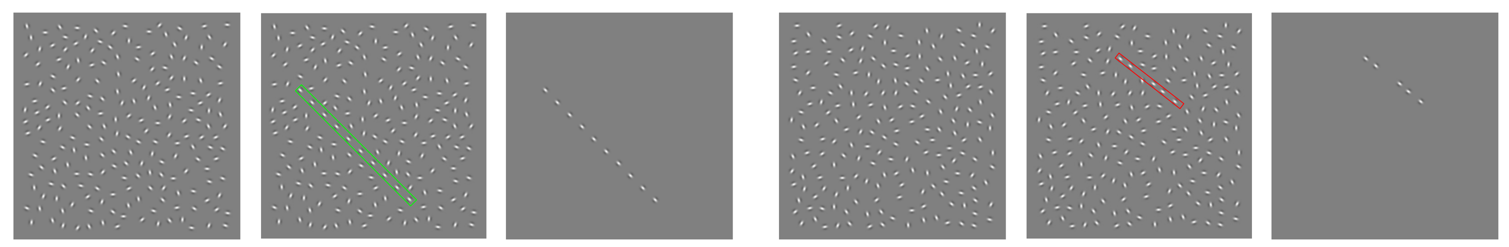
$$\mathrm{NFA}(\{a, b\}) = \mathcal{N}_{tests} \times \min_{i \in \{1,\ldots,5\}} \mathcal{B}(k(p_i), n, p_i) \tag{3}$$

The algorithm detects the structure having the lowest NFA ***if it is less than*** 1.

Results and discussion

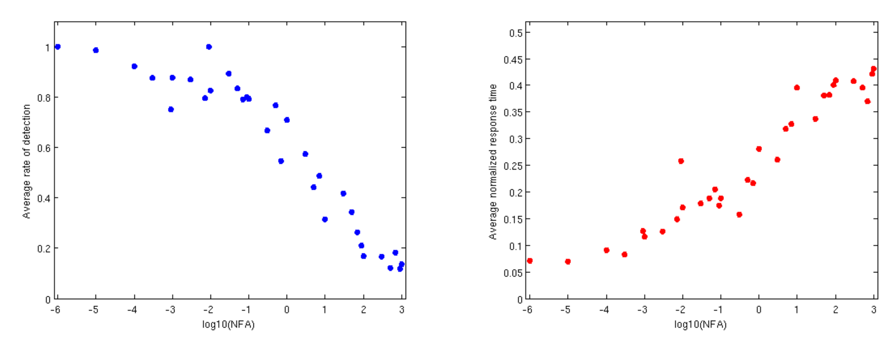
Examples of machine detection

input: length 6 jitter 15° lowest NFA $= 0.36$ detection   input: only noise   lowest NFA $= 99.5$ no detection



input: length 10, no jitter   lowest NFA $= 10^{-5}$ detection   input: length 9, jitter 60° lowest NFA $= 0.995$ "false" detection

Human detection vs. NFA



- 277 started experiments, 229 completed : 7137 trials (5305 positives)
- The $\log_{10}(\mathrm{NFA})$ line is divided into 60 bins. Every stimulus is assigned its lowest NFA. When a positive stimulus is observed during an experiment, it contributes to the detection rate and response time of the corresponding NFA bin.

- Every measured response time is normalized in $[0, 1]$, 0 and 1 corresponding respectively to the user's minimum and maximum response times.

According to the above results, the NFA seems to provide a sensible **measure** of the stimuli **difficulty for human detection**. By **predicting** how likely an alignment of Gabor elements is to be detected by humans, it is an **acceptable model for non-accidentalness** in our experiment.

However, this work is mainly a **starting point** for a more thorough investigation of the potential of *a contrario* formalism in detection modeling.

Future works will consist in (for example) : **1)** improving our detection algorithm, especially as far as the width constancy modelization is concerned ; **2)** running this experiment with different parameters (number of elements per image, size of the image...) to assess how general our framework can be ; **3)** setting up new experiments, such as pre-attentive ones, or with other kinds of stimuli, and run them in a more controlled context (not only on line).