Image segmentation by *a contrario* simulation

Nicolas Burrus [a,b] Thierry M. Bernard [a] Jean-Michel Jolion [b]

[a] *ENSTA - UEI, 32 Boulevard Victor, 75015 Paris, France*

[b] *Université de Lyon, INSA Lyon, LIRIS, UMR CNRS 5205, 69621 Villeurbanne Cedex, France*

Abstract

Segmenting an image into homogeneous regions generally involves a decision criterion to establish whether two adjacent regions are similar. Decisions should be adaptive to get robust and accurate segmentation algorithms, avoid hazardous *a priori* and have a clear interpretation. We propose a decision process based on *a contrario* reasoning: two regions are meaningfully different if the probability of observing such a difference in pure noise is very low. Since the existing analytical methods are intractable in our case, we extend them to allow a mixed use of analytical computations and Monte-Carlo simulations. The resulting decision criterion is tested experimentally through a simple merging algorithm, which can be used as a post-filtering and validation step for existing segmentation methods.

Key words: segmentation, *a contrario* reasoning, statistical image processing, Monte-Carlo simulation

1 Introduction

General purpose image segmentation aims at partitioning an image into homogeneous groups of connected pixels, called regions. This is a long standing problem in computer vision motivated both by its utility in high level vision and by phenomenological experiments coming from the Gestalt theory that shows the importance of perceptual grouping for human visual perception.

This task is difficult for several reasons. First, there are intrinsic ambiguities in perceptual grouping when there is no top-down control and/or strong *a priori*. Thus, different ways of analyzing the image will lead to different segmentations, even if a common decision criterion is used. Second, natural images are highly variable. Thus, it is very difficult to design a general purpose algorithm which takes coherent decisions in all kind of images [1]. Even within a single

image very important variations can be observed, e.g. because of texture and illumination changes.

Thus, many segmentation algorithms have been proposed during the last 30 years [2,3]. They all somehow involve a decision process which states, given two adjacent regions, whether they should be considered as a whole or as distinct entities. They mainly differ in their hypotheses about the image, the analyzed features and the decision thresholds they use [4].

A popular class of methods consider image segmentation as a global energy minimization problem [4–7]. While appealing, these global formulations generally rely on strong quantitative *a priori*, e.g. to control boundaries length versus data fidelity in [6], or to control the number and the sizes of regions in [5]. Moreover, their minimization is often hard and time-consuming, even if great advances have been achieved [8,9].

The other major group of approaches, on which this paper focuses, use explicit predicates on couple of regions. Given two adjacent regions, a criterion measures their difference. This criterion may be empirical [10,11] or statistically motivated [4,12,13]. Then regions are merged if the criterion is below a certain decision threshold.

Because of the high variability within and between images, decision thresholds are difficult to choose and are generally set by the users empirically, at the regional level [10,13,11] or using a global stopping criterion [14]. However, it is desirable for a general purpose segmentation method, whatever its complexity or practical qualities, to provide partitions for which some fundamental properties can be formally ensured on all images. This requires an automatic and adaptive thresholding procedure that avoids quantitative *a priori*, as it seems impossible to define accurate and universal *a priori* quantities [15].

A contrario reasoning, or Computational Gestalt Theory (CGT), has already proved its efficiency for such kind of problems, e.g. for other Gestalt detection [16], motion detection [17], or shape matching [18]. Instead of using necessarily incomplete and approximative models of natural images, *a contrario* methods rely on a — generally trivial — model of pure chance, called the *a contrario* or noise model. Then, according to the so-called Helmholtz principle, events of interest are those whose probability to occur in pure noise is very low. The main interest of this approach comes from its ability to compare event significance accurately using simple *a contrario* models and to automatically determine perceptual thresholds.

In the context of image segmentation, we want to decide whether the differences between two adjacent regions are high enough to keep the regions separate in the final partition. Reasoning *a contrario*, one does not need to model differences occurring in natural images, but only the differences which

can occur by chance. Actually, this has been done in several works where only a background model of homogeneity is provided (generally independent and identically distributed pixel values), and where differences are measured as deviations from this model, e.g. using the Fisher test in [4], the Wilcoxon-Mann-Whitney test in [19,20] or the independent bounded difference inequality in [12].

Still, these works do not provide a grounded way to set the decision thresholds. CGT provides a solution to this problem by estimating the expectation of the number of false alarms as a function of the decision threshold. Then, choosing a bound on the accepted average number of false alarms, one can deduce automatically the optimal thresholds.

Our goal in this paper is to apply this methodology for generic image segmentation and show whether accurate thresholds can be obtained without using any quantitative *a priori* about the images to segment. Robustness will be ensured by the guarantee that in average, whatever the image under analysis, no regions whose difference are only due to noise will be kept separate.

In most previous CGT works, thanks to a significant modeling effort, estimating the number of false alarms could be done analytically. In the segmentation case, data-driven exploration heuristics are used to analyze the set of possible partitions efficiently, and their complex behavior makes it impossible to perform purely analytical computations. Moreover, analytical bounds are difficult to obtain when several criteria (distance measures between regions in our case) are jointly used.

These observations led us to develop a more flexible algorithmic model, by replacing some complex analytical computations by Monte-Carlo simulations. The main idea is to observe what differences can be statistically observed between regions in white noise to deduce significance thresholds automatically. We shall see that the resulting thresholds ensure the same bounds on the expected number of false alarms than the analytical computations of previous works.

This new flexibility allows us to propose a generic segmentation method, which can integrate a large variety of distance measures and most bottom-up heuristics to explore regions (region growing and merging, graph pruning, etc.), while always ensuring the same statistical properties on the resulting partitions.

Section 2 first introduces the notion of ε-reliable segmentation algorithm in an *a contrario* framework. Then Section 3 proposes an example of *a contrario* difference measure, which will be thresholded to make decisions. Section 4 shows how optimal thresholds can be automatically obtained by mixing simulations and analytical computations, and some results are presented in Section 5. The methodology is finally discussed in Section 6.

2 ε-reliable segmentation algorithm

We consider the family of segmentation algorithms composed of two elements:

- A *selection* function which can state, for a couple of adjacent regions, whether they are meaningfully different or not. Thus two regions will be said *selected* — and therefore kept separate — if their differences are considered meaningful.
- A heuristic which explores couples of regions in the image, and builds the final partition according to the decisions of the selection function about the encountered regions. Examples of heuristics are iterative region growing or split and merge based algorithms. To get a valid segmentation algorithm, the heuristic must keep only regions whose adjacent regions in the final partition are meaningfully different, according to the selection function. An example of heuristic will be given in Section 5.

Most classical segmentation algorithms fit into this family, at least indirectly.

Reasoning *a contrario*, we are looking for regions whose differences are very unlikely to be due to chance. First, to define what we mean by "chance", we "*a contrario*" introduce a family of models in which observed differences are considered casual.

Definition 1 *An a contrario model for couples of regions is a model for which pixel values in both regions are independent and identically distributed. We note P_{ac} the probability distribution on couples of regions induced by this family of models.*

In these models, there are no spatial dependencies between pixels, which are organized in a purely random way. Thus, we do not expect any meaningful groups of connected pixels in these models. Note that, for natural images, Nyquist sampling itself implies that only pixels of which the distance is greater than two can be independent [15]. However, for the sake of simplicity, we ignore this detail in the following, but in Section 4, where the Nyquist distance issue will be implicitly taken into account by generating well-sampled white noise images.

With this definition of chance, couples of regions to select and keep separate are those whose differences are unlikely to be a result of an *a contrario* model. Reasoning this way, we can already define a statistically-founded ordering on couples of regions: a couple w_1 of regions is *more* significantly different than a couple w_2 if the probability to observe the difference of w_1 in an *a contrario* model is *lower* than the probability to observe the difference of w_2.

Then the question becomes: below which value of this probability two regions should be considered meaningfully different? This threshold directly influences the rate of false alarms: the lower the threshold, the less likely it is to select regions whose differences are actually a result of an *a contrario* model. However, this threshold is hard to set *a priori*, because it does not have a simple physical interpretation.

This issue was already addressed in previous *a contrario* works, by considering the problem more globally. We follow the same approach here. Segmentation algorithms are applied to whole images, and not standalone couples of regions. Then a sound way to set the above threshold is to relate it to a statistical estimation of how many couples of adjacent regions will be wrongly considered as meaningfully different during the segmentation process, i.e. how many false alarms a segmentation algorithm will make in one image.

The exact probability distribution of this number of false alarms (called NFA in [21]) is hard to compute, because couples of regions analyzed by a segmentation algorithm may overlap, and thus are not independent. For similar reasons, [21] introduced the expectation of this number of false alarms, which is both easier to compute thanks to expectation linearity, and has a more intuitive interpretation.

Applied to segmentation, the expectation of the number of false alarms made by a segmentation algorithm is the average number of adjacent regions considered as meaningfully different in an image, whereas their difference are actually due to chance. Obviously, the lower this expectation, the more reliable the algorithm. This provides an intuitive measure of the reliability of a segmentation algorithm, which will allow us to determine decision thresholds. Hence the notion of ε-reliability for a segmentation algorithm.

Definition 2 *Let $\mathcal{A}$ be a segmentation algorithm composed of an exploration heuristic $\mathcal{H}$ and a selection function S. Let W be the set of couples of regions analyzed by $\mathcal{H}$ and selected as meaningfully different by S on an image of size $N \times M$. Now consider the couples of regions in W whose differences are actually a result of an a contrario model. If their expected number is less than ε, then $\mathcal{A}$ is said to be ε-reliable for $N \times M$ images.*

An ε-reliable algorithm is guaranteed to make, in average, less than ε false alarms in an image. A false alarm occurs when S selects a couple of regions analyzed by $\mathcal{H}$ while their difference is actually a result of an *a contrario* model, i.e. is due to chance. It turns out that the choice of ε is not very sensitive [21], since decision thresholds will only logarithmically depend on it. Thus, a standard choice is $\varepsilon = 1$, which means that in average, less than one false alarm is expected per image.

3 The selection function S_δ

So far, the notion of "region difference" has remained generic. To illustrate our method, we introduce in Section 3.1 a possible set of four distance functions which will be combined to provide a global difference measure between regions. This measure will be related to the probability of observing such high values for the four distances functions in an *a contrario* model. It will then be thresholded to create a selection function in Section 3.2.

3.1 The difference measure F

The difference measure only has to provide an ordering of couples of regions. The relevance of the ordering will determine the detection rate, but in all cases, the number of false alarms will be controlled by the automatic threshold selection of Section 4. However, the difference measure has to be adaptive w.r.t. gray level distributions, as will be required by Proposition 2.

Relying on *a contrario* reasoning, we introduce a difference measure that ranks couple of regions according to the probability of observing their difference by chance. The difference is measured by some distance functions. This paper does not aim at proposing the best distance functions between regions maximizing some evaluation criteria — which are application dependent — but at showing that, given some distance functions, we can combine them quasi-optimally within the *a contrario* framework to build a robust segmentation algorithm. Thus, we exhibit four classical distances, for which it is possible to compute probability distributions in the *a contrario* models. Each distance can be seen as a random variable associated to a couple of regions, and for each variable its distribution in the *a contrario* family of models is given.

3.1.1 Gray level distributions analysis

If pixels in two regions are independent samples of a unique global distribution, what is the probability for the observed distributions in each region to be as different as they are? Regions may be small and are not necessarily normally distributed. The Wilcoxon-Mann-Whitney U test [22] is well-adapted to such a situation. This probabilistic test indirectly relies on the behavior of the difference between the medians of two samples drawn from the same distribution. Let R_1 and R_2 be the two regions to compare, pixel values in both regions are merged and sorted, and a random variable U is computed by summing the ranks of the pixels of R_1 (or R_2). The distribution of U for small samples can be computed exactly, and for large samples (in practice, N_1 or

N_2 greater than 20) it becomes normally distributed:

$$P_{ac}(U \leq u) = \mathcal{N}_{\leq}\left(u, \frac{N_1 N_2}{2}, \sqrt{\frac{N_1 N_2 (N_1 + N_2 + 1)}{12}}\right)$$

with N_1 and N_2 the size of R_1 and R_2, and $\mathcal{N}_{\leq}(x, \mu, \sigma)$ the cumulative distribution function of a normal law with mean μ and deviation σ, evaluated in x. The closer the medians of R_1 and R_2, the closer U to $\frac{N_1 N_2}{2}$. U is used to build the first distance function $D_1 = |U - \frac{N_1 N_2}{2}|$, whose distribution in the family of *a contrario* models is straightforward to deduce:

$$P_{ac}(D_1 \geq d) = 2 \times \mathcal{N}_{\geq}\left(d, 0, \sqrt{\frac{N_1 N_2 (N_1 + N_2 + 1)}{12}}\right)$$

with $\mathcal{N}_{\geq}$ the complementary cumulative distribution function of the normal law.

3.1.2 Variance analysis

If two regions R_1 and R_2 are samples of a unique random process, what is the probability for their respective variances to be that small compared to the variance of $R_1 \cup R_2$? Let σ^2 and μ_4 be the second and fourth moments of the gray level distribution of $R_1 \cup R_2$, S_1^2 and S_2^2 the empirical gray level variance of respectively R_1 and R_2. Then the following approximation holds [23]:

$$P_{ac}(S_1^2 \leq s_1^2) \simeq \mathcal{N}_{\leq}\left(s_1^2, \sigma^2, \frac{\sqrt{\mu_4 - \sigma^4}}{\sqrt{N_1}}\right)$$

The same holds for S_2. Two distance functions can be deduced from S_1 and S_2 : $D_2 = \frac{1}{1+S_1}$ and $D_3 = \frac{1}{1+S_2}$. The higher D_2 and D_3, the more different are the regions, and:

$$P_{ac}(D_2 \geq d_2) = P_{ac}(S_1^2 \leq s_1^2)$$
$$P_{ac}(D_3 \geq d_3) = P_{ac}(S_2^2 \leq s_2^2)$$

Since the approximation is not valid for small samples, this term is only used when N_1 and N_2 are greater than 20.

3.1.3 Frontier contrast analysis

If two regions are samples of a unique random process, how contrasted may be their frontier? A suitable procedure is found in [15], except that the distribution of the gradient $G(x) = P(X \geq x)$ is used on $R_1 \cup R_2$ here, instead of on the full image. Using a 2x2 difference scheme, the gradient values of two pixels

are independent in an *a contrario* model if their distance is greater than 2. Calling D_4 the random variable representing the minimal gradient value obtained by picking one value every two pixels along the frontier and L_{12} the frontier size, then:

$$P_{ac}(D_4 \geq d_4) = G(d_4)^{\frac{L_{12}}{2}}$$

This provides the last distance function D_4.

3.1.4 Combining the distance functions

Ideally, the statistical information brought by each distance functions should be combined through their joint probability. But the distance functions are not necessarily independent, and even though their dependencies might be quantitatively rather small, they are difficult to estimate. Thus, we introduce an approximation of this probability.

Definition 3 *For a couple of regions $w = (R_1, R_2)$, let F be the following a contrario difference function:*

$$F(w) = -log\left(\prod_{k=1}^{4} P_{ac}(D_k \geq D_k(w)|N_1, N_2, L_{12}, G)\right)$$

with N_1 and N_2 the size of R_1 and R_2, L_{12} the frontier size between R_1 and R_2, and G the gradient distribution on $R_1 \cup R_2$.

Region sizes and their global gradient distribution do not convey by themselves interesting information about their differences, so we consider them as known variables. The lower the probability to observe the difference between two regions in an *a contrario* model, the more meaningfully different they are, and the higher F will be. F is only an approximation, but we assume that it will generally respect the ordering given by the joint probability: when a couple of regions is more meaningfully different than another according to the joint probability, then it will also be more meaningful according to F. This can be checked experimentally for the chosen distance functions. Note that the validity of this assumption will not change the reliability of the segmentation algorithm in terms of number of false alarms, but may decrease its sensitivity and relevance.

3.2 Construction of S_δ

A selection function can be easily built on top of F by introducing a decision threshold. This threshold should be chosen so that the selection function, when combined with a given exploration heuristic, provides an ε-reliable segmentation algorithm. However, a unique threshold is not the best choice. Indeed,

in an image, there are much more possible middle-sized regions than big or small ones. For a $N \times M$ image, there are only $N \times M$ regions of one pixel, and one region of $N \times M$ pixels. Between these two extremes, the number of regions increases and then decreases exponentially, quite similarly to binomial coefficients. Thus, statistically, it is more likely to observe by chance high differences among middle-sized regions than among small or big ones.

Another source of variability is the exploration heuristic itself. It might be better at finding large deviations among small regions than among big regions, no matter how many couples of regions are available. For example, Figure 1 shows that a naive heuristic merging regions using a "scanline" order is less likely to find couples of middle-sized regions deviating from noise than couples of small regions.

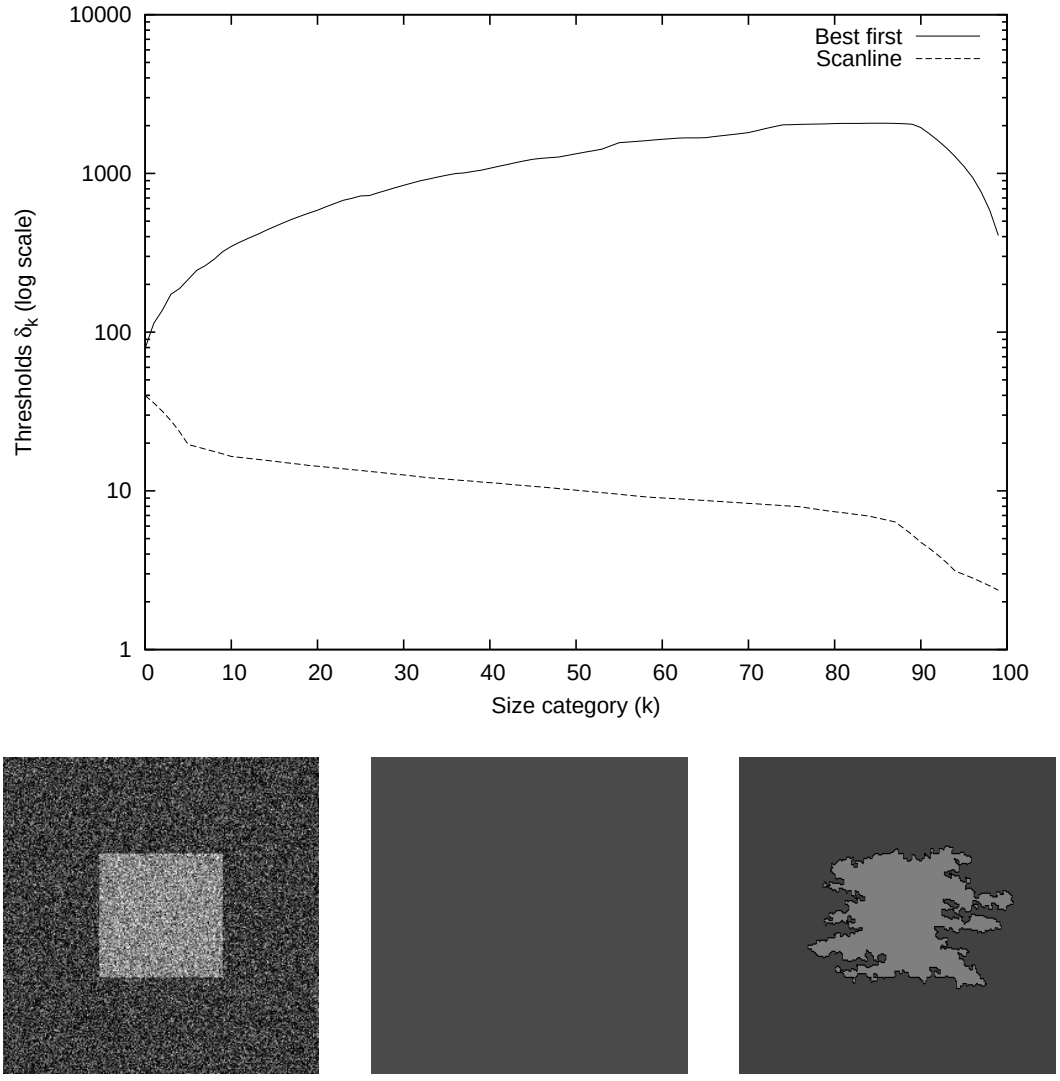
To take these variabilities into account and be more discriminant, the threshold should be adapted to the size of the regions under analysis. This is experimentally confirmed in Section 5. There are about $\frac{(N\times M)^2}{2}$ possible sizes for a couple of regions. This number is too large to introduce one particular threshold for each couple size, because thresholds will be individually estimated by Monte-Carlo simulations in Section 4. The larger the number of thresholds to estimate, the longer the simulation procedure. To reduce the number of thresholds to estimate, we partition the different cases using a quantization function.

Definition 4 *Let $N \times M$ be the image size and K a positive integer constant. A partition function $\mathcal{J}$ is a quantization function mapping couples of region sizes from $\{1 \dots N \times M\}^2$ into $\{1 \dots K\}$.*

Thus, a partition function associates an integer between 1 and K to each couple of region sizes. To build such a partition function, it is quite natural to start with a log-quantization of region sizes. Indeed, a precise pixel count does not make sense for regions above a certain size: only a relative precision is needed. So given the image dimensions N and M, let us introduce $\mathrm{lq}_S(n)$, a function that quantizes a region of size n (between 1 and $N \times M$) onto S levels (between 1 and S):

$$\mathrm{lq}_S(n) = \left\lfloor \frac{S \times \log(n)}{\log(N \times M + 1)} \right\rfloor + 1$$

Then a general way to build a partition function is to associate the couple $(\mathrm{lq}_S(N_1), \mathrm{lq}_S(N_2))$ to each couple of regions of sizes N_1 and N_2. There are S^2 possible outputs. By ordering N_1 and N_2 such that $N_1 \leq N_2$, the number of outputs decreases to $\frac{S(S+1)}{2}$. This eventually partitions the set of region couples into $K = \frac{S(S+1)}{2}$ subsets, and we call $\mathcal{J}_{2d}$ the corresponding partition



(a) (b) (c)

Fig. 1. This figure illustrates the need for thresholds to be adapted to heuristic behaviors and not only to the number of regions they analyze. The plot (top) shows the learned thresholds for Algorithm 1 (cf. Section 5) with a watershed initialization for two different merging orders. "Best first" merges the couple of regions with the lowest F value at each iteration. "Scanline" is a naive strategy that analyzes couples of regions from left to right, top to bottom. While the number of analyzed regions is roughly equivalent in both cases, F values obtained by chance are much higher using the "best first" order. This is further illustrated on the synthetic square image (a). Using the thresholds learned for "best first" with a "scanline" order (b), deviations found in the square area are not large enough and no meaningful region is found. Using the same "scanline" order, but with adapted thresholds (c), the square can be somehow segmented. Similar observations were made in [12] when discussing the merging order (Figure 4 of their paper), but we nuance them here by showing that threshold adaptation can improve the results of the "scanline" order.

function.

This generic quantization function is adapted to any kind of heuristic, as long as K is high enough to ensure a good precision. In practice, $S = 100$ is a reasonable choice to quantize region sizes. This leads to $K = \frac{S(S+1)}{2} = 5050$ different thresholds to estimate. However, an empirical analysis on the heuristics used in this paper shows that this number of thresholds can be further reduced with insignificant loss of accuracy. Indeed, Figure 2 shows that for a fixed minimal region size in a couple of regions, the size of the bigger region does not have much influence, especially for small regions, which are the most sensitive ones. Thus, we introduce the following, simpler partition function which only depends on the size of the smaller region in a couple:

$$\mathcal{J}_{\min}(N_1, N_2) = \lfloor K \times \frac{\log(\min(N_1, N_2))}{\log(\frac{N \times M}{2} + 1)} \rfloor + 1$$

$\mathcal{J}_{\min}$ will be used in all the experiments of this paper. Since only one dimension has to be quantized, Section 5 will show that the number of different size categories can be safely reduced to $K = 100$.

We now define the final selection function, based on a multiple thresholding of the difference function F (Definition 3).

Definition 5 *Let $\delta = \{\delta_1, \delta_2, \ldots, \delta_K\}$ be a set of K real thresholds, and $\mathcal{J}$ a partition function. We call S_δ the function which selects a couple of regions w if and only if:*

$$F(w) > \delta_{\mathcal{J}(N_1, N_2)}$$

where N_1 and N_2 are the number of pixels in the two regions.

This function selects couples of regions whose difference function is higher than the threshold corresponding to its size category. Now we have to set the K thresholds to ensure ε-reliability for a given exploration heuristic.

4 Computing optimal thresholds

4.1 Analytical computations?

In most previous *a contrario* works, the expectation of the number of false alarms and then the optimal thresholds could be computed analytically. This is not possible here for two main reasons.

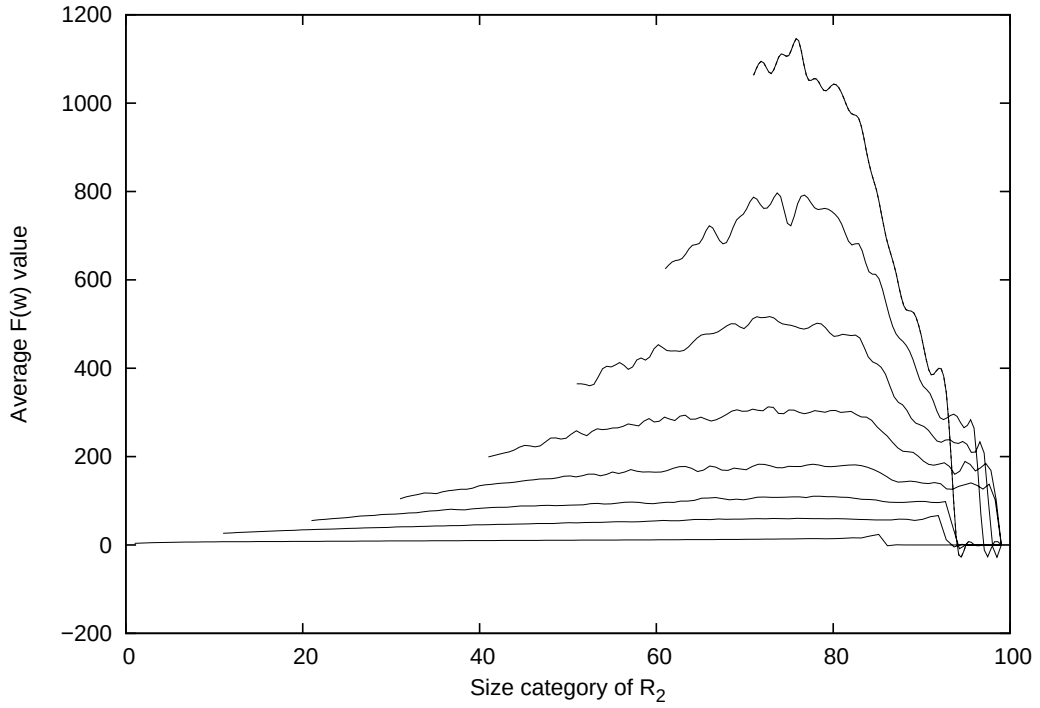


Fig. 2. Average difference values F observed by applying Algorithm 1 (cf. Section 5) with a watershed initialization on 256×256 uniform white noise images. For each couple of regions w, let R_1 be the smaller region and R_2 the bigger one. Each curve corresponds to a given size category for R_1, ranging from 1 (bottom) to 71 (top) by steps of 10 (the last 29 categories, i.e. size categories above 71 are not shown to keep the plot readable). The x-axis gives the size category of R_2. F values only depend marginally on it for a given minimal size. This is especially true for small to medium minimal sizes, which are the most sensitive, leading to the choice of the partition function $\mathcal{J}_{\min}$ in Section 3.2.

First, several discriminative features are used to compute the difference function. Thus, the probability distribution of F is a distribution of a product of random variables, which is difficult to compute analytically. In previous *a contrario* works, either a single feature was used [15,21], or features were normalized and projected onto a single one using a min or max operator [18].

Second, an important hypothesis in these works was the independence between the procedure of extraction of the entities to analyze (e.g. level lines in [15]) and the measured properties (minimal contrast in [15]). This hypothesis does not hold here since segmentation algorithms use exploration heuristics, which are generally data-driven and select couples of regions to analyze depending on their differences. Thus, the statistical properties of the analyzed couples of regions depend on the exploration heuristic, and this dependency is hard to estimate.

In [24], a data-driven heuristic was also used to find meaningful rectangles

efficiently. The expected number of false alarms was computed as if the algorithm analyzed all the possible rectangles of the image. Since the heuristic was very good, the detection thresholds remained accurate. This approach is not possible for image segmentation for two reasons. First, segmentation heuristics are far from being optimal. Considering the huge number of possible regions in an image, the thresholds would be highly pessimistic. Second, even if this solution was acceptable, evaluating the number of possible couples of regions in an image remains a hard problem of enumerative combinatorics.

In [25], a region merging algorithm was developed for stereo analysis using Computational Gestalt Theory and faced a rather similar issue. The bias created by a data-driven heuristic was ignored and the expectation of the number of false alarms was estimated as a function of the number of regions analyzed by the heuristic. However, Figure 1 shows that two different heuristics, even if they analyze a similar number of regions, can lead to dramatically different statistical properties for the analyzed regions and therefore require different thresholds. The reason is that, after a few iterations, regions analyzed by a data-driven heuristic will be among the most meaningful of the image, and thus observed differences become statistically higher than for regions selected arbitrarily. This bias is not critical in [25] because a relative merging criterion is used to compare the expectation of the number of false alarms associated to keeping two regions separate and merging them. Thus, the absolute number of false alarms does not need to be estimated as accurately as in our case.

These observations led us to develop a hybrid methodology to estimate the expectation of the number of false alarms, by mixing analytical computations and Monte-Carlo simulations.

4.2 Simulation-based thresholds computation

The following proposition allows to find the thresholds independently for each size category.

Proposition 1 *If a segmentation algorithm $\mathcal{A}$ is $\frac{\varepsilon}{K}$-reliable for each of the K size categories given by the partition function $\mathcal{J}$, then $\mathcal{A}$ is ε-reliable.*

The proof is straightforward using expectation linearity: if an algorithm makes less than $\frac{\varepsilon}{K}$ errors in average for each size category, then it globally makes less than ε errors in average. Let us now find the optimal thresholds for a given heuristic $\mathcal{H}$ and for a given image size $N \times M$. It is possible to repeat the given procedure identically for any other heuristic or image size. To get an ε-reliable algorithm, the thresholds must guarantee that, in average, less than ε couples of regions which are actually a result of an *a contrario* model will be selected in one image.

We first look for thresholds ensuring ε-reliability for uniform white noise (u.w.n.) images. Then we will show that under some conditions, the obtained thresholds also ensure ε-reliability for any other kind of images.

Thanks to Proposition 1, it is just needed to prove $\frac{\varepsilon}{K}$-reliability for each size category. We propose the following algorithm:

(1) Initialize a set of thresholds $\delta = \{\delta_k\}_{k=1...K}$
(2) Generate Q well-sampled uniform white noise images of size $N \times M$, by sampling a uniform white noise signal whose high frequencies have been removed by an ideal low-pass filter to take into account the Nyquist distance issue (see Section 2).
(3) For each image:
   - Run the heuristic $\mathcal{H}$ with S_δ, and store the number of couples of regions selected by S_δ for each size category.
(4) For each size category k:
   - Compute the empirical mean m_k and deviation s_k of the number of couples of regions selected by S_δ
   - Compute a confidence interval on the expectation μ_k using the classical property:
   $$P(M_k \leq m_k) = F_{Q-1}(\frac{m_k - \mu_k}{s_k}\sqrt{Q-1})$$
   with $F_n(x)$ the cumulative distribution function of a Student law with n degrees of freedom.
   - For a chosen confidence level, if the estimated upper bound of μ_k is greater than $\frac{\varepsilon}{K}$, increase δ_k otherwise decrease δ_k.
(5) Repeat until convergence of the thresholds.

At each step, μ_k is the expectation of the number of false alarms made by the algorithm for the k-th category of sizes. After threshold convergence, we know with the chosen confidence that the segmentation algorithm is ε-reliable for well-sampled uniform white noise images.

In practice, a very good initialization can be obtained by first running this algorithm once with infinite thresholds (thus preventing any selection) and storing computed F values on regions analyzed by $\mathcal{H}$ at step (3). Then, the maximal values for each size category are excellent initial threshold values.

Furthermore, since F grows exponentially with the distance functions, ε is not a very sensitive value. All computations are done in logarithmic scale, and thresholds are only required to ensure a reliability of the same order of magnitude as ε to get satisfying results.

Thus, using the former initialization, the above algorithm turns into a validation procedure rather than an estimation procedure, and thresholds convergence is generally obtained in one or two iterations using a geometrical factor

of 1.01 to increase/decrease δ_k.

Our method requires a new set of thresholds for each different size of images. This is not so problematic since image sizes are mostly standard and one can pre-compute thresholds for usual sizes. Furthermore, thresholds are evolving smoothly with image size, as shown in Figure 3, so we can even use interpolation for non-standard sizes.

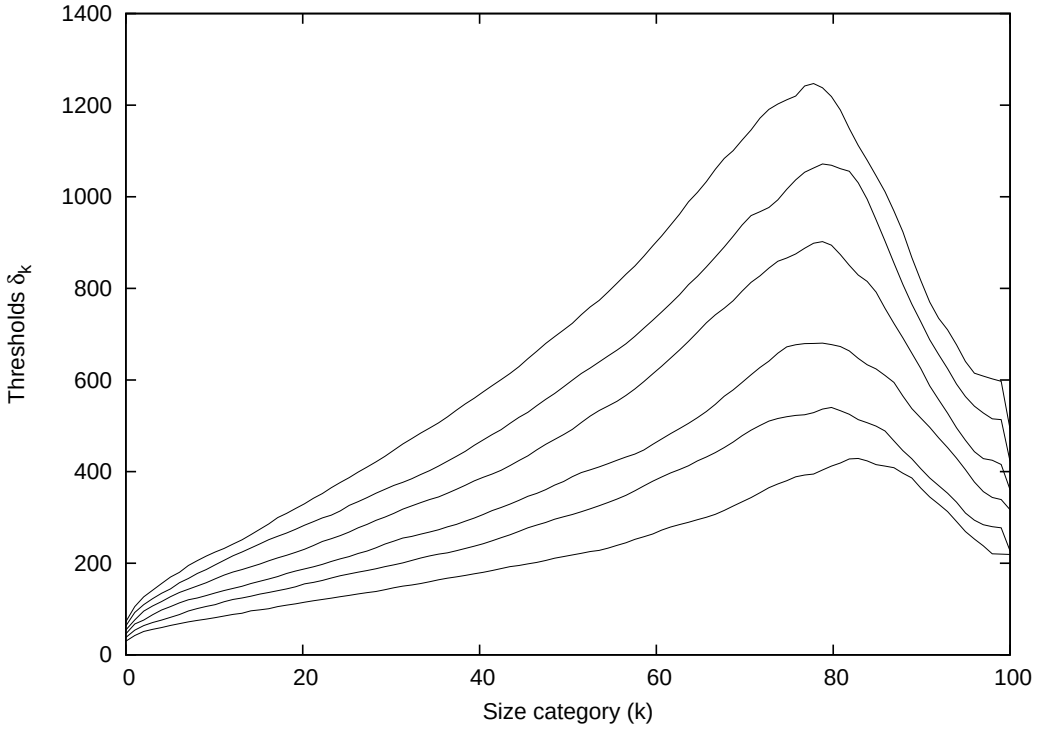


Fig. 3. F values are obtained by applying Algorithm 1 (cf. Section 5) with a watershed initialization on 1000 uniform white noise images with different sizes. Curves are stacked by increasing image sizes, starting with 100×100 pixels for the lowest one to 200×200 for the highest one, with steps of 20 pixels on each dimension. Since values evolve smoothly with image size, they may be safely interpolated for intermediate sizes.

4.3 Conditions for ε-reliability on arbitrary image

The thresholds computed in Section 4.2 ensure that, in average, less than ε couples of regions will be considered meaningfully different in a uniform white noise (u.w.n.) image. This ensures the ε-reliability of the segmentation algorithm on u.w.n. images. What about other kind of images? A u.w.n. image only contains *a contrario* couples of regions, which is the worst case, since any detection would be a false alarm. Thus, one can expect that if a segmentation algorithm is not biased by the uniformity of the white noise distribution (the *a contrario* model only assume i.i.d. pixels), neither biased by the fact that

the whole image is pure noise, then it can only produce less false alarms on images which are not pure noise. Hence the following proposition, which gives formal conditions to ensure that an ε-reliable algorithm on u.w.n. images is also ε-reliable on other kinds of images.

Proposition 2 *Let $\mathcal{A}$ be a segmentation algorithm composed of a heuristic $\mathcal{H}$ and the selection function S_δ. Let $\Gamma(w)$ denote the gray level distribution of a couple of regions w, and $\Gamma_u(w)$ denote the special case of a uniform gray level distribution. $\mathcal{A}$ is ε-reliable for $N \times M$ images under the following hypotheses:*

(1) $\mathcal{A}$ is ε-reliable for $N \times M$ well-sampled uniform white noise images.

(2) $\forall \delta, \quad P_{ac}(F(w) > \delta \mid \Gamma(w)) \leq P_{ac}(F(w) > \delta \mid \Gamma_u(w))$

(3) $\forall \delta, \forall \Gamma(w), \quad P_{ac}(\mathcal{H}(w) \mid F(w) > \delta, \Gamma(w)) \leq P_{ac}(\mathcal{H}(w) \mid F(w) > \delta, \Gamma_u(w))$

(4) Let R be a region in an image where pixels are independent and identically distributed. Let $\Omega = I \backslash R$ denote the remaining part of the image. Then we must have, for a couple of regions w included in R:

$$\forall \delta, \quad P_{ac}(\mathcal{H}(w) \mid F(w) > \delta, \Omega) = P_{ac}(\mathcal{H}(w) \mid F(w) > \delta)$$

A mathematical proof is given in Appendix A. Here, we just clarify the meaning of each of the four conditions. The first one is the ε-reliability of $\mathcal{A}$ on uniform white noise images, which can be asserted using the thresholding procedure of Section 4.2. The second one is that F values in an image (independently from the heuristic) should not be biased by the gray level distribution of the couple of regions, knowing that the pixels are independent and identically distributed (i.i.d.). F is a product of probabilities, and each of these probabilities adapts to the global distribution. Thus, the only remaining influence is the entropy of the distribution, which may not allow large deviations to occur. For example, in the extreme case of a Dirac distribution, all pixels will have the same value and only F values of $-log(1) = 0$ will be observed. Since the uniform distribution of white noise has the maximal entropy, we expect the inequality to be satisfied.

The third condition states that the heuristic must not be better at finding couples of i.i.d. regions with a high F value when the distribution of gray levels on the regions is not uniform. In other words, the distribution of deviating regions should not help the heuristic to find them.

Conditions (2) and (3) can be asserted empirically using simulations in white noise images with different distributions. This is done for the watershed initialization in Figure 4. More precisely, what is asserted by this experiment is that the joint probability of having high F values and of being selected by the

heuristic for an *a contrario* couple of regions only gets lower or equal when the couple's distribution is not uniform.

The fourth condition is more difficult to assert formally. Given an i.i.d. region R included in an image I, it means that the ability of the heuristic to find deviations from chance in R should not be biased by the properties of the surrounding regions. Intuitively, this condition will be met at least for bottom-up heuristics, where exploration in R will not be deeply influenced by the surroundings of R.

In practice, these conditions are stricter than necessary since the order of magnitude of F values one can get in an *a contrario* model is way lower than typical F values for different regions in natural images, as illustrated in Figure 5. This enables a large choice of exploration heuristics, and Section 5 will show three satisfying examples.

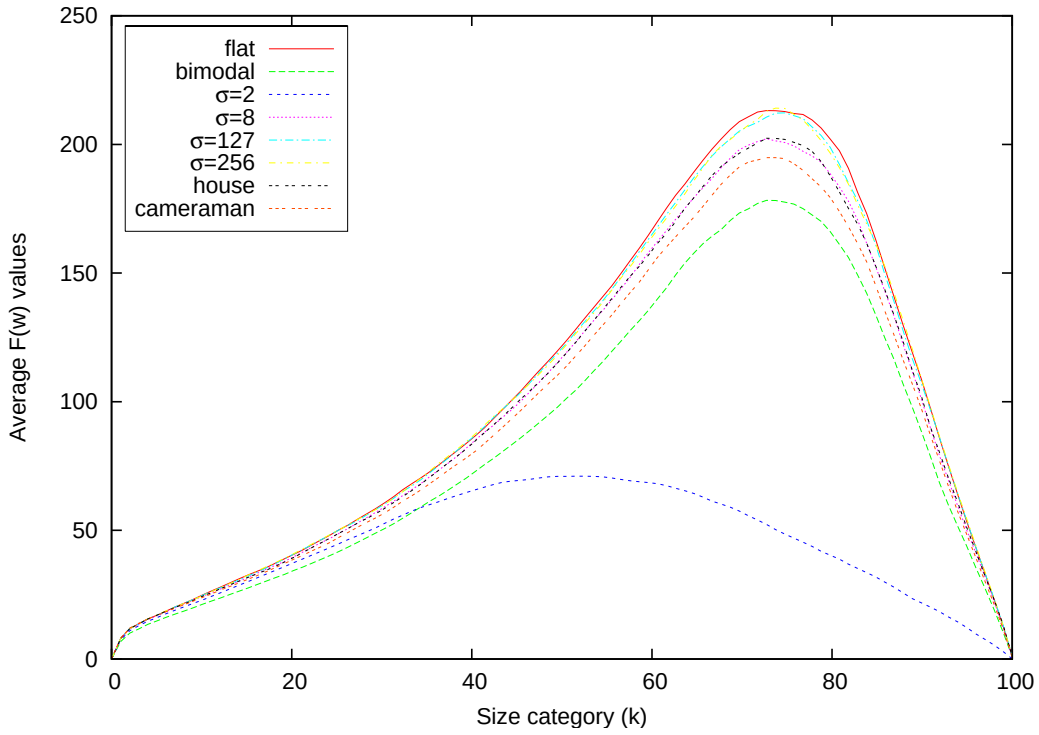


Fig. 4. Average F values observed by applying Algorithm 1 (cf. Section 5) with a watershed initialization on 1000 128×128 pure noise images with different gray level distributions. Bimodal is a binary distribution with $P(X = 0) = P(X = 255) = 0.5$. Flat corresponds to uniform white noise. "house" is the distribution of the image of Figure 7, and "cameraman" the distribution of the image of Figure 8. Other curves correspond to Gaussian distributions centered on 127, with various deviations σ. $\sigma = 2$ is almost $P(X = 127) = 1$ and $\sigma = 256$ is almost flat. Except for extreme cases (bimodal and $\sigma = 2$), curves are quite similar to the flat one, showing a good independence of F w.r.t regions gray level distribution.

5 Results

The proposed segmentation method may exploit any existing exploration heuristic as long as it matches the conditions of Proposition 2. We choose to illustrate it using Algorithm 1, based on a greedy, merge-only exploration heuristic. Starting from an initial segmentation, the algorithm iteratively merges adjacent regions which are not meaningfully different according to S_δ. At each iteration, it merges the couple w for which $F(w)$ is minimal. This algorithm was combined with three different initializations: a classical watershed with gaussian pre-filtering [26], the graph-based algorithm of [10] (called EGBIS), and the region merging algorithm of [12] (called SRM). For EGBIS and SRM the binaries provided by the authors were used. The source code used for these experiments is freely available on the authors' webpage.

Algorithm 1: Fusion-based segmentation algorithm using S_δ with a merge-only exploration heuristic. Initial partitioning is free, it may be any over-segmenting algorithm (at worst one pixel per region).

```
Initialize regions ;
Compute F for every adjacent couple of regions;
while Some couples are not selected by S_δ do
    Merge the couple with the lowest F value;
    Update F for adjacent couples of regions ;
end
```

In all experiments, ε is set to 1 and the number K of size categories for $\mathcal{J}_{\min}$ is set to 100, ensuring enough partitioning. The confidence for the threshold estimation of Section 4 is set to 0.99. The three initialization algorithms have their own parameters, which are set to $\sigma = 0.8$ for the watershed gaussian filter, $\sigma = 0.8$, $k = 150$, minarea $= 20$ for EGBIS, and $Q = 256$ for SRM. These parameters aim at capturing the maximal amount of details in the image. EGBIS with the minimal area parameter set to 0 provides similar results after post-processing, but filtering small regions *a priori* highlights the non-trivial post-processing task.

A contrario post-processing time depends on the degree of over-segmentation given by the initialization. Observed processing times for 256×256 images do not exceed 500 ms for the watershed initialization, 200 ms for EGBIS and 100 ms for SRM on an Intel Core 2 Duo run at 2.4GHz, without using multicore parallelism.

Figure 5 shows the thresholds obtained for the watershed initialization, and all the observed F values during the segmentation process of the "house" image. One can see many large deviations to the *a contrario* model, materialized by F values way above the thresholds. These deviations correspond to meaning-

fully different regions. The order of magnitude of the deviations confirms the robustness of the approach.

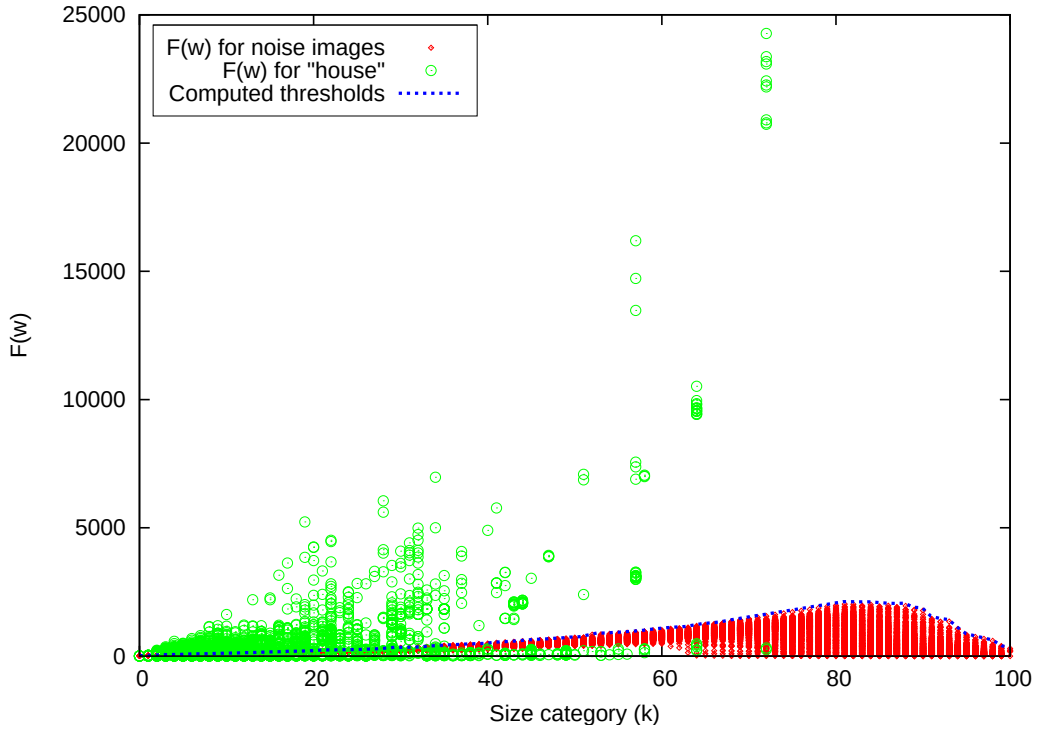
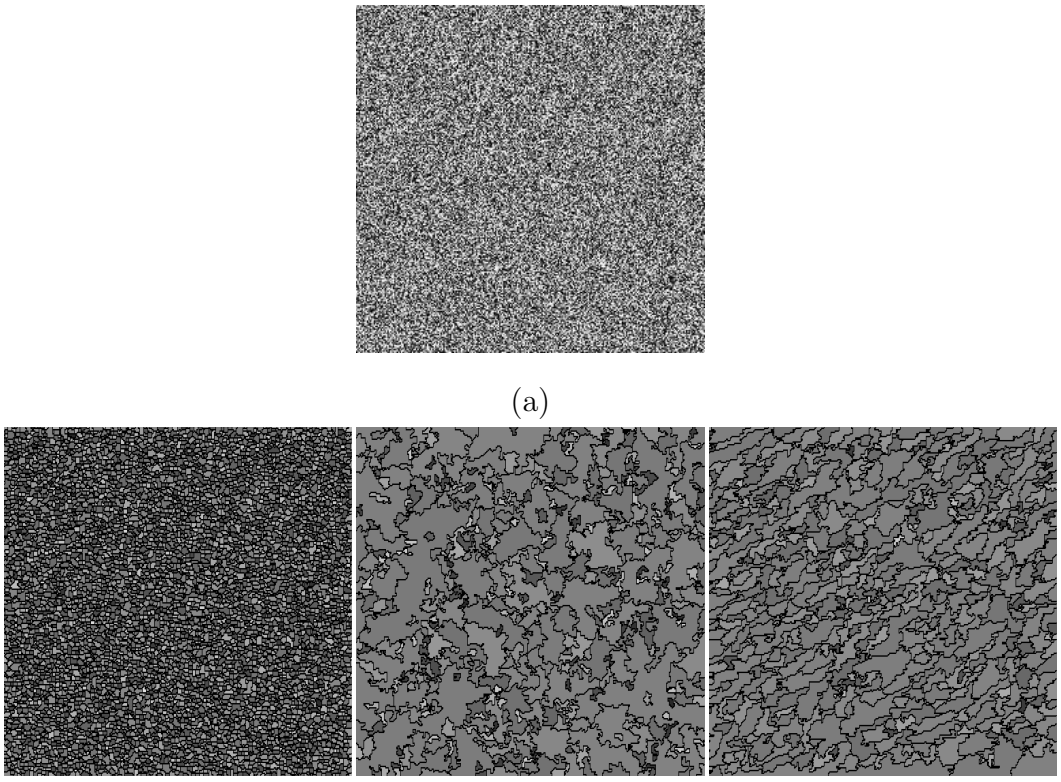


Fig. 5. F values obtained by applying Algorithm 1 with a watershed initialization. The algorithm is run on 1000 random images. For each image, the maximal F value for each size family is plotted (in red). Optimal thresholds according to Section 4 roughly correspond to the globally maximal values. F values observed in the "house" image are also depicted (black circles). Values over the threshold curve correspond to meaningfully different couples of regions. Very large deviations are observed, showing the ability of F to easily discriminate the consequences of some physical phenomena from the consequences of chance.

Some results are presented in Figures 6, 7, 8, 9, 10, 11 and 12. The watershed initialization outputs strong over-segmentations, which are well corrected by *a contrario* post-processing. However, the "best first" heuristic of Algorithm 1 is sensitive to local ambiguities and tends to create regions with complex boundaries. Better results are obtained with the EGBIS initialization, which provides better initial regions. With the parameters we chose, EGBIS keeps most details, but the counterpart is a high number of false alarms. These are efficiently filtered out by *a contrario* post-processing, while most details are still preserved. Even with a high complexity parameter Q, the SRM algorithm loses many detail, but outputs fewer false alarms. In this case, *a contrario* filtering keeps almost every region, except in the noisy cases of Figures 6 and 9.

Our *a contrario* method detects deviations from the *a contrario* models according to some distance measures, whatever the cause of the deviations. The

distance functions of Section 3 are sensitive to illumination gradients and shadows, so our method detects them. However, with an initialization discarding illumination gradients, which is the case for EGBIS, those detections are avoided, as illustrated in Figure 7. Application-specific needs can be integrated by choosing adapted distance functions or with an adapted initialization.



(a)

(b) (c) (d)

Fig. 6. (a) 256×256 uniform white noise image (b) Watershed segmentation : 6229 regions (c) EGBIS segmentation: 349 regions (d) SRM segmentation: 489 regions The three initialization algorithms we experiment produce false alarms on pure noise images. In each case, no meaningful couple of regions is kept after *a contrario* post-processing with Algorithm 1. This is a sanity check for our approach.

6 Discussion

We have presented a new methodology that brings more flexibility to previous purely analytical *a contrario* computational models. This allowed us to design a generic and statistically-founded segmentation method. Our contributions are two-fold:

- Modeling the segmentation problem *a contrario* allowed to use a difference measure between regions taking into account many parameters, e.g. region

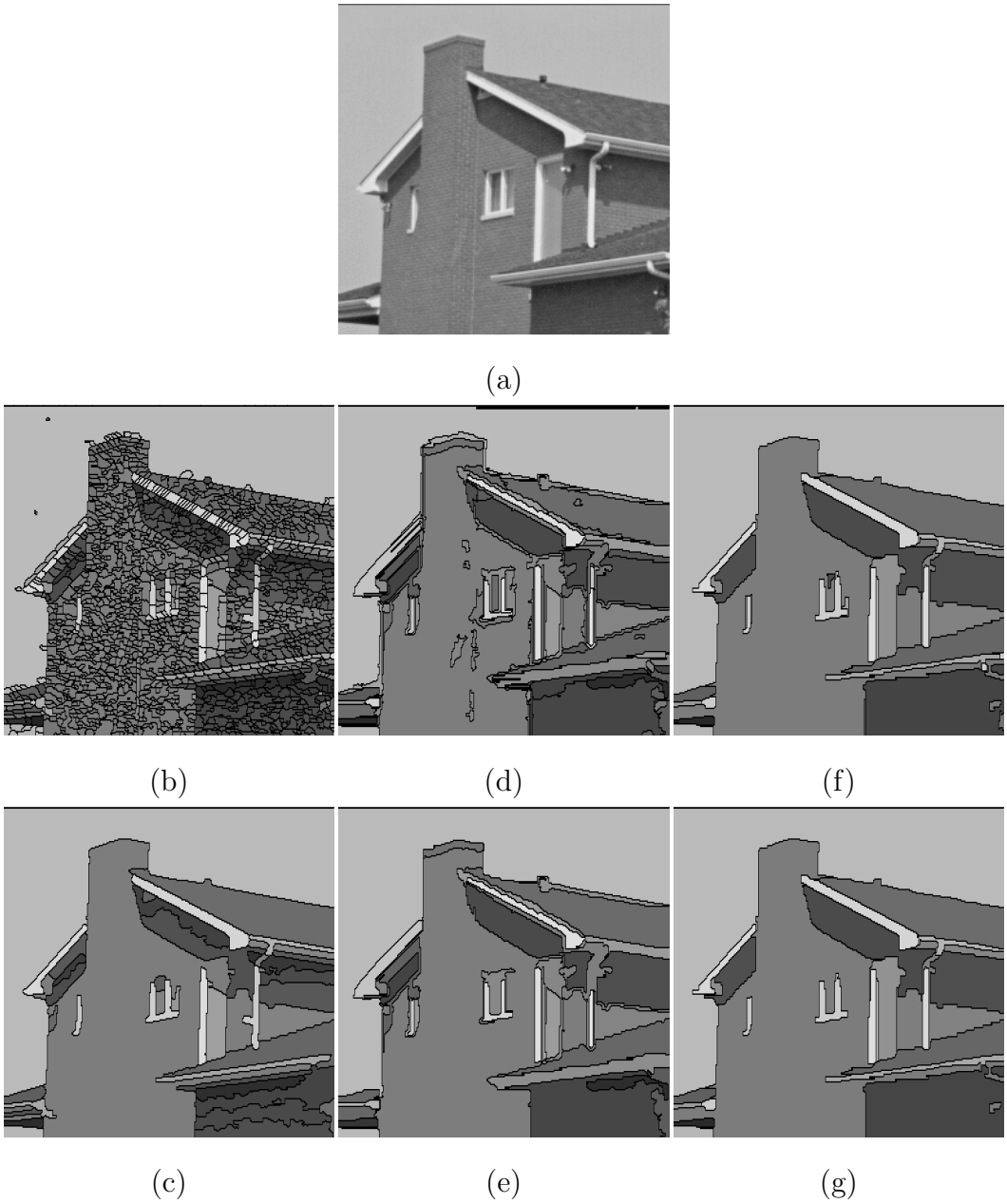


Fig. 7. (a) "House" image (b) Watershed segmentation : 1839 regions (c) After post-processing with Algorithm 1: 47 regions. (d) EGBIS segmentation: 130 regions (e) After post-processing with Algorithm 1: 34 regions. (f) SRM segmentation: 43 regions (g) After post-processing with Algorithm 1: 31 regions.

sizes, global distributions of pixels, gray level difference, variance difference and frontier contrast. These parameters and measures are thus combined in a sensible way using simple yet powerful models of chance, without requiring any quantitative *a priori*.

- Unlike previous *a contrario* works, our approach is too complex to be solved by purely analytical calculus, and led us to develop a new algorithm to automatically find the best decision thresholds, based on *a priori* simulations.

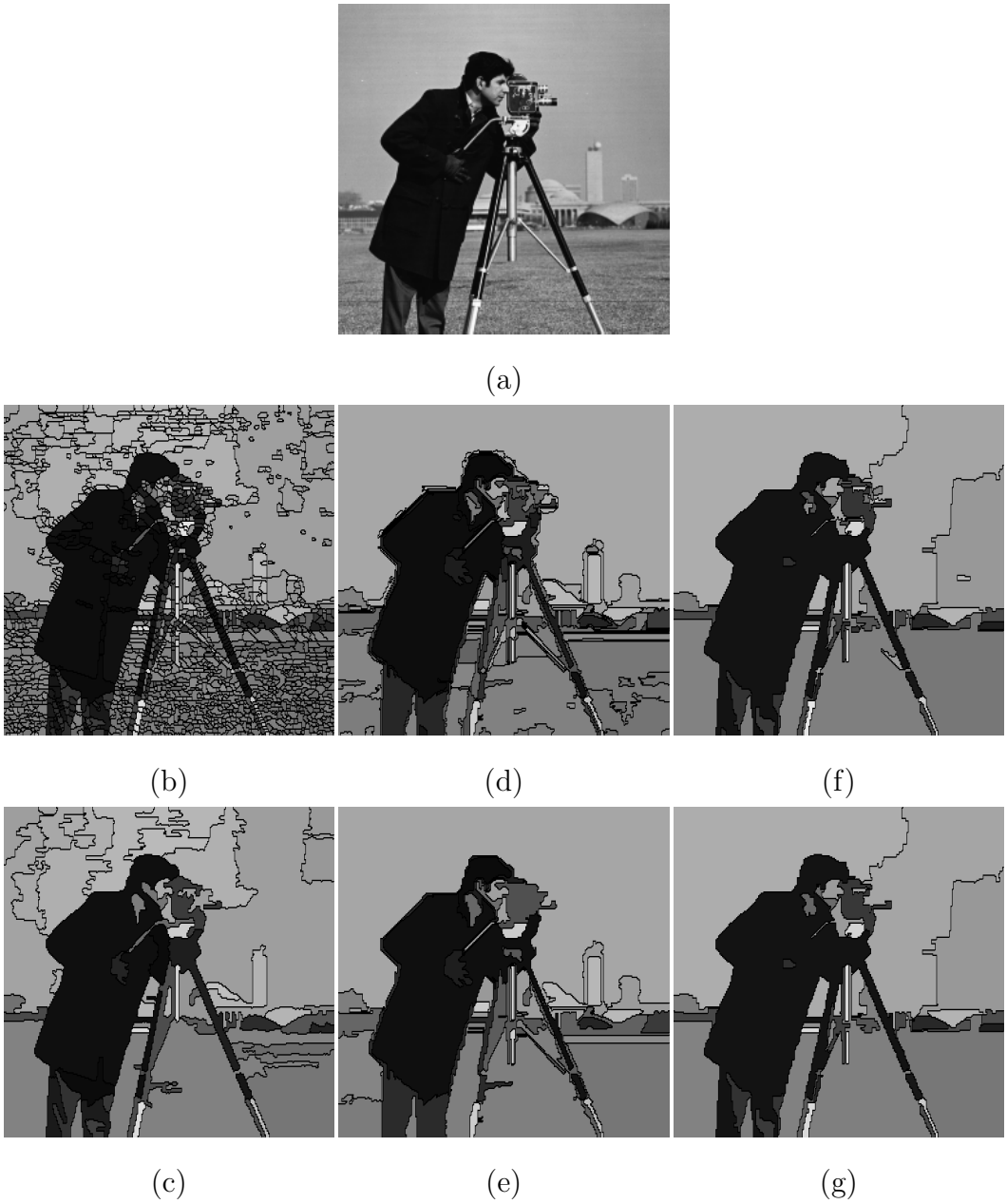


Fig. 8. (a) "Cameraman" image (b) Watershed segmentation: 1784 regions (c) After post-processing with Algorithm 1: 64 regions. (d) EGBIS segmentation: 184 regions (e) After post-processing with Algorithm 1: 36 regions. (f) SRM segmentation: 64 regions (g) After post-processing with Algorithm 1: 42 regions.

This method keeps the good properties of previous works, while bringing more flexibility.

This new paradigm is not limited to segmentation and is valuable in at least two cases:

- When several properties (distance functions in our case) have to be jointly

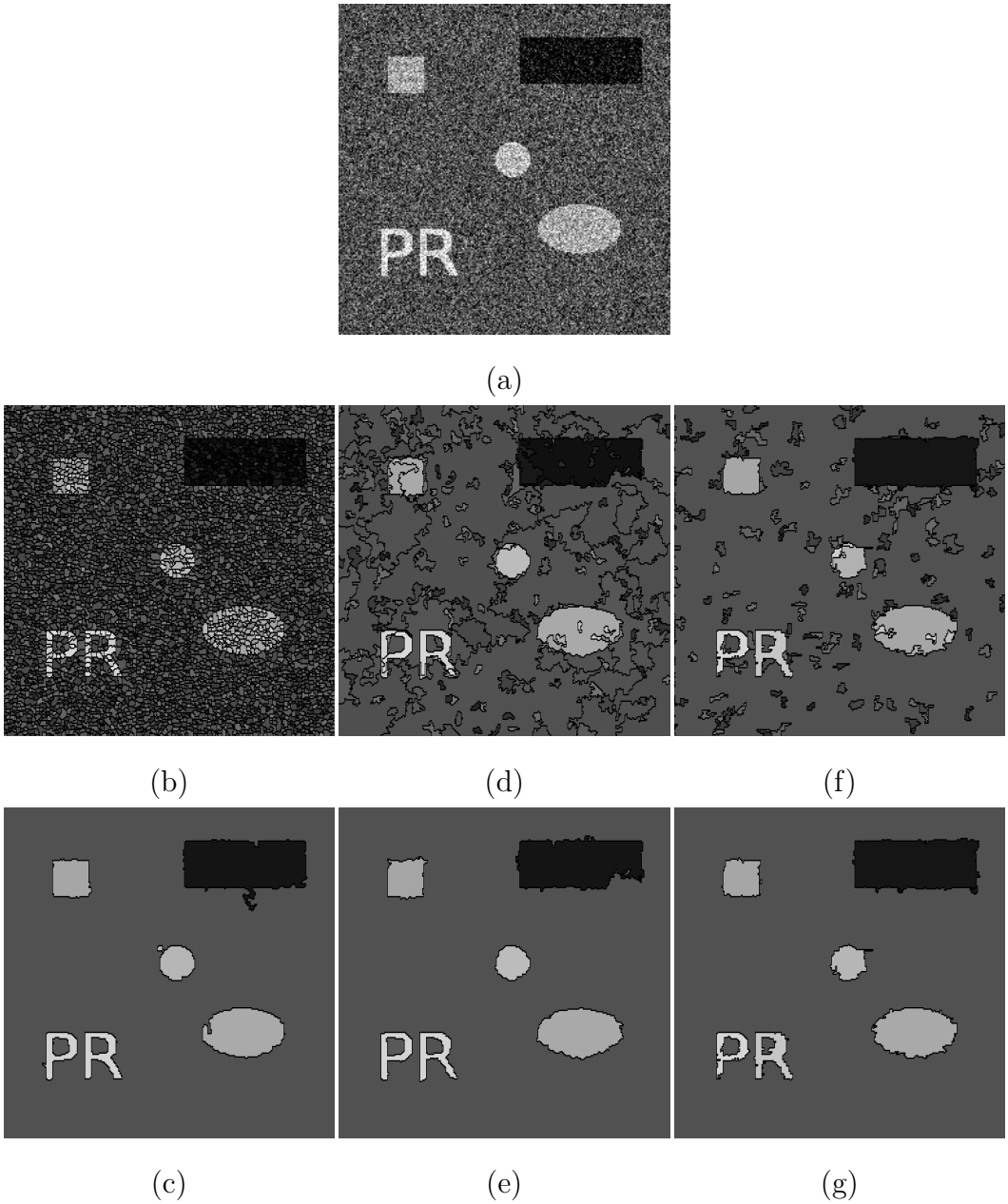


Fig. 9. (a) "Noise" image (b) Watershed segmentation: 5662 regions (c) After post-processing with Algorithm 1: 9 regions. (d) EGBIS segmentation: 232 regions (e) After post-processing with Algorithm 1: 9 regions. (f) SRM segmentation: 184 regions (g) After post-processing with Algorithm 1: 9 regions.

used, but are not statistically independent, making their joint probability difficult to compute. An approximation by their product still contains most of the information, but disable purely analytical computations.

- When entities to analyze are chosen depending on their properties, e.g. via a data-driven exploration heuristic. For example in Algorithm 1, the set of analyzed couples of regions depends on region differences.

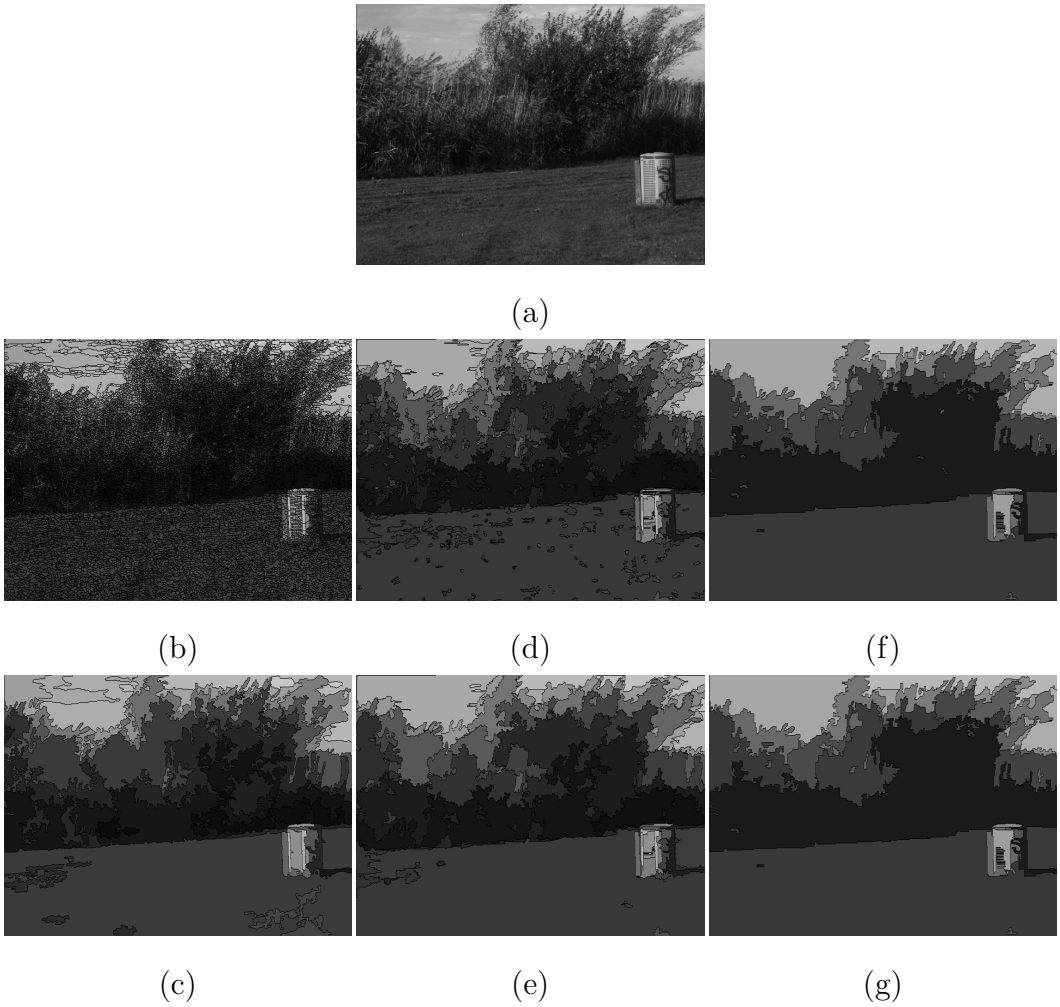


Fig. 10. (a) "Trash" image (b) Watershed segmentation: 14677 regions (c) After post-processing with Algorithm 1: 269 regions. (d) EGBIS segmentation: 682 regions (e) After post-processing with Algorithm 1: 87 regions. (f) SRM segmentation: 62 regions (g) After post-processing with Algorithm 1: 48 regions.

With these improvements we have proposed a segmentation method which can integrate many distance functions and exploration heuristics, and can be used as a post-processing filter for existing segmentation methods. In all cases, it is guaranteed that the final partitions will contain, in average, less than ε couples of regions whose differences are actually due to chance.

From an experimental point of view, results obtained with a simple merging algorithm confirm our hopes. The only assumption of our approach is that there are no meaningful differences in a set of independent and identically distributed pixels, and it is enough to automatically obtain accurate perception thresholds: experimental results show that most regions that are meaningfully different, at least for a human observer, are kept by *a contrario* filtering, while false alarms are removed. Thanks to the method flexibility, the same process could be applied to three different heuristics, corresponding to three different

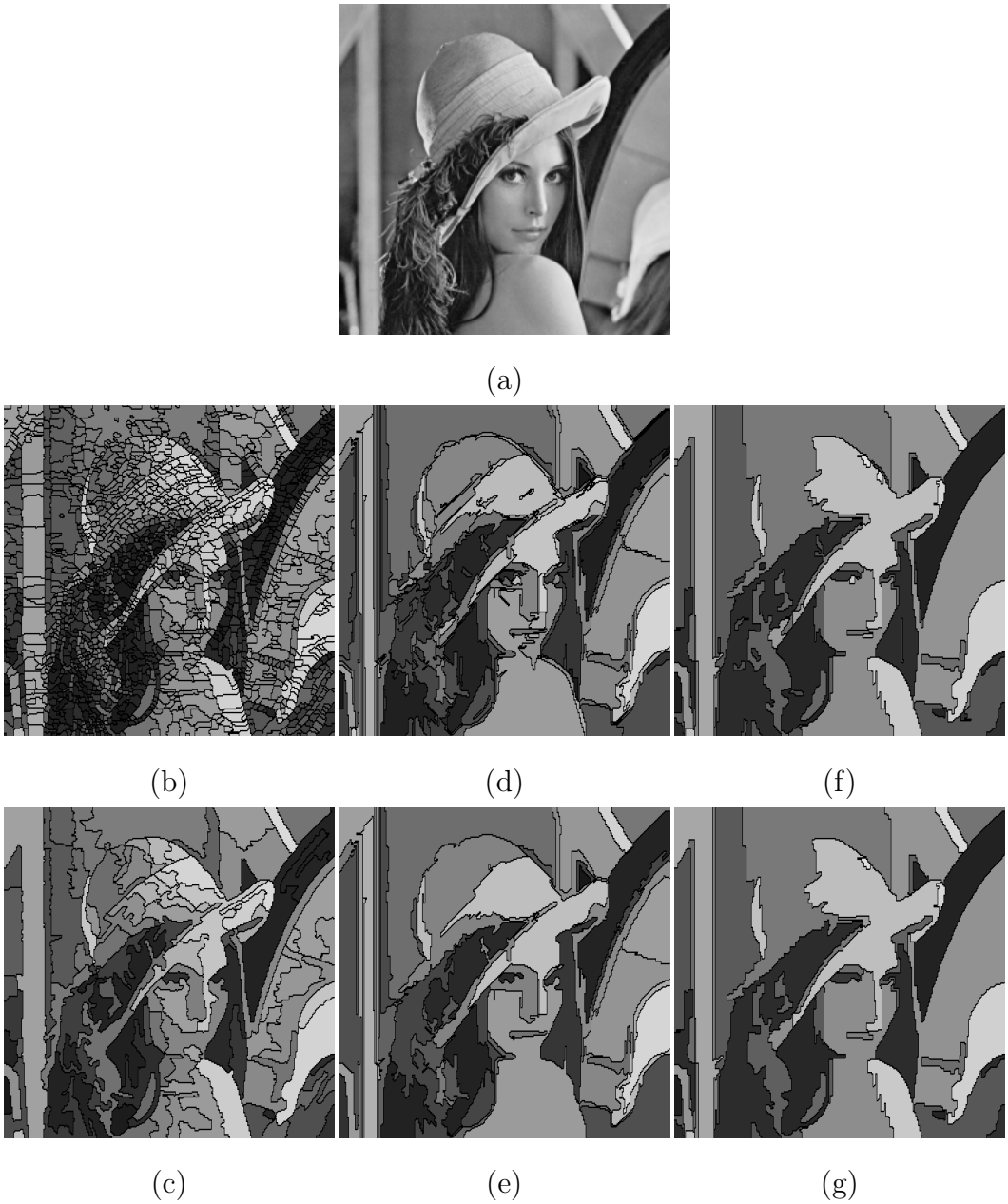


Fig. 11. (a) "Lena" image (b) Watershed segmentation: 1855 regions (c) After post-processing with Algorithm 1: 94 regions. (d) EGBIS segmentation: 205 regions (e) After post-processing with Algorithm 1: 50 regions. (f) SRM segmentation: 62 regions (g) After post-processing with Algorithm 1: 48 regions.

initializations of Algorithm 1, and similar conclusions were drawn in each case.

Many improvements are possible. First, more sophisticated exploration heuristic should bring significantly better results. In particular, the decision criterion is intrinsically well-suited to multi-scale approaches, since two meaningfully different regions may also be meaningfully different from other adjacent regions once merged. Algorithm 1 is bottom-up and never merges meaningfully

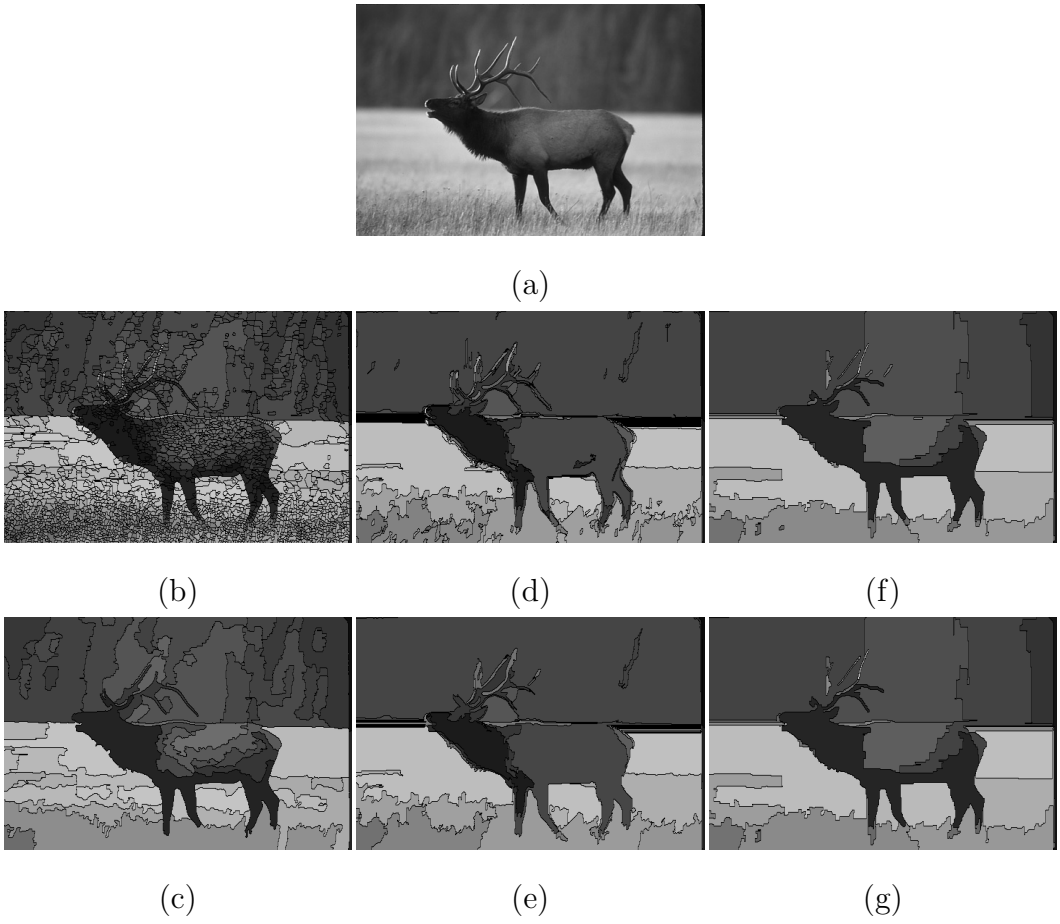


Fig. 12. (a) "Elan" image (b) Watershed segmentation: 3191 regions (c) After post-processing with Algorithm 1: 234 regions. (d) EGBIS segmentation: 234 regions (e) After post-processing with Algorithm 1: 50 regions. (f) SRM segmentation: 44 regions (g) After post-processing with Algorithm 1: 36 regions.

different regions. Thus, it provides partitions keeping as many details as possible. It would be possible to continue the merging process though, and build a tree of meaningfully different regions. In addition, by comparing the significance of meaningful couples of regions, it would become possible to directly extend the notion of maximal significance of [21] and automatically select the most significant scale for a set of regions.

Second, the actual range of detected physical phenomena depends on the ability of the chosen measures to reflect their particular properties. Choosing illumination gradient independent measures could be more relevant for some applications. Additional measures could also be performed to bring more discriminative power, such as texture analysis, shape convexity or regularity. The measures could also be extended rather directly to color distributions.

Finally, the proposed methodology is by no mean restricted to image segmentation. The flexibility brought by the simulation-based thresholds estimation

greatly simplifies the modeling work and enables new problems to be easily solved *a contrario*. In a simplified and not yet formalized version, it was already applied by the authors to meaningful segment extraction [27] and we believe that other applications could advantageously use it.

A Proof of Proposition 2

Let $W = \{w_1, w_2, \ldots, w_n\}$ be the set of all possible adjacent couples of regions in a $N \times M$ image. $\#W$, the cardinal of W, is very big and not known analytically, but only depends on N and M. We consider an image as an observation of some image model, e.g. a uniform white noise model, and by extension, each region of an image is seen as an observation of some regional probabilistic model. Let us introduce the following notations:

- Θ_{ac} denotes the independent and identically distributed (i.i.d.) family of models for an image. When W is extracted from an observation of an i.i.d. model, it is noted by $\Theta_{\text{ac}}(W)$. Θ_{acu} denotes the uniform white noise model for an image (case where the distribution is uniform). Thus, Θ_{acu} is a subcase of Θ_{ac}. When W is extracted from an observation of uniform white noise, it is noted by $\Theta_{\text{acu}}(W)$.
- θ_{ac} denotes the family of *a contrario* models for couples of regions, and θ_{acu} denotes the uniform white noise model for couples of regions. Thus, θ_{acu} is a subcase of θ_{ac}. When a couple of regions w_i is an observation of an *a contrario* model, it is noted $\theta_{\text{ac}}(w_i)$, and when it is an observation of the uniform white noise model, it is noted $\theta_{\text{acu}}(w_i)$.
- $\mathcal{H}(w_i)$ means that a couple of regions w_i is analyzed by the heuristic $\mathcal{H}$.
- $S_\delta(w_i)$ means that a couple of regions w_i is meaningfully different according to S_δ.
- $\mathbb{E}(X)$ denotes the expectation of X.
- $\#\Omega$ denotes the cardinal of a set Ω.
- The expectation of the number of false alarms made by the algorithm $\mathcal{A}$ on W is noted:

$$\mathbb{E}_W = \mathbb{E}\left(\#\left\{w_i \in W; \mathcal{H}(w_i) \text{ and } S_\delta(w_i) \text{ and } \theta_{\text{ac}}(w_i)\right\}\right)$$

  This is the expected number of couples of regions analyzed by $\mathcal{H}$ and considered as meaningfully different by S_δ, whereas they were actually observations of an *a contrario* model.

Let us recall and reformulate the conditions of Proposition 2:

(1) $\mathcal{A}$ is ε-reliable for uniform white noise images:

$$\mathbb{E}_W \mid \Theta_{\text{acu}}(W) < \varepsilon$$

(2) $P(S_\delta(w_i) \mid \theta_{\text{ac}}(w_i)) \leq P(S_\delta(w_i) \mid \theta_{\text{acu}}(w_i))$
(3) $P(\mathcal{H}(w_i) \mid S_\delta(w_i), \theta_{\text{ac}}(w_i)) \leq P(\mathcal{H}(w_i) \mid S_\delta(w_i), \theta_{\text{acu}}(w_i))$
(4) $P(\mathcal{H}(w_i) \mid S_\delta(w_i), \Theta_{\text{ac}}(W)) = P(\mathcal{H}(w_i) \mid S_\delta(w_i), \theta_{\text{ac}}(w_i))$ and in particular $P(\mathcal{H}(w_i) \mid S_\delta(w_i), \Theta_{\text{acu}}(W)) = P(\mathcal{H}(w_i) \mid S_\delta(w_i), \theta_{\text{acu}}(w_i))$

We want to show that under these conditions, $\mathcal{A}$ is ε-reliable for any kind of images, i.e. that $\mathbb{E}_W < \varepsilon$. Let X_i be the Bernoulli variable equal to one when there is a false alarm, i.e. when $\mathcal{H}(w_i)$, $S_\delta(w_i)$ and $\theta_{\text{ac}}(w_i)$ simultaneously occur. Using expectation linearity:

$$\begin{aligned}\mathbb{E}_W \mid \Theta_{\text{acu}}(W) &= \mathbb{E}\left(\sum_{i=1}^{\#W} X_i \mid \Theta_{\text{acu}}(W)\right) \\ &= \sum_{i=1}^{\#W} P_{\text{acu}}^{X_i}\end{aligned}$$

where $P_{\text{acu}}^{X_i} = P(X_i = 1 \mid \Theta_{\text{acu}}(W)) = P(\mathcal{H}(w_i), S_\delta(w_i), \theta_{\text{ac}}(w_i) \mid \Theta_{\text{acu}}(W))$.

Since W is extracted from an observation of uniform white noise, we know that every couple of regions is a result of a uniform white noise model, and thus of an *a contrario* model, so:

$$\begin{aligned}P_{\text{acu}}^{X_i} &= P(\mathcal{H}(w_i), S_\delta(w_i) \mid \Theta_{\text{acu}}(W)) \\ &= P(\mathcal{H}(w_i) \mid S_\delta(w_i), \Theta_{\text{acu}}(W)) \times P(S_\delta(w_i) \mid \Theta_{\text{acu}}(W))\end{aligned}$$

Using condition (4) and then (3):

$$\begin{aligned}P(\mathcal{H}(w_i) \mid S_\delta(w_i), \Theta_{\text{acu}}(W)) &= P(\mathcal{H}(w_i) \mid S_\delta(w_i), \theta_{\text{acu}}(w_i)) \\ &\geq P(\mathcal{H}(w_i) \mid S_\delta(w_i), \theta_{\text{ac}}(w_i))\end{aligned}$$

Knowing the probabilistic model of a couple of regions, the observed differences do not depend upon the global image model:

$$P(S_\delta(w_i) \mid \Theta_{\text{acu}}(W)) = P(S_\delta(w_i) \mid \theta_{\text{acu}}(w_i))$$

Using condition (2):

$$P(S_\delta(w_i) \mid \theta_{\text{acu}}(w_i)) \geq P(S_\delta(w_i) \mid \theta_{\text{ac}}(w_i))$$

Thus:

$$P_{\text{acu}}^{X_i} \geq P(\mathcal{H}(w_i) \mid S_\delta(w_i), \theta_{\text{ac}}(w_i)) \times P(S_\delta(w_i) \mid \theta_{\text{ac}}(w_i))$$

Besides:

$$\mathbb{E}_W = \sum_{i=1}^{\#W} P(X_i = 1)$$

And:

$$\begin{aligned} P(X_i = 1) &= P(\mathcal{H}(w_i), S_\delta(w_i), \theta_{\text{ac}}(w_i)) \\ &= P(\mathcal{H}(w_i) \mid S_\delta(w_i), \theta_{\text{ac}}(w_i)) \\ &\quad \times P(S_\delta(w_i) \mid \theta_{\text{ac}}(w_i)) \\ &\quad \times P(\theta_{\text{ac}}(w_i)) \end{aligned}$$

where the *a priori* probability $P(\theta_{\text{ac}}(w_i))$ is unknown, but necessarily less than one. Then:

$$\begin{aligned} P(X_i = 1) &< P(\mathcal{H}(w_i) \mid S_\delta(w_i), \theta_{\text{ac}}(w_i)) \times P(S_\delta(w_i) \mid \theta_{\text{ac}}(w_i)) \\ &< P_{\text{acu}}^{X_i} \\ &< P(X_i = 1 \mid \Theta_{\text{acu}}(W)) \end{aligned}$$

It follows that $\mathbb{E}_W < \mathbb{E}_W \mid \Theta_{\text{acu}}(W)$ and, according to condition (1), $\mathbb{E}_W < \varepsilon$. □

In conclusion, the four conditions of Proposition 2 are sufficient to guarantee the ε-reliability we look for.

References

[1] S. Shah, Performance modeling and algorithm characterization for robust image segmentation, International Journal of Computer Vision To appear.

[2] N. Pal, S. Pal, A review on image segmentation techniques, Pattern Recognition 26 (9) (1993) 1277–1294.

[3] L. Lucchese, S. Mitra, Color Image Segmentation: A State-of-the-Art Survey, Proceedings of the Indian National Science Academy 67 (2) (2001) 207–221.

[4] S. Zhu, A. Yuille, Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (9) (1996) 884–900.

[5] Z. Tu, S. Zhu, Image segmentation by data-driven markov chain monte carlo, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (5) (2002) 657–673.

[6] D. Mumford, J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems, Communications in Pure and Applied Mathematics 42 (5) (1989) 577–685.

[7] J. Shi, J. Malik, Normalized Cuts and Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 888–905.

[8] J. Kim, K. Hong, A new graph cut-based multiple active contour algorithm without initial contours and seed points, Machine Vision and Applications 19 (3) (2008) 181–193.

[9] Y. Pan, J. Birdwell, S. Djouadi, Bottom-up hierarchical image segmentation using region competition and the mumford-shah functional, in: Proceedings of the 18th International Conference on Pattern Recognition, Vol. 2, 2006, pp. 117–121.

[10] P. Felzenszwalb, D. Huttenlocher, Efficient Graph-Based Image Segmentation, International Journal of Computer Vision 59 (2) (2004) 167–181.

[11] V. Rehrmann, L. Priese, Fast and robust segmentation of natural color scenes, in: Proceedings of the 3rd Asian Conference on Computer Vision, 1998, pp. 598–606.

[12] R. Nock, F. Nielsen, Statistical region merging, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (11) (2004) 1452–1458.

[13] D. Comaniciu, P. Meer, Mean shift: A robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 603–619.

[14] T. Adamek, N. E. O'Connor, Semantic Multimedia, Springer, 2007, Ch. Stopping Region-Based Image Segmentation at Meaningful Partitions, pp. 15–27.

[15] A. Desolneux, L. Moisan, J.-M. Morel, Edge detection by Helmholtz principle, Journal of Mathematical Imaging and Vision 14 (3) (2001) 271–284.

[16] A. Desolneux, L. Moisan, J.-M. Morel, A grouping principle and four applications, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (4) (2003) 508–513.

[17] T. Veit, F. Cao, P. Bouthemy, An a contrario Decision Framework for Region-Based Motion Detection, International Journal of Computer Vision 68 (2) (2006) 163–178.

[18] P. Musé, F. Sur, F. Cao, Y. Gousseau, J.-M. Morel, An a contrario decision method for shape element recognition, International Journal of Computer Vision 69 (3) (2006) 295–315.

[19] Y. Chang, X. Li, Adaptive image region-growing, IEEE Transactions on Image Processing 3 (6) (1994) 868–872.

[20] L. Wolf, X. Huang, I. Martin, D. Metaxas, Patch-based texture edges and segmentation, in: Proceedings of the European Conference on Computer Vision, Vol. 2, Springer, 2006, pp. 481–493.

[21] A. Desolneux, L. Moisan, J.-M. Morel, Meaningful Alignments, International Journal of Computer Vision 40 (1) (2000) 7–23.

[22] H. Mann, D. Whitney, On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other, The Annals of Mathematical Statistics 18 (1) (1947) 50–60.

[23] G. Saporta, Probabilités, analyse des données et statistique, Editions Technip, 1990.

[24] R. Grompone, J. Jakubowicz, J.-M. Morel, G. Randall, LSD: A line segment detector, CMLA preprint (2008-15).

[25] L. Igual, J. Preciozzi, L. Garrido, A. Almansa, V. Caselles, B. Rouge, Automatic Low Baseline Stereo In Urban Areas, Inverse Problems and Imaging 1 (2) (2007) 319–348.

[26] S. Beucher, F. Meyer, The morphological approach to segmentation: the watershed transformation, Mathematical Morphology in Image Processing (1993) 433–481.

[27] N. Burrus, T. M. Bernard, Adaptive Vision Leveraging Digital Retinas: Extracting Meaningful Segments, Advanced Concepts for Intelligent Vision Systems, 8th International Conference (2006) 220–231.