A PSYCHOPHYSICAL EVALUATION OF THE *A CONTRARIO* DETECTION THEORY

S. Blusseau[⋆], *A. Carboni*[†], *A. Maiche*[†], *J. M. Morel*[⋆], *R. Grompone von Gioi*[⋆]

[⋆] CMLA, ENS Cachan, France
[†] CIBPsi, Facultad de Psicología, Universidad de la República, Uruguay

ABSTRACT

The *a contrario* theory is a mathematical formalization of the so-called *non-accidentalness* principle of perception, which states that an observed configuration is relevant only when it is unlikely to appear just by chance. It has been successfully applied to several image processing and computer vision problems. In this paper, human vision is compared to an *a contrario* based algorithm in a simple perceptual task. For this aim, a psychophysical experiment was set up, in which subjects and the algorithm were asked to detect alignments of Gabor patches. We found that the proposed algorithm predicted accurately the subjects' responses, therefore providing an interpretation to perceptual thresholds.[1]

Index Terms— *A-contrario* detection, Psychometric testing, Computer vision, Detection algorithms.

1. INTRODUCTION

Psychophysical studies can help computer vision researchers by suggesting general principles for the design of their algorithms [1]. Early statistical models of vision were formulated in the 1950s [2]. Since the late 1990s, the Bayesian formalism has proved relevant to model simple perceptual grouping [3] or much more complex and general mechanisms [4, 5, 6].

The *non-accidentalness* principle was introduced independently in the computer vision [7, 8] and perception [9] literature. The idea is that features that are probable to appear just by chance are not relevant, and should be rejected by setting thresholds accordingly. In the words of Lowe, "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." [8, p. 39].

The *a contrario* framework is a mathematical formulation of the non-accidentalness principle [10]. Instead of computing the *probability* that an observation arises by accident, the focus is placed on the *total number* of accidental detections per image. An *a contrario* detection algorithm estimates for each observed feature of interest its *expected* number of occurrences under a background probability model. If this number is low, the feature is termed non-accidental (in the background model) and therefore significant.

A whole family of unsupervised algorithms in computer vision are based on the *a contrario* theory. The framework is used to detect line segment and curves [10, 11, 12], vanishing points [13], aliasing [14], shots in videos [15], as well as to perform shape matching [16], change detection [17, 18], image segmentation [19], feature point matching [20], homographic registration [21], clustering [22], mirror symmetry detection [23]. Our hypothesis is that the success of these unsupervised algorithms is due to the *a contrario* theory's consistency with certain perceptual grouping mechanisms, and the present work is meant to test this hypothesis.

A few experimental studies have confronted computer vision theories to human perception [24]; among them, [25] centered on the *a contrario* theory. The latter study already proved that a quantitative definition of non-accidentalness gives an interpretation and quite accurate predictions of the perceptual thresholds. Its most remarkable insight is to translate several detection-affecting parameters into one unique measure of rareness, the Number of False Alarms (NFA), which correlates well with the subjects detection performance. Yet the psychophysical protocols in [25] had several limitations that also limited the strength of its conclusions.

First, the experimental set-up in [25] relied on "yes-no" questions to assess the detection by subjects. Here, we used a refined protocol in which subjects indicate *where*, instead of *whether*, a structure is seen. We also recorded reaction times, providing an additional measure of confidence to the subjects' responses, and followed a paradigm that is commonly used in psychophysical studies of good continuation [26, 27].

Second, [25] did not compare *directly* human subjects to an algorithm. In contrast, we took inspiration from the innovative work by Fleuret et al. [28] where human and machine performed the same visual categorization task, "side by side", on each stimulus. In a nutshell, the algorithm became an ar-

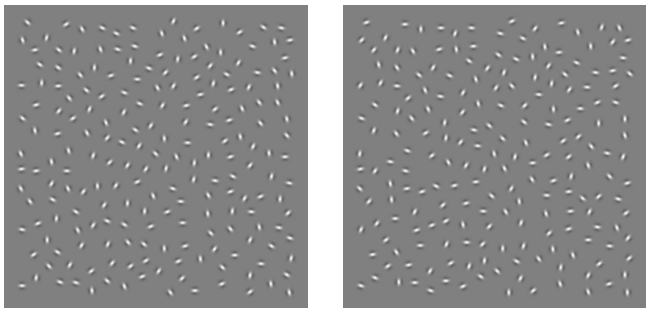

Fig. 1. Gabor arrays like those displayed during the experiment. The coordinates of the elements are the same in both images, and only their orientations change; the level of angular jitter added to the 10 aligned elements varies from 0 rads (left), to $\frac{\pi}{2}$ rads (right), which corresponds to the model of the random Gabor array (defs. 1 and 2).

tificial subject. Similarly, our experiment compared human and machine vision in a simple perceptual task: the detection of line patterns among a set of Gabor patches (Fig. 1(left)).

Arrays of Gabor patches are a classic tool in perceptual grouping to study good continuation [26, 27]. Gabor functions ensure a control on the spectral complexity of the stimuli and on the spatial scale of the contours. They provide a flexible and easy way of generating a great variety of stimuli. Additionally, they are a good model for the receptive field of simple cells in the visual cortex [29]. It has been verified that human vision is better at grouping oriented elements when they are roughly tangent to a smooth contour [26, 24, 27].

However, these studies were limited to measuring the detection performance as a function of parameters, without testing a theoretical and interpretative model. Our aim is to explain such measurements and we will show that the *a contrario* theory gives a quantitative answer.

This paper is organized as follows: Sect. 2 describes the perceptual task, including the experimental set up. Our detection algorithm is presented in Sect. 3. We discuss the results of the comparison between subjects and the algorithm in Sect. 4, and draw conclusions in Sect. 5.

2. PERCEPTUAL TASK

We measured the detectabilty, for human vision, of a structure embedded in an image, as a function of the parameters used to generate it. We then compared this data to a detection algorithm's output to assess whether the detection thresholds of perception and the a contrario theory match. In this study, we only tested the detection of line-segments.

The stimuli displayed in the experiments were Gabor arrays like those of Fig. 1, and were generated with the software GERT (v1.1) [30]. The N oriented blobs contained in an array are called Gabor elements or patches. Each patch represents a symmetrical Gabor function, which is a zero-phase sinusoid convolved with a Gaussian. Each array includes n aligned elements, regularly spaced by a distance r_a. Their orientations were chosen as follows: given the line's direction θ_l, the n angles were uniformly and independently sampled in $[\theta_l - \alpha, \theta_l + \alpha]$, with $\alpha \in [0, \frac{\pi}{2}]$. This is equivalent to adding uniform noise, also called *angular jitter*, to orientations equal to θ_l. The alignment position and direction θ_l were selected randomly in each image. The $N - n$ non-aligned elements, called *background elements*, were randomly placed but respected a minimal distance r_b from each other and from the aligned patches. Their orientations were uniformly and independently sampled in $[0, \pi)$.

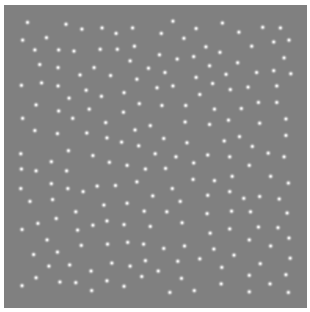

Fig. 2. These dots have the same coordinates as the Gabor patches of Fig. 1. It is almost impossible to detect the ten aligned elements without the orientation information.

Distances r_a and r_b were set as functions of N to fulfill two requirements. First, r_b was tuned so that all N elements fit into the image and fill it homogeneously, avoiding clusters and empty regions. Second, we chose r_a larger than r_b to make the alignment almost impossible to detect from the coordinates of the elements only, that is to say from the proximity and width constancy cues only (Fig. 2). Then the difficulty to detect the aligned Gabors was essentially ruled by their number n, and by the angle α. The larger n and the smaller α, the more conspicuous the alignment.

Each experimental session counted 126 trials divided into three blocks of 42, with pauses after the first and the second block. In each trial, the subject was shown on a screen a Gabor array of 496×496 pixels, containing $N = 200$ elements, n of which were aligned ($n \in \{4, \dots, 9\}$) and affected by an angular jitter of certain intensity $\alpha \in \{0, \frac{\pi}{12}, \frac{\pi}{8}, \frac{\pi}{6}, \frac{\pi}{4}, \frac{\pi}{2}\}$. For this value of N and this image size, we had $r_a = 34.5$ and $r_b = 22.5$ pixels. Four trials were conducted for each couple of conditions (n, θ) with $\theta \leq \frac{\pi}{4}$, and one trial per maximal jitter conditions $(n, \frac{\pi}{2})$. The subjects knew that every stimulus contained an alignment, but did not know where, how long and how jittered it was. They were asked to click on an element they perceived as part of the alignment, and just to give their best guess in case they did not see any clear alignment. We wanted the task to be *attentive*. Thus, the stimulus remained on the screen until an answer was given, although subjects were advised to spend at most 20 seconds per stimulus. The coordinates of the click and the response time were recorded. After the subject's click, a transition gray image

was displayed for 500 ms and then the next stimulus appeared.

Twelve subjects, six women and six men, took the experiment through a web interface (`http://bit.ly/ac_alignments`). They were all naive to the purpose of this study. Each subject used their own computer device. Thus, the monitor properties, the distance to the screen and the illumination conditions could vary from one subject to the other. Our bet was that, for the scope of our study, such variability would not have significant inter-subject effects. To help understand the task, subjects were first shown two training sequences of 5 and 10 trials, with the corresponding ground-truths at the end of each sequence.

3. ALGORITHM

This section defines the alignment detection algorithm to be matched to human perception. Without loss of generality, in our model we take $[0,1]^2$ as the domain of the images.

Definition 1. *We call* N-Gabor array *a set* $\mathbf{g} = \{(x_i, \theta_i), i = 1 \dots N\}$ *where* N *is an integer,* $x_i \in [0,1]^2$ *and* $\theta_i \in [0, \pi)$. *Each element* $\gamma_i = (x_i, \theta_i)$ *is called* Gabor element.

Definition 2. *A* N*-Gabor array* $\mathbf{G} = \{(X_i, \Theta_i), i = 1 \dots N\}$ *where* X_i *and* Θ_i *are random variables, and the* Θ_i *are independently and uniformly distributed in* $[0, \pi)$*, is called* random N-Gabor array *in the remaining of the paper.*

Fig. 1 shows two representations of Gabor arrays. Whereas the right hand image is a realization of a random Gabor array, the left hand one clearly deviates from the *a contrario* model for its 10 elements with constrained orientations. Note that in Def. 2, no assumption is made on the Gabor elements' coordinates, since we tried to minimize their influence by constructing stimuli as explained in Sect. 2.

Definition 3. *Let* $a \neq b \in [0,1]^2$, $w > 0$. *We call* w-wide rectangle associated to $\{a, b\}$, *and note* $\mathcal{R}(a, b, w)$, *the rectangle going from* a *to* b, *of width* w *(Fig. 3).*

The input to the algorithm is a N-Gabor array $\mathbf{g}$. The first step is to compute the average Euclidean distance $d_{avg}(\mathbf{g})$ of the elements of $\mathbf{g}$ to their nearest neighbor. Then, for each pair $\{\gamma_i, \gamma_j\}$ of $\mathbf{g}$, the rectangle $r = \mathcal{R}(x_i, x_j, \frac{d_{avg}(\mathbf{g})}{10})$ is considered. The number n_{obs} of elements of $\mathbf{g}$ belonging to r, is compared to an ideal number $n_{th} = \left\lfloor \frac{|x_i - x_j|}{d_{avg}(\mathbf{g})} + \frac{1}{2} \right\rfloor + 1$. If $n_{obs} < 0.7 \times n_{th}$, then the rectangle r is rejected because it is not considered to be dense enough to contain a visible alignment. Otherwise, the rectangle is further examined. The algorithm computes the difference between the orientation θ_k ($1 \leq k \leq n_{obs}$) of each Gabor element in r, to the one of the rectangle, noted θ_r. Dividing these angular errors by $\frac{\pi}{2}$ yields normalized errors $e_k \in [0,1]$ such that $e_k = 0$ when $\theta_k - \theta_r = 0 \mod \pi$ and $e_k = 1$ when $\theta_k - \theta_r = \frac{\pi}{2} \mod \pi$ (Fig. 3). The maximal error for the rectangle $e_{\max} \doteq \max\{e^*, e_1, \dots, e_{n_{obs}}\}$ is measured up to a minimal precision $e^* = \frac{1}{10}$. Finally, the Number of False Alarms associated to the pair $\{\gamma_i, \gamma_j\}$, is defined as:

$$\mathrm{NFA}(\{i,j\}, \mathbf{g}) = \frac{N(N-1)}{2} \cdot (e_{\max})^{n_{obs}} .$$

As usual [10], a large NFA value corresponds to a likely (and therefore insignificant) configuration in the *a contrario* model, whereas a small one indicates a rare and interesting event. The algorithm's output is the set of Gabor elements that belong to the rectangle associated to the pair of smallest NFA (Fig. 4). The following proposition shows that this method satisfies the non-accidentalness principle.

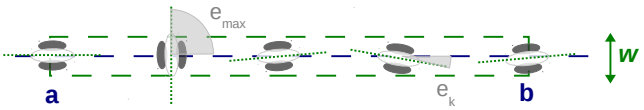

Fig. 3. An example of rectangle; $n_{obs} = 5$ and $e_{max} = 1$.

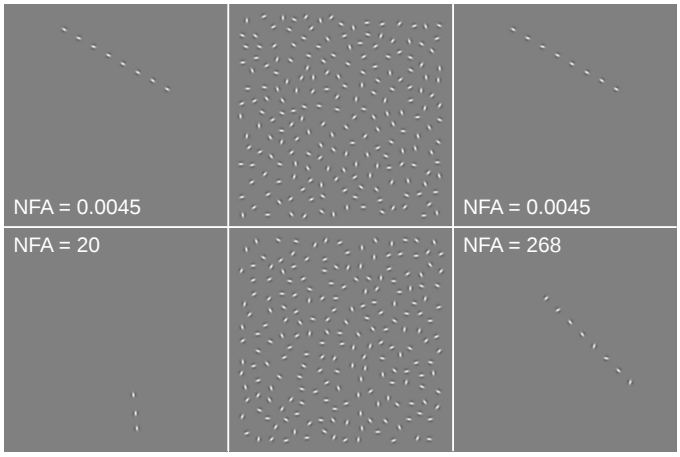

Fig. 4. Detections by the algorithm (Sect. 3). **Center:** input. **Left:** output. **Right:** ground truth.

Proposition 1. *Let* $\varepsilon > 0$, *and* $\mathbf{G} = \{(X_i, \Theta_i)\}_{i=1 \dots N}$ *a random* N*-Gabor array. Then the expected number of pairs* $\{i,j\}$ *such that* $\mathrm{NFA}(\{i,j\}, \mathbf{G}) < \varepsilon$, *is less than* ε.

4. RESULTS AND DISCUSSION

Each click made by a subject is associated to the nearest Gabor element in the image, and counted as a valid detection when that element belongs to the ground truth. The same is done for the algorithm, by selecting a random element among the returned ones. This count permits to define the detection rate for humans and for the algorithm. Figs. 5 (a) and (b) plot the detection rate as a function of the jitter level and the number of aligned elements. Figs. 5 (c) and (d) plot the detection rate and the reaction time as a function of $\log_{10}(\mathrm{NFA})$ of the

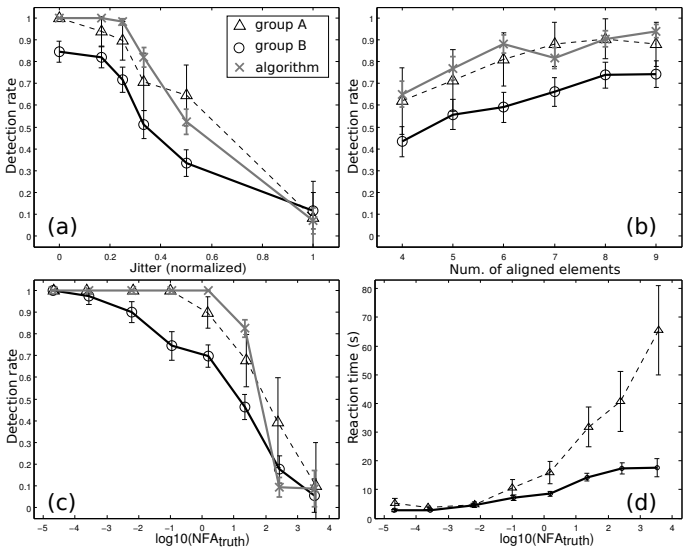

Fig. 5. Detection rate as a function of: **(a)** the normalized jitter level $\frac{\alpha}{\pi/2}$ of the ground truth alignment; **(b)** the number n of elements in the ground truth alignment; **(c)** the $\log_{10}$ of the ground truth's NFA. Plot **(d)** shows the reaction times, in seconds, as a function of $\log_{10}(\mathrm{NFA})$. **Circles**, **Crosses**, **Triangles**: group A, B and algorithm's responses, respectively.

ground truth. The abscissa is always divided into bins: every point represents an average value for the corresponding trials. The error bars give 95 % confidence; they are defined as $[\overline{x} - 2\frac{\sigma}{\sqrt{q}}, \overline{x} + 2\frac{\sigma}{\sqrt{q}}]$, where $\overline{x}, \sigma$ and q are respectively the mean, standard deviation and number of trials of the bin.

A Kruskal-Wallis test on detection results indicates that all 12 subjects did not give homogeneous answers ($p \approx 1.8 \cdot 10^{-5}$), but can be split into two significantly homogeneous groups: on one side, group A is composed of subjects 8 and 11 ($p \approx 1$), both achieving a detection rate of 80%; on the other side, group B is formed by the remaining 10 subjects ($p \approx 0.27$), with detection rates between 52% and 67%.

Figs. 5(a) and (b) confirm a previously observed phenomenon [26, 27]: the angular jitter affects the detection performance, but this effect gets weaker when the number of aligned elements increases (at constant number of background elements). More striking is the fact that the NFA of the ground truth alignments is an *accurate* one-dimension measure of the stimuli's difficulty. Indeed, Figs. 5(c) and (d) show decreasing detection rates and increasing reaction times with respect to the NFA. This yields the following interpretation: points with $\log_{10}(\mathrm{NFA})$ lower than -2 in graphics (c) and (d) represent stimuli in which the ground truth alignment is expected to occur less than once every 100 random Gabor arrays. At the other end of the NFA axis, on the right of $\log_{10}(\mathrm{NFA}) = 2$ for example, stand the images whose hidden alignment is an event that could occur, on average, more than 100 times in a single random Gabor array. Thus it seems that, in our experiment, detectability and *non-accidentalness* are strongly correlated.

S.	1	2	3	4	5	6
τ	-0.30	-0.32	-0.30	-0.21	-0.27	-0.32
$\mathbf{p}_\tau <$	10^{-6}	10^{-7}	10^{-6}	10^{-3}	10^{-5}	10^{-7}
ρ	0.54	0.68	0.61	0.47	0.63	0.57
$\mathbf{p}_\rho <$	10^{-10}	10^{-300}	10^{-13}	10^{-7}	10^{-14}	10^{-11}

7	8	9	10	11	12	**all**
-0.24	-0.26	-0.35	-0.29	-0.23	-0.30	-0.28
10^{-4}	10^{-4}	10^{-8}	10^{-5}	10^{-4}	10^{-6}	10^{-300}
0.69	0.73	0.65	0.69	0.65	0.62	0.60
10^{-300}	10^{-300}	10^{-15}	10^{-300}	10^{-15}	10^{-13}	10^{-300}

Table 1. Kendall's τ between subjects' answers and the NFA, and Spearman's ρ between reaction times and the NFA, for each subject (**S**). The p-values were computed with a normal approximation (for τ) and a Student test (for ρ).

$\log_{10}(\mathrm{NFA})$	$(-4.7, 0]$	$(0, 1]$	$(1, 2]$	$(2, 4.17]$
nb. of trials	661	322	333	196
Detection rate Subj. (%)	87	71	45	17
Detection rate Algo. (%)	100	100	77	5.6
Agreement rate (%)	87	71	53	80

Table 2. Detection and agreement rates for 4 NFA intervals.

In order to assess the significance of these behaviors, we computed the Kendall's rank correlation coefficient (noted τ) between each subject's answers and the NFA, and Spearman's rank correlation coefficient between each subject's reaction times and the NFA (Table 1). The p-values associated to Kendall's τ were obtained with a non parametric test based on a normal approximation, whereas Spearman coefficient's significance was checked with a Student test. As Table 1 shows, the significance of the observed trend is overwhelming.

We finally compared the subjects to the algorithm. What appears at first glance is the similarity of shape between the black curves and the gray one, in Figs. 5 (a), (b) and (c). The results of group A are particularly similar to the algorithm's ones, and plot (d) shows that these subjects took significantly more time than average. Furthermore, we looked at the agreement between a subject and the algorithm (i.e. both detect or both fail). Among the 1512 trials of the 12 sessions, the agreement rate is 74.8%. Splitting the NFA range into 4 well chosen intervals (Table 2) helps understand the strengths and flaws of the algorithm at imitating human subjects. On the one hand, the algorithm keeps performing well even in the interval $(1, 2]$ where the subjects' performance is already quite low. On the other hand, subjects are still able to find almost 20% of the alignments when $\log_{10}(\mathrm{NFA}) > 2$, whereas the algorithm's detection rate drops down to less than 6%. This difference is materialized by the area between black and gray curves in graphic 5 (c). A likely explanation is that the algorithm processes accurately the *orientation* cue whereas subjects are less accurate but may use other cues, such as the constant spacing between aligned elements.

5. CONCLUSION

Our results suggest two main conclusions. First, the NFA provides a quantitative law linking length and jitter to alignments' visibility. Second, despite its simplicity, the proposed algorithm imitates quite well, and thus predicts, the average behavior of human subjects. These results support the *a contrario* theory from a perceptual point of view. They suggest that it is possible to control accurately the false alarm rates of computer vision algorithms. In future studies, this experiment will be held in a controlled environment, and our method generalized to more complex detection tasks.

6. REFERENCES

[1] H. J. Chang, M. S. Park, Jeong H., and J.Y. Choi, "Tracking failure detection by imitating human visual perception," in *ICIP*, 2011, pp. 3293–3296.

[2] F. Attneave, "Some informational aspects of visual perception.," *Psychological Review*, vol. 61, no. 3, pp. 183–193, May 1954.

[3] J. Feldman, "Bayesian contour integration," *Attention, Perception, & Psychophysics*, vol. 63, no. 7, pp. 1171–1182, 2001.

[4] D. C. Knill and W. Richards, *Perception as Bayesian inference*, Cambridge University Press, 1996.

[5] R. P. N. Rao, B. A. Olshausen, and M. S. Lewicki, *Probabilistic models of the brain: Perception and neural function*, Mit Press, 2002.

[6] D. Heyer and R. Mausfeld, *Perception and the physical world*, Wiley Online Library, 2002.

[7] A. P. Witkin and J. M. Tenenbaum, "On the role of structure in vision," in *Human and Machine Vision*, J. Beck, B. Hope, and A. Rosenfeld, Eds., pp. 481–543. Academic Press, 1983.

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

[9] I. Rock, *The Logic of Perception*, The MIT Press, 1983.

[10] A. Desolneux, J. M. Morel, and L. Moisan, *From Gestalt Theory to Image Analysis, a Probabilistic Approach*, Springer, 2008.

[11] R. Grompone von Gioi, J. Jakubowicz, J. M. Morel, and G. Randall, "LSD: a line segment detector," *IPOL*, 2012.

[12] V. Pătrăucean, P. Gurdjos, and R. Grompone von Gioi, "A parameterless line segment and elliptical arc detector with enhanced ellipse fitting," *ECCV*, pp. 572–85, 2012.

[13] A. Almansa, A. Desolneux, and S. Vamech, "Vanishing point detection without any a priori information," *PAMI*, vol. 25, no. 4, pp. 502–507, 2003.

[14] B. Coulange and L. Moisan, "An aliasing detection algorithm based on suspicious colocalizations of fourier coefficients," in *ICIP*, 2010, pp. 2013–2016.

[15] P. Hellier, V. Demoulin, L. Oisel, and P. Perez, "A contrario shot detection," in *ICIP*, 2012, pp. 3085–3088.

[16] P. Musé, F. Sur, F. Cao, and Y. Gousseau, "Unsupervised thresholds for shape matching," in *ICIP*, 2003.

[17] T. Veit, F. Cao, and P. Bouthemy, "An a contrario decision framework for region-based motion detection," *IJCV*, vol. 68, no. 2, pp. 163–178, 2006.

[18] A. Robin, L. Moisan, and S. Le Hegarat-Mascle, "An a-contrario approach for subpixel change detection in satellite imagery," *PAMI*, vol. 32, no. 11, pp. 1977–1993, 2010.

[19] J. Cardelino, V. Caselles, M. Bertalmio, and G. Randall, "A contrario hierarchical image segmentation," in *ICIP*, 2009, pp. 4041–4044.

[20] J. Rabin, J. Delon, and Y. Gousseau, "A contrario matching of sift-like descriptors," in *ICPR*, 2008, pp. 1–4.

[21] L. Moisan, P. Moulon, and P. Monasse, "Automatic homographic registration of a pair of images, with a contrario elimination of outliers," *IPOL*, vol. 2012, 2012.

[22] M. Tepper, P. Musé, and A. Almansa, "Meaningful clustered forest: an automatic and robust clustering algorithm," *CoRR*, vol. abs/1104.0651, 2011.

[23] V. Pătrăucean, R. Grompone von Gioi, and M. Ovsjanikov, "Detection of mirror-symmetric image patches," in *CVPRW*, 2013, pp. 211–216.

[24] W. S. Geisler, J. S. Perry, B. J. Super, and D. P. Gallogly, "Edge co-occurrence in natural images predicts contour grouping performance," *Vision research*, vol. 41, no. 6, pp. 711–724, 2001.

[25] A. Desolneux, L. Moisan, and J. M. Morel, "Computational gestalts and perception thresholds," *Journal of Physiology - Paris*, vol. 97, pp. 311–324, 2003.

[26] D. J. Field, A. Hayes, and R. F. Hess, "Contour integration by the human visual system: Evidence for a local association field," *Vision Research*, vol. 33, no. 2, pp. 173 – 193, 1993.

[27] J. Wagemans, T. Van Looy, and G. E. Nygard, "The influence of orientation jitter and motion on contour saliency and object identification," *Vision Research*, vol. 49, pp. 2475–2484, 2009.

[28] F. Fleuret, T. Li, C. Dubout, E. K. Wampler, S. Yantis, and D. Geman, "Comparing machines and humans on a visual categorization test," *PNAS*, vol. 108, no. 43, pp. 17621–17625, 2011.

[29] J.P. Jones and L.A. Palmer, "An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex," *Journal of neurophysiology*, vol. 58, no. 6, pp. 1233–1258, 1987.

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