Image denoising by statistical area thresholding

D. Coupier, A. Desolneux, and B. Ycart

January 29, 2004

MAP5, FRE CNRS 2428, Université René Descartes, Paris

E-Mail addresses : {coupier,desolneux,ycart}@math-info.univ-paris5.fr

Corresponding author : A. Desolneux
Mail address : MAP5, UFR Math-Info,
Université Paris 5, 45 rue des Saints-Pères
75270 PARIS Cedex 06, FRANCE.
Telephone : 33 1 44 55 35 26
Fax : 33 1 44 55 35 35

Abstract

Area openings and closings are morphological filters which efficiently suppress impulse noise from an image, by removing small connected components of level sets. The problem of an objective choice of threshold for the area remains open. Here, a mathematical model for random images will be considered. Under this model, a Poisson approximation for the probability of appearance of any local pattern can be computed. In particular, the probability to observe a component with size larger than k in pure noise has an explicit form. This permits to define a statistical test on the significance of connected components, thus providing an explicit formula for the area threshold of the denoising filter, as a function of the noise intensity. Finally, using threshold decomposition, an adaptive denoising algorithm for gray level images is proposed.

Key words : image denoising, mathematical morphology, area opening and closing, random image, threshold function, Poisson approximation, lattice animals.

AMS Subject Classification : 68U10, 62H35

1 Introduction

The general problem of image denoising consists of deciding what comes from the image and has to be kept, and what comes from the noise, and must be removed. Many different criteria can be used to detect the noise-induced structures. For example, the oscillations due to an additive gaussian noise can be measured in terms of the wavelet coefficients. The noise may then be removed by a thresholding in the wavelet domain. Donoho and Johnstone [6] gave an explicit way to choose the threshold as a function of the variance of the noise. Their claim is that "denoising, with high probability, rejects pure noise completely". The underlying idea is that in a pure noise, all the structures that actually belong to the image could not appear; or else, the structures coming from the image itself can be defined as those "objects" which would have a very small probability to appear in a pure noise. This idea was implemented in [4] and [5] for the detection of alignments and meaningful level lines in an image.

Here, we shall focus on the size of connected components of the level sets of the image. Removing small components is a classical and efficient way to remove impulse noise from an image. This method, known as "the grain filter", was first introduced in the framework of Mathematical Morphology [15] by Vincent in [19] as morphological area openings and closings (see also [20] and [7]). This filter is sometimes called the "extrema killer". It was then generalized by Masnou and Morel in [11], and by Monasse in [13]. In [14] a similar filter was used, in the framework of gradient percolation, for recovering fuzzy images.

But the main question remains: how should the threshold for the area of the components that have to be kept, be chosen? A natural idea, imported from statistical inference, consists in fixing an a priori risk level ε (e.g. $\varepsilon = 0.001$), and deciding that anything that has probability lower than ε of occuring under a pure noise hypothesis cannot come from the noise and hence should be kept in the image. Thus for the threshold area, one will choose the integer $s(n, p, \varepsilon)$, such that a connected component of size $k \geqslant s(n, p, \varepsilon)$ has a probability less than ε to appear in a pure noise image with intensity p and size $n \times n$. Applying a grain filter with area threshold $s(n, p, \varepsilon)$ will ensure that, with probability larger than $1 - \varepsilon$, pure noise is eliminated. To implement this, one must be able to compute the probability for any connected component of size k to appear in a pure noise image. An exact computation is not feasible. However an approximation can be given if the image is large: our main theoretical result (Theorem 2.4) gives a Poisson approximation for the probability of occurrence for any image property which is local in the sense that its definition involves only a fixed number of connected pixels.

Our plan is the following. Section 2 is devoted to the probabilistic model of noise on binary images: all pixels are independent, black with probability p or white with probability $1-p$. The Poisson approximation result will be stated (Theorem 2.4) and an outline of its proof will be given (technical details will be postponed to the Appendix). Section 3 is devoted to applications. We will first explain how Theorem 2.4, together with combinatorial results on lattice animals, can be used to obtain an explicit formula for the size threshold $s(n, p, \varepsilon)$. An example of denoising for a binary image will be given. Then we shall extend the method to gray level images through threshold decomposition: the binary image corresponding to each gray level is treated separately, then all denoised binary images are recombined. Some experiments and a discussion of the obtained results come last.

2 Probability of a local property

Our probabilistic model for random images is the following. Let n be a positive integer. Consider the *pixel set* $\Xi_n = \{1, \ldots, n\}^2$. A *binary image of size* n is a mapping from Ξ_n to $\{0, 1\}$ (black/white). Their set is denoted by E_n. It is endowed with the probability distribution $\mu_{n,p}$ defined by: each pixel is black with probability p or white with probability $1-p$, and all the pixel colors are independent. A *random image of size* n *and intensity* p, denoted by $\mathcal{I}_{n,p}$, is a random element of E_n with probability distribution $\mu_{n,p}$.

The pixel set Ξ_n is embedded in $\mathbb{Z}^2$ and naturally endowed with a *graph structure*. We consider in this paper the case of the 4-connectivity (2 horizontal and 2 vertical neighbors). For purely technical reasons, it will be convenient that all pixels have the same neighborhood: this is why we impose periodic boundary conditions, deciding that $(1, j)$ is neighbor with (n, j) and $(j, 1)$ with (j, n). Thus the graph is a 2-dimensional torus. As usual, the *graph distance* d is defined as the minimal length of a path between two pixels. We shall denote by $B(x, r)$ the ball of center x and radius r:

$$B(x, r) = \{y \in \Xi_n;\ d(x, y) \leqslant r\} .$$

For $r < n/2$, $B(x, r)$ is a lozenge containing $2r^2 + 2r + 1$ pixels. For the rest of this section, the radius r is a fixed integer, and the image size n is larger than $2r$.

The image properties we are interested in are all local, in the sense that they can be described inside balls of a fixed radius. All balls are translations of each other. We shall choose a ball of radius r, say $B(0, r)$, and fix a translation τ_x, from $B(0, r)$ to $B(x, r)$ for all x. We call *pattern*, and denote by D, an image defined on $B(0, r)$, and determined by its set of black pixels, denoted by $\beta(D)$. Of course, $B(0, r) \setminus \beta(D)$ is the set of white pixels. We shall denote by $b(D)$ the cardinality of $\beta(D)$ (number of black pixels in the pattern). We shall deal with rather small levels of noise, seen as relatively sparse black pixels on a white background. This is of course a mere convention: swapping black and white, together with p and $1-p$ does not change the model. Thus, in what follows, we will always assume that $p \leqslant \frac{1}{2}$.

If D is a pattern on $B(0, r)$ and τ is a translation of pixels, we shall denote by $\tau(D)$ the pattern on $B(\tau(0), r)$, whose set of black pixels is $\tau(\beta(D))$. If $\tau(0) = x$, we denote by $D(x)$ the property: "the restriction of the image to $B(x, r)$ is $\tau(D)$". The property we are actually interested in is

$$\widetilde{D} = (\exists x \in \Xi_n\ ,\ D(x)) .$$

In other words $\widetilde{D}$ means: "a copy of pattern D can be found somewhere in the image".

The patterns D are the building blocks of all local properties. Indeed, there exists only a finite number of such patterns (precisely 2^{2r^2+2r+1}): let us denote their set by $\mathcal{D}$. Any *assertion* relative to the pixels in $B(0,r)$ will be called "local": it can be expressed in a unique way as a disjunction (logical "or", denoted by $\vee$) of distinct patterns. The following definitions will be used in the counting of occurrences of a local property in an image.

Definition 2.1 *Let ψ be a local assertion, relative to the pixels in $B(0,r)$.*

1. *The* definition set *of ψ, denoted by $\mathcal{D}(\psi)$, is the subset of $\mathcal{D}$ such that*

$$\psi = \bigvee_{D \in \mathcal{D}(\psi)} D \ .$$

2. *The* black index $b(\psi)$ *of ψ is the integer $b(\psi)$ defined by*

$$b(\psi) = \min_{D \in \mathcal{D}(\psi)} \{b(D)\} \ .$$

3. *A* meaningful definition set *of ψ, denoted by $\mathcal{D}_0(\psi)$, is a subset of $\mathcal{D}(\psi)$ such that*

   (a) $\forall D \in \mathcal{D}_0(\psi) \ , \quad b(D) = b(\psi) \ ,$

   (b) If τ is a translation, then $D, D' \in \mathcal{D}_0(\psi)$ and $\tau(\beta(D)) = \beta(D')$ imply $D = D'$,

   (c) $D \in \mathcal{D}(\psi)$ and $b(D) = b(\psi)$ imply $\exists \tau \, , \exists D' \in \mathcal{D}_0(\psi) \, ,$ s. t. $\tau(\beta(D)) = \beta(D')$.

   All meaningful definition sets have the same cardinality, which will be called the meaningful index *of ψ, and denoted by $e(\psi)$.*

The black index $b(\psi)$ is the minimal number of black pixels, in a pattern that satisfies ψ. One can see the meaningful index $e(\psi)$ as the maximal number of patterns with exactly $b(\psi)$ black pixels that satisfy ψ, up to possible translations. Both will be used to count occurrences of the *local property* based on ψ.

Definition 2.2 *Let ψ be a local assertion, and $\psi(x)$ its localization on the ball centered at x :*

$$\psi(x) = \bigvee_{D \in \mathcal{D}(\psi)} D(x) \ .$$

We call local property *based on ψ, and denote by $\widetilde{\psi}$ the property*

$$\widetilde{\psi} = (\exists x \, , \ \psi(x)) \ .$$

Our basic example of a local property $\widetilde{\psi}$ is: "there exists a connected component of k black pixels". A connected component of size k is always included in a ball of radius $r \geqslant k/2$. The local assertion ψ is "there exists a connected component of size k in $B(0,r)$". The description set is the set of all patterns on $B(0,r)$, having at least k connected black pixels. The black index is the minimal number of black pixels necessary for ψ to be satisfied (obviously k in our example). The meaningful index is the number of connected components of size k, up to translations (see Section 3).

The size n of the random image $\mathcal{I}_{n,p}$ tends to infinity. For a fixed level p with $0 < p < 1$, by the independence of pixels, it is easy to see that asymptotically any pattern will be present in a random image with a probability tending to 1 (see [3] for more precise results). Therefore the asymptotic probability for the random image $\mathcal{I}_{n,p}$ to satisfy $\widetilde{\psi}$ is 1, whatever ψ. That asymptotic probability can be different from 1 only if $p = p(n)$ tends to 0 as n tends to infinity. Thus our images will have a relatively small proportion of black pixels.

A classical object of the theory of random graphs (see [1, 18] as general references), is the notion of *threshold function*. It describes the appearance of a given subgraph in a random graph. The notion of threshold function easily adapts to random images. Let $\mathcal{A}$ be an image property. The function $\theta(n)$ is called a threshold function of $\mathcal{A}$ if for $p(n) \leqslant 1/2$ then

$$\lim_{n\to\infty} \frac{p(n)}{\theta(n)} = 0 \Longrightarrow \lim_{n\to\infty} \mu_{n,p(n)}(\mathcal{A}) = 0 \ ,$$

and

$$\lim_{n\to\infty} \frac{p(n)}{\theta(n)} = \infty \Longrightarrow \lim_{n\to\infty} \mu_{n,p(n)}(\mathcal{A}) = 1 \ .$$

Notice that a threshold function is not unique. For instance if $\theta(n)$ is a threshold function for $\mathcal{A}$, then so is $c\theta(n)$ for any positive constant c. It is customary to ignore this and talk about "the" threshold function of $\mathcal{A}$. The following lemma was proved in [3].

Lemma 2.3 *The threshold function of the local property $\widetilde{\psi}$ is $n^{-\frac{2}{b(\psi)}}$.*

In other words, the appearance of a local property mainly depends on its black index: if $p(n)$ is small compared to $n^{-\frac{2}{b}}$, then the probability of any local property that needs b black pixels to be satisfied is small. If $p(n)$ is large compared to $n^{-\frac{2}{b}}$, then the probability is large. The particular case $b(\psi) = 0$ corresponds to the appearance of a white ball. For $p(n) \leqslant \alpha < 1$, its probability of being present in the random image always tends to 1: there is no threshold function. From now on, we will always assume that the black index of ψ is positive.

Lemma 2.3 suggests that the correct scaling for $p(n)$ when one studies a local property $\widetilde{\psi}$ is $p(n) = cn^{-\frac{2}{b(\psi)}}$. Our main result shows that with this scaling, the probability of $\widetilde{\psi}$ in a random image converges to a non trivial limit.

Theorem 2.4 *Let ψ be an assertion on $B(0, r)$, with black index $b(\psi)$ and meaningful index $e(\psi)$. Let $p(n) = cn^{-\frac{2}{b(\psi)}}$, where c is a positive constant. Then*

$$\lim_{n\to\infty} \mu_{n,p(n)}(\widetilde{\psi}) = 1 - \exp(-e(\psi)c^{b(\psi)}) \ . \tag{1}$$

The reason why such a result is called a Poisson approximation becomes clear if one considers the property "there exists a black pixel". Let X_n be the total number of black pixels. Since all pixels are independent, the random variable X_n follows the binomial distribution with parameters n^2 and $p(n)$. In particular the probability that there exists a black pixel is

$$\mathbb{P}[X_n > 0] = 1 - (1 - p(n))^{n^2} \ .$$

Here the black index is 1 and the threshold function is n^{-2}. Take $p(n) = cn^{-2}$. Then the binomial distribution of X_n converges to the Poisson distribution with parameter c, and the probability that there exists a black pixel ($X_n > 0$) tends to $1 - \exp(-c)$.

The situation is not so simple as soon as the black index is larger than 1. Consider for instance $\widetilde{\psi}$: "there exist two connected black pixels". On the ball of radius $r = 1$, the definition set is composed of all those patterns on $B(0,1)$ whose center is black, and at least one of the 4 neighbors is also black (15 patterns). Consider the number of occurrences of any of those patterns, somewhere in the random image. It is a sum of Bernoulli random variables. However they are not independent: patterns on balls centered at two adjacent pixels have one pixel in common. The same can be said of any local property $\widetilde{\psi}$: the number of occurrences of $\psi(x)$ can be viewed as a sum of (dependent) Bernoulli random variables. The sum of a large number of Bernoulli r.v.'s converges in distribution to a Poisson distribution, provided the dependencies between the variables are not too large. In the theory of random graphs, similar results are frequent (see e.g. [18] Lecture 1 p.296, Lecture 2 p.303 or Lecture 5 p.314).

Proof of Theorem 2.4: There are several ways to prove a Poisson approximation result. We chose the famous "moment method" based on the following result ([1], Chapter 1 p.25).

Lemma 2.5 *Let $(X_n)_{n\in\mathbb{N}^*}$ be a sequence of integer valued, nonnegative random variables and λ be a strictly positive real. For all $n, l \in \mathbb{N}^*$ define the quantity*

$$E_l(X_n) = \sum_{k \geqslant l} \mathbb{P}(X_n = k) \frac{k!}{(k-l)!} \ .$$

If, for all $l \in \mathbb{N}^*$, $\lim_{n\to\infty} E_l(X_n) = \lambda^l$ then (X_n) converges in distribution to the Poisson distribution with parameter λ.

In our case, X_n counts the number of occurrences in the random image of some patterns, to be precised later. The "moment" $E_l(X_n)$ is the expected number of ordered l-tuples of occurrences of those patterns.

Firstly, one should observe that patterns in the definition set of ψ cannot be all treated equally: since $p(n) = cn^{-\frac{2}{b(\psi)}}$, by Lemma 2.3 any pattern with more than $b(\psi)$ black pixels has a vanishing probability of being observed. Hence we can reduce the set of patterns to those having exactly $b(\psi)$ black pixels. In the example of two connected pixels with $r = 1$, $\mathcal{D}(\psi)$ has 15 different patterns, but only 4 of them have exactly 2 black pixels.

Now, one has to take care of multiple counts. Among the 4 patterns on $B(0,1)$ that have 2 black pixels, 2 patterns have two horizontal black neighbors, and the 2 other patterns have two vertical black neighbors. Assume the image has only one occurrence of two horizontal black neighbors. If we examine all possible pixels x, we will find two adjacent centers for which $\psi(x)$ is satisfied. In order to obviate this problem, we need to count patterns *up to possible translations*. We say that two patterns with black index $b(\psi)$ are equivalent if their sets of black pixels are translations of each other. The number of equivalence classes is the meaningful index $e(\psi)$ of Definition 2.1. (In the example of two connected black pixels, there are two equivalence classes: horizontal or vertical neighbors).

We choose a meaningful set, i.e. we fix a pattern for each equivalence class:

$$\mathcal{D}_0(\psi) = \{\bar{D}_1, \ldots, \bar{D}_{e(\psi)}\} .$$

The counting variable X_n to which Lemma 2.5 will be applied is the total number of occurrences of one of the patterns $\bar{D}_1, \ldots, \bar{D}_{e(\psi)}$, in the random image $\mathcal{I}_{n,p(n)}$:

$$X_n = \sum_{x \in \Xi_n} \sum_{i=1}^{e(\psi)} \mathbb{I}_{\bar{D}_i(x)}(\mathcal{I}_{n,p(n)}) ,$$

where $\mathbb{I}$ denotes the indicator function of an event. The expectation of X_n is

$$\mathbb{E}(X_n) = n^2 \, e(\psi) \, (p(n))^{b(\psi)} (1 - p(n))^{2r^2+2r+1-b(\psi)} .$$

As n tends to infinity, it tends to $e(\psi) \, c^{b(\psi)}$, which is the parameter of the Poisson approximation in formula (1). In order to apply Lemma 2.5 to X_n, one has to check that the hypothesis holds.

Lemma 2.6

$$\forall l \in \mathbb{N}^* , \qquad \lim_{n \to \infty} E_l(X_n) = \left(e(\psi) c^{b(\psi)}\right)^l .$$

The proof of Lemma 2.6 is rather technical and will be given in the Appendix.

Now Lemma 2.5 implies that X_n converges in distribution to the Poisson distribution with parameter $e(\psi) c^{b(\psi)}$ therefore $\mu_{n,p(n)}(X_n > 0)$ tends to $1 - \exp(-e(\psi) c^{b(\psi)})$. It is clear that $X_n > 0$ implies that $\mathcal{I}_{n,p(n)}$ satisfies $\widetilde{\psi}$. Hence $\mu_{n,p(n)}(X_n > 0) \leqslant \mu_{n,p(n)}(\widetilde{\psi})$. Moreover, the event $(\widetilde{\psi} \setminus (X_n > 0)) = (\widetilde{\psi} \cap (X_n = 0))$ implies the appearance of a pattern with at least $b(\psi) + 1$ black pixels in a ball of radius r, and then by Lemma 2.3, its probability tends to 0 as n tends to infinity. Therefore,

$$\lim_{n \to \infty} \mu_{n,p(n)}(X_n > 0) = \lim_{n \to \infty} \mu_{n,p(n)}(\widetilde{\psi}) = 1 - \exp(-e(\psi) c^{b(\psi)}) .$$

It should be noticed that the asymptotics of X_n does not depend on the choice of the meaningful definition set $\{\bar{D}_1, \ldots, \bar{D}_{e(\psi)}\}$. It does not depend either on the radius r of the ball. Consider for instance the property $\widetilde{\psi}$ "the image contains two horizontally connected black pixels". Its definition set for the ball $B(0, r)$ has $r^2 2^{2r^2+2r}$ elements. Among these, only $2r^2$ have exactly 2 black pixels, and there is only one equivalence class up to translations, whatever r. Therefore r is a phantom parameter, as should be expected. It serves only to ensure that properties remain local. □

3 Application to image denoising

In the previous section, we computed the asymptotic probability of appearance of any local property in a random binary image. This provides a statistical test to decide if an observed pattern in an image may be due to noise or not, and this test can be applied for image denoising. In this section, all the considered images will be corrupted by the same kind of noise, namely an impulse noise. This type of noise models for example the fact that some (unknown) part of the data is lost. We will assume that the intensity of the noise is known. We will first start with the denoising of binary images, and then extend it to gray level images using their threshold decomposition.

3.1 Binary images

Let I_0 be the original (non degraded) binary image with size $n \times n$. This original image I_0 is then corrupted by an impulse noise, which has an intensity p in the white components and an intensity q in the black ones (see Figure 2 for an example). We shall see in the next section why it is important to allow black and white pixels to be destroyed with a different probability. Thus the noisy image I is given by

$$\forall x, \;\; I(x) = I_0(x) \cdot (1 - \zeta_p(x)) + (1 - I_0(x)) \cdot \zeta_q(x), \tag{2}$$

where the $\zeta_p(x)$'s (resp. $\zeta_q(x)$'s) are independent Bernoulli random variables with parameter p (resp. q). In other words, we have the following conditional probabilities

$$\mathbb{P}(I(x) = 0 \,|\, I_0(x) = 1) = p \quad \text{and} \quad \mathbb{P}(I(x) = 1 \,|\, I_0(x) = 0) = q.$$

As it can be noticed on Figure 2, the impulse noise creates small black and white connected components. These small components will be removed using a statistical decision based on their size ("size", in this paper, always means "area"). We are first interested in the black connected components (with respect to 4-connectivity). The results of the previous section give us the threshold function and also the probability of appearance of such components. More precisely, the threshold function for a given (fixed) black component of size k is $\theta(n) = n^{-2/k}$ and its asymptotic appearance probability in a $n \times n$ image of noise with intensity $p(n) = c\theta(n)$, as n goes to infinity, is equal to

$$1 - e^{-c^k}.$$

Now, if we are interested in the appearance of a component of size k (i.e. any of them, not only a given one), Theorem 2.4 claims that the asymptotic (as n is large) probability of appearance is

$$1 - e^{-a_k c^k} \;,$$

where a_k is the number of 4-connected components one can make with exactly k pixels (up to translations). Writing $c^k = n^2 p^k$, we thus have an approximation for the probability of appearance of a component of size k in the $n \times n$ image, with a proportion p of black pixels. We denote by $\mathrm{PA}(n,k,p)$ this approximation :

$$\mathrm{PA}(n,k,p) = 1 - e^{-n^2 a_k p^k} \ .$$

The 4-connected components are known in the combinatorics literature as "square lattice animals" or "polyominoes". Counting these animals is a difficult combinatorial problem and there is no general expression for a_k. However, some asymptotic results are known: a concatenation argument [9] shows that there exists a constant a, called *growth constant*, such that:

$$\lim_{k\to\infty} (a_k)^{\frac{1}{k}} = \sup_{k\geqslant 1} (a_k)^{\frac{1}{k}} = a \ .$$

The exact value of a is unknown, numerical estimates give $a \simeq 4.06$ and the best published rigorous bounds for it are $3.9 < a < 4.65$ (see [2, 8, 10]). But thanks to some numerical studies[1], the first values of the sequence $(a_k)_{k\geqslant 1}$ are known up to $k = 47$, which will be enough in practice for denoising applications. The first terms are: $a_1 = 1$, $a_2 = 2$, $a_3 = 6$, $a_4 = 19$, etc. Furthermore, numerical computations show that for $p \leqslant p_{max} \simeq 0.2$, one has $a_{k+1}p \leqslant a_k$ for $k \in [1,47]$, which ensures that $\mathrm{PA}(n,k+1,p) \leqslant \mathrm{PA}(n,k,p)$. This means that the probability of appearance of an animal is a decreasing function of its size. This is rather reasonable: for fixed values of p and n, it would not make much sense to keep a connected component of size k and to remove one of size $k' \geqslant k$.

Let us fix a (small) positive real ε which will be our risk probability, in the sense of statistical testing. If the size k of a connected component observed in a noisy image I is such that $\mathrm{PA}(n,k,p) \leqslant \varepsilon$, then we will consider that it comes from the original image I_0, and keep it. If $\mathrm{PA}(n,k,p) > \varepsilon$, it will be regarded as noise and removed. Thus the size threshold for the components we keep is defined by:

$$s(n,p,\varepsilon) = \inf\{k\,;\ \mathrm{PA}(n,k,p) = 1 - e^{-n^2 a_k p^k} \leqslant \varepsilon\}. \tag{3}$$

A component with size $k \geqslant s(n,p,\varepsilon)$ has a very low probability (less than ε) to appear in a pure noise image. On Figure 1.b, we plot the size threshold $s(n,p,\varepsilon)$ as a function of the noise intensity p, for a fixed value of $n = 256$ and three different values of ε: 10^{-1}, 10^{-2} and 10^{-3}.
The algorithm for the binary image denoising can be decomposed in four steps:

1. Compute all the black 4-connected components of the noisy image I.
2. Remove the ones which have a size less than $s(n,p,\varepsilon)$ (i.e. change their pixels into white). Obtain a new binary image.
3. Compute all the white 4-connected components of this new image.

[1] for up-to-date informations on the topic, see the web-site of the "On-line Encyclopedia of Integer Sequences", `http://www.research.att.com/~njas/sequences/` and references therein.

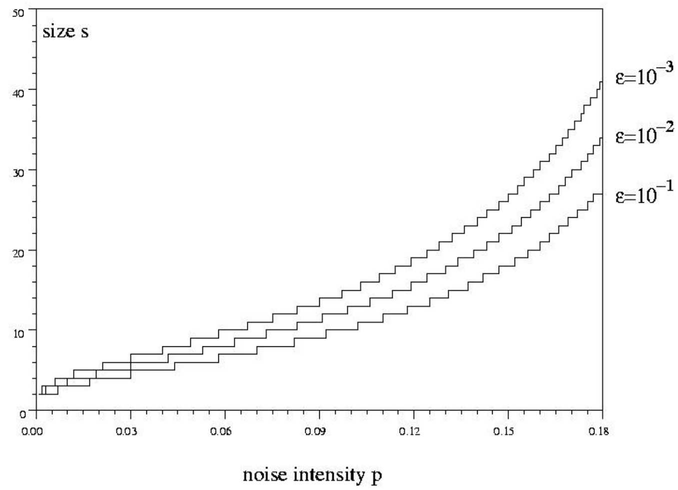

Figure 1: *The size threshold* $s(n,p,\varepsilon)$ *as a function of the noise intensity* $p \in [0, 0.18]$*, for* $n = 256$ *and* $\varepsilon = 10^{-1}$*,* 10^{-2} *and* 10^{-3}.

4. Remove the ones which have a size less than $s(n,q,\varepsilon)$ (i.e. change their pixels into black), to obtain the final denoised image denoted by $\tilde{I} = TI$.

To summarize, this denoising filter T can be written as:

$$T = T^{+}_{s(n,q,\varepsilon)} \circ T^{-}_{s(n,p,\varepsilon)} ,$$

where T^+_s (resp. T^-_s) is the morphological area opening (resp. closing) of size s defined by L. Vincent in [19].
See Figure 2 for an example of obtained result. Before explaining how this method will be extended to gray level images, let us make a few general comments.

- The method is valid when p is not too large, since we need $a_k p^k$ to be small. In practice, we are limited to $p \leqslant p_{max} \simeq 0.2$.
- The dependence in ε is low since it is in fact a $\log(\varepsilon)$-dependence. Indeed, $1 - e^{-n^2 a_k p^k}$ is equivalent to $n^2 a_k p^k$ when the value of this expression is small. If we replace a_k by a^k, the threshold for the minimal size of the components we keep is approximately given by
$$s(n,p,\varepsilon) \simeq \frac{\log\varepsilon - 2\log n}{\log a + \log p} .$$
- The boundaries of the remaining components are not smoothed. This comes from the fact that when some noise is at the boundary of a component, it becomes part of

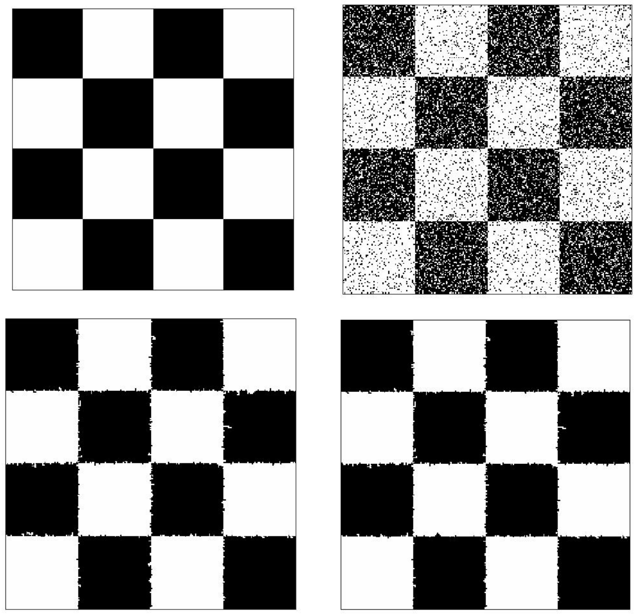

Figure 2: *First row: on the left, the original binary image I_0 with size 256×256; on the right, the corrupted image I. White pixels have been changed with probability $p = 0.1$ and black ones with probability $q = 0.2$. Second row: on the left, the result of the denoising algorithm when removing first black components and then white ones (i.e. applying $T^+_{s(n,q,\varepsilon)} \circ T^-_{s(n,p,\varepsilon)}$); on the right, denoising with first, removal of white components and then, black ones (i.e. applying $T^-_{s(n,p,\varepsilon)} \circ T^+_{s(n,q,\varepsilon)}$). The two images are not the same, illustrating the fact that the two operators T^-_s and T^+_s do not commute. However in both cases, small black and white components due to noise have been removed. Only the boundaries of the remaining ones are different.*

it. In order to remove it, one would need another *a priori* knowledge of the original image (like smooth or straight boundaries as in the case of Figure 2). It is actually a general problem of image denoising: one has to define some *a priori* model for the image. Here the underlying model is that the original binary image is made of "large" (as compared to the noise) black and white connected components.

- The two filters T^-_s and T^+_s do not commute (see Figure 2). This was already noticed by Vincent in [19]. One solution he proposed is to use them in alternating sequential filters [16, 17] with increasing sizes of area. This may not be a real issue, since he also noticed that $T^-_s \circ T^+_s$ and $T^+_s \circ T^-_s$ are visually extremely close (this will be even more true for gray level images). Another solution, proposed by Masnou and Morel in [11], and then formalized by Monasse [12], is to process simultaneously upper and lower level sets. This grain filter denoted G_t (where t is the area threshold) is done by a pruning of the tree of all level sets, built thanks to the inclusion principle (this algorithm which is very fast is called the Fast Level Set Transform [13]).

- If the original image I_0 is all white ($I_0 \equiv 0$), and if it is corrupted by some noise with intensity p as described by equation (2), we obtain an image I which is a "pure noise". The probability that it contains a connected component with size larger than $s(n,p,\varepsilon)$ is (by definition of $s(n,p,\varepsilon)$ and thanks to theorem 2.4) less than ε. Thus,

  $$\mathbb{P}(T^-_{s(n,p,\varepsilon)} I = I_0) \geqslant 1-\varepsilon \ ,$$

  which means that, with probability larger than $1-\varepsilon$, pure noise is completely removed.

3.2 Gray-level images

Let u be a gray level image, with size $n \times n$ and gray level values in the range $[0,255]$. Assume that this image is corrupted by an impulse noise with intensity p. This means that the observed noisy image v may be written under the form:

$$\forall x, \ \ v(x) = (1-\zeta_p(x)) \cdot u(x) + \zeta_p(x) \cdot \nu(x), \tag{4}$$

where the $\zeta_p(x)$'s are independent Bernoulli random variables with parameter p and the $\nu(x)$'s are i.i.d.r.v.'s, uniformly distributed on $[0,255]$.

For each level $\lambda \in [0,255]$, we can consider the thresholded images $u_\lambda = \mathbb{I}_{u \geqslant \lambda}$ and $v_\lambda = \mathbb{I}_{v \geqslant \lambda}$. The gray-level images may then simply be recovered by $u = \sum_\lambda u_\lambda$ and $v = \sum_\lambda v_\lambda$. The binary noisy image v_λ is a corrupted version of the binary image u_λ; they are related by

$$\mathbb{P}(v_\lambda(x) = 0 \,|\, u_\lambda(x) = 1) = p \times \frac{\lambda}{256} \ \text{ and } \ \mathbb{P}(v_\lambda(x) = 1 \,|\, u_\lambda(x) = 0) = p \times (1 - \frac{\lambda}{256}).$$

We are thus back in the framework described for binary images with parameters $p_\lambda = p\lambda/256$ and $q_\lambda = p(1-\lambda/256)$. The image v_λ can be denoised following the method described in the previous subsection. Finally, we reconstruct a gray-level image by simply adding the binary ones: $\tilde{v} = \sum_\lambda \tilde{v}_\lambda$. This can be summarized by the formula

$$\tilde{v} = Tv = \sum_{\lambda=0}^{255} T^+_{s(n,q_\lambda,\varepsilon)} \circ T^-_{s(n,p_\lambda,\varepsilon)}(v_\lambda), \quad \text{where } p_\lambda = p\frac{\lambda}{256} \text{ and } q_\lambda = p\left(1 - \frac{\lambda}{256}\right). \tag{5}$$

Figures 3 and 4 give two examples of results obtained by this filtering.

One natural question that can be asked is whether the filter T defined by formula (5) is a morphological filter. Unfortunately, the answer is negative. For two gray levels $\lambda \geqslant \lambda'$, one has $v_\lambda \leqslant v_{\lambda'}$, and for a fixed area threshold t one would have $T^-_t(v_\lambda) \leqslant T^-_t(v_{\lambda'})$ (because area openings and closings are morphological operators). Now, the two thresholds $s(n,p_\lambda,\varepsilon)$ and $s(n,p_{\lambda'},\varepsilon)$ can be different, i.e. $s(n,p_\lambda,\varepsilon) > s(n,p_{\lambda'},\varepsilon)$ and thus it is not necessarily true that $T^-_{s(n,p_\lambda,\varepsilon)}(v_\lambda) \leqslant T^-_{s(n,p_{\lambda'},\varepsilon)}(v_{\lambda'})$. This happens when v_λ and $v_{\lambda'}$ both contain a same small black connected component of size k such that $s(n,p_\lambda,\varepsilon) > k >$

Figure 3: *Left: image v obtained with an impulse noise of intensity $p = 0.15$ on the Lena image. Middle: thresholded image v_λ for the gray level $\lambda = 150$. Right: denoised image $\tilde{v}$ obtained by the noise adaptive grain filter T with $\varepsilon = 10^{-3}$.*

$s(n, p_{\lambda'}, \varepsilon)$. However, in the experimental results, we noticed that this rarely happens: for most values of λ, one has $\tilde{v}_\lambda \leqslant \tilde{v}_{\lambda-1}$.

In order to illustrate the interest of an adaptive area threshold, we treated the same image using our method, then using fixed area threshold (for this we used the algorithm developed by Monasse in [12]). The results are those of Figures 4 and 5. Figure 5 shows the result of the usual grain filter, denoted by G_t, for two different values of the area threshold: $t = 10$ and $t = 20$. One can notice that the parameter value $t = 10$ seems too low since there is still some remaining noise (for example on the coat of the cameraman). On the other hand the value $t = 20$ seems too large, since some of the original structures have disappeared (this is, for example, the case of the white parabola at the top of the building) and too low (there is some remaining noise on the coat). These results have to be compared with the one of Figure 4-c. This last figure shows that thanks to the adaptive area threshold $s(n, p_\lambda, \varepsilon)$ a small white component can be kept and in the same time, a larger gray component removed. These results also illustrate what we have proposed in this paper, namely an adaptive and automatic way to choose the right parameter for the area openings and closings.

4 Conclusion

We have introduced a mathematical model for random images, in which we were able to compute the probability of appearance of any "local pattern" (Theorem 2.4). This was then used to give an explicit formula for the size threshold $s(n, p, \varepsilon)$, such that the probability of appearance of a component of size $k \geqslant s(n, p, \varepsilon)$ in a $n \times n$ image of pure noise with intensity p is less than ε. Using this value of $s(n, p, \varepsilon)$ for the area openings and closings defined by Vincent will ensure that, with probability larger than $1 - \varepsilon$, pure noise is

Figure 4: *From left to right, top to bottom: (a) the original cameraman image u (size* 256×256*); (b) degraded image v, with impulse noise intensity* $p = 0.2$*; (c) filtered image* $\tilde{v}$*, obtained with* $\varepsilon = 10^{-3}$*; (d) image of the difference* $u - \tilde{v}$*. It shows that most of the noise has been removed, except at the boundaries of the objects and also in the grass texture.*

completely removed. This denoising process was then extended to gray level images using their threshold decomposition. There, the proposed area threshold depends on both the intensity p of the impulse noise and the gray level λ of the level set.
Now, some questions remain, that have not been adressed in this paper: if the intensity p of the impulse noise is unknown, what is the best way to estimate it ? For a binary pure noise image, the best estimate of p is simply the ratio of the number of black pixels to the area of the image. Then, by analogy, a first answer for binary images (like the chessboard for example) is to compute the relative number of black pixels outside a dilation of the "large" black components. Now, it is not clear how this can be extended to gray level images, since they often contain textures creating small components which over-estimate p.
Another remaining question is: how can this be extended to other models of noise ?

Figure 5: *Result of the filtering of the noisy image v with the usual grain filter G_t with area threshold $t = 10$ on the left and $t = 20$ on the right.*

Acknowledgements

We would like to thank Mireille Bousquet-Mélou for all the references about the enumeration of square lattice animals.

References

[1] B. Bollobás. *Random Graphs.* Academic Press, London, 1985.

[2] A.R. Conway and A.J. Guttmann. On two-dimensional percolation. *J. Phys. A: Math. Gen.*, 28:891–904, 1995.

[3] D. Coupier, A. Desolneux, and B. Ycart. A zero-one law for first-order logic on random images. Technical Report 2003-03, MAP5, 2003.

[4] A. Desolneux, L. Moisan, and J.-M. Morel. Meaningful alignments. *Int. J. Computer Vision*, 40(1):7–23, 2000.

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

[6] D.L. Donoho and I.M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of Amer. Stat. Assoc.*, 90:1200–1224, 1995.

[7] H.J.A.M. Heijmans. Connected morphological operators for binary images. *Computer Vision and Image Understanding*, 73(1):99–120, 1999.

[8] I. Jensen and A.J. Guttmann. Statistics of lattice animals (polyominoes) and polygons. *J. Phys. A: Math. Gen.*, 33:L257–L263, 2000.

[9] D.A. Klarner. Cell growth problems. *Canad. J. Math.*, 19:851–863, 1967.

[10] D.A. Klarner and R.L. Rivest. A procedure for improving the upper bound for the number of n-ominoes. *Canad. J. Math.*, 25:585–602, 1973.

[11] S. Masnou and J.-M. Morel. Image restoration involving connectedness. In *Proc. DIP'97, Vienna, Austria*, volume 3346. SPIE, 1998.

[12] P. Monasse. Contrast invariant representation of digital images and application to registration. PhD thesis, University Paris 9, 2000.

[13] P. Monasse and F. Guichard. Fast computation of a contrast invariant image representation. *IEEE Transactions on Image Processing*, 9(5):860–872, 2000.

[14] B. Sapoval and M. Rosso. Gradient percolation and fractal frontiers in image processing. *Fractals*, 3:23–31, 1995.

[15] J. Serra. *Image analysis and mathematical morphology, Part. I.* Academic Press, New York, 1982.

[16] J. Serra. *Image analysis and mathematical morphology, Part. II: Theoretical Advances.* Academic Press, London, 1988.

[17] J. Serra and L. Vincent. An overview of morphological filtering. *Circuits, Systems and Signal Processing*, 11(1):47–108, 1992.

[18] J. Spencer. Nine lectures on Random Graphs. In P. Bernard, editor, *Ecole d'été de probabilité de Saint-Flour XXI*, volume 1541 of *L.N. in Mathematics*, pages 293–343. Springer-Verlag, New York, 1991.

[19] L. Vincent. Morphological area opening and closing for grayscale images. In *Proc. NATO Shape in Picture Workshop, Driebergen, The Netherlands*, pages 197–208. Springer-Verlag, 1992.

[20] L. Vincent. Grayscale area openings and closings: their applications and efficient implementation. In *Proc. EURASIP Workshop on Mathematical Morphology and its Applications to Signal Processing, Barcelona, Spain*, pages 22–27, 1993.

A Appendix

Proof of Lemma 2.6:

Fix $l \in \mathbb{N}^*$. Recall that X_n counts the number of occurrences of the meaningful patterns $\bar{D}_1, \ldots, \bar{D}_{e(\psi)}$ in the random image $\mathcal{I}_{n,p(n)}$ where $p(n) = cn^{-\frac{2}{b(\psi)}}$. We are interested in:

$$E_l(X_n) = \sum_{k \geqslant l} \mathbb{P}(X_n = k) \frac{k!}{(k-l)!} \ .$$

We need to prove that $E_l(X_n)$ tends to $(e(\psi)c^{b(\psi)})^l$ as n tends to infinity.
One can see $E_l(X_n)$ as the average number of ordered l-tuples of copies of the patterns $\bar{D}_1, \ldots, \bar{D}_{e(\psi)}$ in $\mathcal{I}_{n,p(n)}$. Thus, we can write:

$$\begin{aligned} E_l(X_n) &= \mathbb{E}\left(\sum_{\substack{x_1,\ldots,x_l \\ x_i \neq x_j}} \sum_{1 \leqslant j_1,\ldots,j_l \leqslant e(\psi)} \mathbb{I}_{\bar{D}_{j_1}(x_1) \wedge \ldots \wedge \bar{D}_{j_l}(x_l)}(\mathcal{I}_{n,p(n)}) \right) \\ &= \sum_{s=1}^{l} \sum_{\substack{(x_1,\ldots,x_l) \\ \in \mathcal{C}(s)}} \sum_{\substack{1 \leqslant j_1,\ldots,j_l \\ \leqslant e(\psi)}} \mu_{n,p(n)}(\bar{D}_{j_1}(x_1) \wedge \ldots \wedge \bar{D}_{j_l}(x_l)) \ , \end{aligned}$$

where, for $s = 1, \ldots, l$, $\mathcal{C}(s)$ represents the set of the $l-$tuples $(x_1, \ldots, x_l)$ of pixels of Ξ_n such that the set of the balls $B(x_1, r), \ldots, B(x_l, r)$ is composed of s equivalence classes for the $4-$connectivity relation.
The term corresponding to $s = l$ in the last sum will be denoted by $E_l^{'}(X_n)$ and the rest by $E_l^{''}(X_n)$. The quantity $E_l^{'}(X_n)$ can be seen as the average number of ordered l-tuples of copies of $\bar{D}_1, \ldots, \bar{D}_{e(\psi)}$, on non-overlapping balls. We will first show that:

$$\lim_{n \to \infty} E_l^{'}(X_n) = (e(\psi)c^{b(\psi)})^l \ . \tag{6}$$

Then we will prove that $E_l^{''}(X_n)$ tends to 0 as n tends to infinity.

We want to choose l pixels $x_1, \ldots, x_l$ such that the balls of radius r centered on those pixels are two by two disjoint. For the first pixel x_1, there are n^2 possibilities. Let $2 \leqslant j \leqslant l$ and suppose pixels $x_1, \ldots, x_{j-1}$ have been chosen. For the j-th choice, the set of all pixels

x such that $d(x, x_k) \leqslant 2r$ for some $1 \leqslant k \leqslant j-1$, must be avoided. The cardinality of this set is bounded by $(j-1) \times (8r^2 + 4r + 1)$ whatever $x_1, \ldots, x_{j-1}$. This bound does not depend on n. So, asymptotically the number of choices for the j−th element is n^2, and consequently the cardinality of $\mathcal{C}(l)$ is equivalent to n^{2l}. On the other hand, if two balls $B(x, r)$ and $B(x', r)$ are disjoint, then for all $1 \leqslant j, j' \leqslant e(\psi)$, the random variables $\mathbb{I}_{\bar{D}_j(x)}$ and $\mathbb{I}_{\bar{D}_{j'}(x')}$ are independent. Therefore, we obtain the first limit (relation (6)):

$$E_l^{'}(X_n) \sim n^{2l}(e(\psi)p(n)^{b(\psi)}(1-p(n))^{2r^2+2r+1-b(\psi)})^l \sim (e(\psi)c^{b(\psi)})^l \ .$$

The factor $e(\psi)^l$ comes from the choice of the $e(\psi)$ patterns $\bar{D}_1, \ldots, \bar{D}_{e(\psi)}$ for the l chosen balls.

There remains to prove that $E_l^{''}(X_n)$ tends to 0 as n tends to infinity. The intuition is that if two patterns occur in overlapping balls, then locally more than $b(\psi)$ black pixels are present in a ball of radius $2r$. This has vanishing probability, by Lemma 2.3.
Let $1 \leqslant s \leqslant l-1$ and $(x_1, \ldots, x_l)$ be an element of $\mathcal{C}(s)$. Let $C_1, \ldots, C_s$ represent the connected components of the set $\cup_{k=1}^l B(x_k, r)$. Then by independence between them (they concern disjoint pixel sets):

$$\mu_{n,p(n)}(\bar{D}_{j_1}(x_1) \wedge \ldots \wedge \bar{D}_{j_l}(x_l)) = \prod_{m=1}^{s} \mu_{n,p(n)}(\bigwedge_{k;B(x_k,r)\in C_m} \bar{D}_{j_k}(x_k)) \ .$$

As a consequence of $s \leqslant l-1$, there exists at least one connected component, say C_1, having at least two elements. Since the black pixel sets of two different patterns of $\mathcal{D}_0(\psi)$ cannot be translated of each other, there must be at least $b(\psi)+1$ black pixels in C_1. Thus we have

$$\mu_{n,p(n)}(\bigwedge_{k;B(x_k,r)\in C_1} \bar{D}_{j_k}(x_k)) \leqslant p(n)^{b(\psi)+1} \ .$$

For the other connected components, we simply bound

$$\mu_{n,p(n)}(\bigwedge_{k;B(x_k,r)\in C_m} \bar{D}_{j_k}(x_k)) \leqslant \mu_{n,p(n)}(\bar{D}_{j_{k_m}}(x_{k_m})) \leqslant p(n)^{b(\psi)} \ ,$$

for any index k_m such that $B(x_{k_m}, r) \in C_m$. Therefore, we obtain the following result:

$$\mu_{n,p(n)}(\bar{D}_{i_1}(x_1) \wedge \ldots \wedge \bar{D}_{i_l}(x_l)) \leqslant p(n)^{sb(\psi)+1}.$$

Finally, the set $\mathcal{C}(s)$ has only $O(n^{2s})$ elements and the number of ways to choose l elements among $\bar{D}_1, \ldots, \bar{D}_{e(\psi)}$ does not depend on n. Consequently, the desired result follows:

$$E_l^{''}(X_n) \leqslant \sum_{s=1}^{l-1} O(n^{2s} \times n^{-\frac{2(sb(\psi)+1)}{b(\psi)}}) = \sum_{s=1}^{l-1} O(n^{-\frac{2}{b(\psi)}}) = o(1) \ .$$

□