On Region Merging: The Statistical Soundness of Fast Sorting, with Applications

Frank Nielsen
Sony CS Labs, FRL
3-14-13 Higashi Gotanda
Shinagawa-Ku, Tokyo 141-0022, Japan
http://www.csl.sony.co.jp/person/nielsen/

Richard Nock
DSI-Univ. Antilles-Guyane
Campus de Schoelcher, B.P. 7209
97275 Schoelcher, France
http://www.univ-ag.fr/∼rnock

Abstract

This work explores a statistical basis for a process often described in computer vision: image segmentation by region merging following a particular order in the choice of regions. We exhibit a particular blend of algorithmics and statistics whose error is, as we formally show, close to the best possible. This approach can be approximated in a very fast segmentation algorithm for processing images described using most common numerical feature spaces. Simple modifications of the algorithm allow to cope with occlusions and/or hard noise levels. Experiments on grey-level and color images, obtained with a short C-code, display the quality of the segmentations obtained.

1. Introduction

It is established since the Gestalt movement in psychology that perceptual grouping plays a fundamental role in Human perception. Even though this observation is rooted in the early part of the XXth century, the adaptation and automation of the segmentation (and more generally, grouping) task with computers has remained so far a tantalizing and central problem for image processing. Roughly speaking, the problem can be presented as the transformation of the collection of pixels of an image into a meaningful arrangement of regions and objects. There are four large categories of approaches to image segmentation [1], one of which is of direct interest to us : region growing and merging techniques. In region merging, regions are sets of pixels with homogeneous properties and they are iteratively grown by combining smaller regions or pixels, pixels being elementary regions. Region growing/merging techniques usually work with a statistical test to decide the merging of regions [1]. A merging predicate uses this test, and builds the segmentation on the basis of (essentially) local decisions.

Figure 1. A natural RGB image and the segmentation found by our segmentation method (regions white bordered & averaged inside).

This locality in decisions has to preserve global properties, such as those responsible for the perceptual units of the image [2]. In Figure 1, the grassy region below the castle is one such unit, even when its variability is high compared to the other regions of the image. In that case, a good region merging algorithm has to find a good balance between preserving this unit and the risk of overmerging for the remaining regions. The right image in Figure 1 shows the result of our approach.

As long as the approach is greedy, two essential components participate in defining a region merging algorithm: the merging predicate, and the order followed to test the merging of regions. There is a lack of theoretical results on the way these two components interact together, and can benefit from each other. This might partially be due to the fact that most approaches use assumptions on distributions, more or less restrictive, which would make any theoretical insight

into how region merging works restricted to such settings, and therefore of possibly smaller interest (see *e.g.* [3] for related criticisms).

Following [4], we chose to tackle the segmentation problem with a model which, basically, only keeps independence as its main assumption, and is otherwise essentially hypothesis-free. Therefore, there is no *ad hoc* prior on the nature of the distributions. Our main result is a region merging algorithm tailored to this model, and a proof of its performance guarantees. Our proof completes previous results on this line of work [4], in particular by showing that the algorithm has an accuracy in segmentation close to the optimum, up to low order terms. The algorithm has some desirable features: it relies on a simple interaction between a merging predicate easily implementable, and an order in merging approximable in linear time. Furthermore, it can be adapted to most numerical feature description spaces. The practical standpoints emphasize some desirable features of the algorithm: its implementation requires only a few kilobytes C code, and no pre-processing prior to segmentation is needed (neither any extensive tuning of its parameters). Finally, its simplicity makes it, as we show, a simple tool for extensions to segmentations handling hard noise and/or significantly occluded images, at very affordable algorithmic complexity.

The next section details our model of image generation. It is followed by a section presenting and analyzing the merging predicate. Section 4 proves our main Theorem on the accuracy of the resulting segmentation algorithm. Section 5 presents our experimental setups for the algorithm's implementation. Section 6 presents and discusses general experimental results, as well as results on adapting the algorithm to occlusion and/or noise handling.

2. Image Generation Model

Throughout this paper, the notation $|.|$ stands for cardinal. The observed image, I, contains $|I|$ pixels, each containing one of g grey-levels. For the sake of simplicity, our primary analysis relies on the grey-level setting, *i.e.* for a single color band. On this basis, the extension of the results to more numerical channels, such as **RGB**, does not require an involved analysis. It is presented in Section 5. Our model of image generation and our basic machinery follow [4].

A theoretical "image" (or scene) I^*, represents the true objects (or true regions) as distributions. We aim at approximating the contour of these objects through the observation of an image, I, which is sampled from I^*. We suppose that there are s_{opt} true objects in I^*. I^* is composed of theoretical pixels, to each of which corresponds a pixel of I. Each theoretical pixel is a set of $Q > 0$ independent random variables (r.v.). These r.v. are supposed discrete, and take positive values so as to sum for each pixel in the set defining the grey-levels (say, $\{1, 2, ..., g\}$). The true objects in I^* satisfy the following properties:

- an *homogeneity* property: for each true object, the sum of expectations of the Q r.v. of each of its theoretical pixels is the same.
- a *separability* property: for any two theoretical pixels belonging to adjacent true objects, the sum of expectations of their Q r.v. are different.

Let us consider a segmentation as a partition of an image into 4-connex regions. Our model allows to define the *ideal* segmentation of I, $s^*(I)$: it is the partition of I according to the s_{opt} true objects of I^*. Our objective is to build an observed segmentation of I, $s(I)$, approximating as best as possible $s^*(I)$, that is, having the lowest error. The quality of $s(I)$ with respect to $s^*(I)$ is estimated by the following quantity:

$$Err(s(I), s^*(I)) = \boldsymbol{E}_{R\cap O, R\in s(I), O\in s^*(I)} \left(|\mathbf{E}(O) - \mathbf{E}(R)|\right)$$

Here, $\boldsymbol{E}$ (slanted) denotes the expectation with associated probability measure $\mu(R \cap O) = |R \cap O|/|I|$. The notation $\mathbf{E}(T)$ for some arbitrary region $T \in s(I) \cup s^*(I)$ is the expectation over all corresponding theoretical pixels of I^* of their sum of expectations of their Q r.v. . For the sake of clarity, the letter R (eventually with subscripts) refers to regions of $s(I)$, whereas the letter O (eventually with subscripts) refers to regions of $s^*(I)$. Finally, notation $\overline{R}$ (resp. $\overline{O}$) refers to the observed grey-level average on image I of some $R \in s(R)$ (resp. $O \in s^*(I)$).

Notice that the frequent i.i.d. assumption is relaxed in this model to that of ordinary independence. Inside a true region, it can be the case that all distributions associated to each pixel are different, as long as the homogeneity property is satisfied. This freedom has a counterpart, which led us to introduce Q, not necessarily to make our model more general, but essentially for practical purposes : the conventional choice $Q = 1$ would make it actually hard to estimate reliably anything for small regions, or equivalently, would make it necessary to consider very large images to improve the segmentation's statistical accuracy. Notice that Q is a parameter which makes sense: it quantifies the complexity of the scene, the generality of the model, and the statistical hardness of the task as well. Experimentally, Q turns out to be a parameter that the user can chose to tune to control the coarseness of the segmentation, even if we shall see that this tuning does not absolutely seem to be necessary, as an intermediate value ($Q = 32$) brings accurate results for a large body of images from different domains.

3. Merging Predicate

A first result states that, with high probability, the observed average of any region in I shall not deviate too much

from its theoretical expectation.

Theorem 1 *[4] Let R be a region in I. Let $\mathcal{R}_l$ be the set of regions having l pixels in I. Then the probability that there exists a region in $\mathcal{R}_{|R|}$ such that $\left|\overline{R} - \boldsymbol{E}(R)\right| \geq g\sqrt{\frac{1}{2Q|R|}\ln\frac{2|\mathcal{R}_{|R|}|}{\delta'}}$ is no more than δ'.*

The use of Theorem 1 is straightforward. The probability that there exists a region in I (regardless of its size) for which its average deviates from its expectation by more than the right-hand-side of the ineq. in Theorem (1) is no more than $|I|\delta'$, since the region can have size 1, 2, ..., or $|I|$. If we fix

$$\delta \quad = \quad |I|\delta' \tag{1}$$

and

$$b(R) \quad = \quad g\sqrt{\frac{1}{2Q|R|}\left(\ln\frac{2}{\delta'} + \ln|\mathcal{R}_{|R|}|\right)}\ , \tag{2}$$

then we know that, with high probability (*i.e.* $> 1-\delta$), for any region R, $\overline{R} \in [\mathbf{E}(R) - b(R), \mathbf{E}(R) + b(R)]$. In that case, because of the triangle inequality, we know that for any two regions R and R' having the same expectation ($\mathbf{E}(R') = \mathbf{E}(R) = q$), we have $|\overline{R'} - \overline{R}| \in [-(b(R) + b(R')), b(R) + b(R')]$, since $|\overline{R'} - \overline{R}| \leq |\overline{R'} - q| + |\overline{R} - q|$. From a practical point of view, if $|\overline{R'} - \overline{R}| \leq b(R) + b(R')$, then we can suppose that R and R' belong to the same true region in I^*, and merge them. The merging predicate for a single color level is therefore:

$$\mathcal{P}(R,R') = \left\{ \begin{array}{ll} \texttt{true} & \text{if } |\overline{R'} - \overline{R}| \leq b(R) + b(R') \\ \texttt{false} & otherwise \end{array} \right. .$$

We also replace $|\mathcal{R}_{|R|}|$ in $b(.)$ by an upperbound $|\mathcal{R}_{|R|}| \leq (|R|+1)^g$ [4]. To complete our discussion on the importance of Q in section 2, let us consider a simple numerical case. Suppose that $|R| = 30 \times 30$ (a typical segment size), with $|I| = 1000 \times 1000$ (a large image). If we fix $\delta = 2^{-7}$, then $b(R) \approx 256\ (= g)$ for $Q = 1$, a choice clearly not reasonable. However, if we fix *e.g.* $Q = 32$, then $b(R) \approx 45$, a reasonable bound.

4. Algorithm and Analysis

Given the predicate $\mathcal{P}$, it basically remains to give the order in which the region merging tests should be carried out. [4] shows that there exists a theoretical order in the merging tests which leads to an interesting global property on the segmentation's error. More precisely, consider algorithm $\mathcal{A}$ which satisfies the following:

- $\mathcal{A}$ makes the merging tests in such a way that when any test between two (parts of) true regions occurs, that means that all tests inside each of the two true regions have previously occurred.

Informally, the merging predicate on I defines a total ordering, admitting a partial sub-ordering with a simple property relying on I^*. It is important to notice that this assumption does not make *a priori* $\mathcal{A}$ an efficient algorithm at all (even theoretically). In particular, the algorithm still faces the three kind of segmentation errors. First, while some ideal regions may coincide with the regions found in $s(I)$, some ideal regions may be grouped together into regions of $s(I)$, which represents over-merging (more merges than necessary are done). Second, there is the "opposite" case where true regions can be split into many small regions in $s(I)$, a situation to which we refer as under-merging (not enough merges). Third, there is a hybrid case, in which some regions obtained in $s(I)$ span sub-parts of different true regions. These three sources of errors cohabit together in $s(I)$, and obtaining a good segmentation with low error can only be obtained by the control of all of them. It is possible to show that, when $\mathcal{A}$ is associated to $\mathcal{P}$, only one source of errors (out of the three) remains.

Theorem 2 *[4] With probability $> 1-\delta$ (eq. (1)), the segmentation on I is an over-merging of I^*, that is: $\forall O \in s^*(I), \exists R \in s(I) : O \subseteq R$.*

What Theorem 2 says is that with high probability, the only source of error will be through over-merging, and neither from under-merging nor the splitting of true regions of $s^*(I)$ between more than one region found. It limits the sources of error, but unfortunately not the error incurred by these sources. We now show that this error can effectively be upperbounded. The remaining of this Section is devoted to the proof and discussion of the following Theorem.

Theorem 3 *For any images I^* (with $s_{opt} > 1$) and I, $\forall 0 < \delta' < 1$, suppose that we use our merging predicate $\mathcal{P}$ together with algorithm $\mathcal{A}$. Then, with probability $> 1-\delta$ where $\delta = |I|\delta'$, the segmented image suffers an error with respect to the optimal segmentation satisfying*

$$\begin{array}{l} Err(s(I), s^*(I)) \\ \quad \leq \quad \mathcal{O}\left(g\sqrt{\frac{s_{opt}\ln s_{opt}}{|I|Q}\left(\ln\frac{1}{\delta'} + g\ln|I|\right)}\right) . \end{array}$$

Since Theorem 2 holds, for any region R obtained in I, we define $n(R) = |\{O \in s^*(I) : O \subseteq R\}|$. Notice that because we are in an over-merging setting, any region R for which $n(R) = 1$ does not participate in increasing the error. Therefore, w.l.o.g., we concentrate only on regions R for which $n(R) > 1$. We introduce a more adequate representation for regions in $s(I)$, which shall come

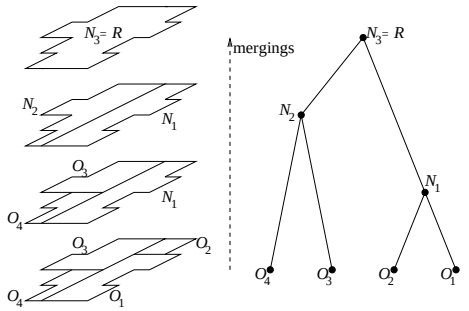

Figure 2. The tree-representation of a region R (with $n(R) = 4$) in $s(I)$, depicting the stepwise construction of R by merging four regions O_1, O_2, O_3, O_4 of $s^*(I)$.

in handy for our proof. A region R is represented by a binary tree, the leaves of which are the ideal regions contained in R, and labeled $O_1, O_2, ...$ and so on. The root represents the region R, and each internal node has two children, representing the two regions in I that $\mathcal{A}$ has merged to obtain their parent node. Figure 2 displays the tree of a simple region R containing four ideal regions. For any node C of this tree, we also use the notation "i-node(C)" to represent the set of all internal nodes of the subtree rooted at C (therefore, containing also C). In Figure 2, we would have i-node(O_1)= i-node(O_2)= i-node(O_3)= i-node(O_4)= $\emptyset$, i-node(N_1)= $\{N_1\}$, i-node(N_2)= $\{N_2\}$, and i-node(N_3)= i-node(R)= $\{N_1, N_2, N_3\}$= $\{N_1, N_2, R\}$. In our over-merging setting, denote $\forall O \in s^*(I)$:

$$I(O) \quad = \quad R \in s(I) : O \subseteq R \ . \tag{3}$$

For any region T, either in $s(I)$, or in $s^*(I)$, we rewrite $b(.)$ (section 3) using the upperbound on $|\mathcal{R}_{|R|}|$ as

$$b(T) \quad = \quad g\sqrt{\frac{1}{2Q|T|}\left(\ln\frac{2}{\delta'} + g\ln(|T|+1)\right)} \ . \tag{4}$$

We also fix $\mathcal{S} = \{O \in s^*(I) : O \neq I(O)\}$ in the next Lemmas.

Lemma 1 $Err(s(I), s^*(I)) \leq \sum_{O\in\mathcal{S}} \frac{|O|}{|I|}(b(I(O)) + b(O)) + \sum_{O\in\mathcal{S}} \frac{|O|}{|I|}|\overline{I(O)} - \overline{O}|$.

Proof: $\forall R \in s(I), \forall O \in s^*(I)$, we have by the triangle inequality $|\mathbf{E}(O) - \mathbf{E}(R)| \leq |\mathbf{E}(O) - \overline{O}| + |\overline{O} - \overline{R}| + |\overline{R} - \mathbf{E}(R)|$. Since the segmentation is an over-merging of I^*, we have $Err(s(I), s^*(I)) = \sum_{O\in s^*(I), O\neq I(O)} \frac{|O|}{|I|} |\mathbf{E}(O) - \mathbf{E}(I(O))|$, and since we are in the $> 1-\delta$ probability event that for each region $R \in s(I)$ its theoretical grey-level expectation does not deviate by more than $b(R)$ from its observed average, we have also $|\mathbf{E}(O) - \overline{O}| \leq b(O)$ and $|\overline{I(O)} - \mathbf{E}(I(O))| \leq b(I(O))$. Putting this altogether, we obtain the statement of the Lemma. □

We upperbound separately the two terms of the error in Lemma 1.

Lemma 2

$$\sum_{O\in\mathcal{S}} \frac{|O|}{|I|}|\overline{I(O)} - \overline{O}| \leq 2g\sqrt{\frac{s_{opt}\log(s_{opt})}{|I|Q}\left(\ln\frac{2}{\delta'} + 2g\ln|I|\right)}.$$

Proof: We have $\sum_{O\in s^*(I), O\neq I(O)} |O||\overline{I(O)} - \overline{O}|/|I| \leq \sum_{(R\in s(I), n(R)>1)} \sum_{(O\in s^*(I), O\subset R)} |O||\overline{I(O)} - \overline{O}|/|I|$. Let us consider the local error for some fixed region R with $n(R) > 1$ in the last summation : $\sum_{O\in s^*(I), O\subset R} \frac{|O|}{|I|}|\overline{R} - \overline{O}|$. Consider the two true regions O_1 and O_2 that were merged at first prior to creating R (see Figure 2). Their participation to the local error due to R is $|O_1||\overline{R} - \overline{O_1}|/|I| + |O_2||\overline{R} - \overline{O_2}|/|I|$. Straightforward derivations yield (i) $|O_1||\overline{R} - \overline{O_1}|/|I| \leq |O_1|\left(|\overline{R} - \overline{O_1\cup O_2}| + |\overline{O_1\cup O_2} - \overline{O_1}|\right)/|I|$, (ii) $|\overline{O_1\cup O_2} - \overline{O_1}| \leq |O_2||\overline{O_2} - \overline{O_1}|/|O_1 \cup O_2|$ (the same relationships hold when permuting O_1 and O_2), (iii) $|\overline{O_1} - \overline{O_2}| \leq b(O_1) + b(O_2)$ (because O_1 and O_2 were merged). Putting altogether all these inequalities, we get:

$$\begin{aligned}
&\frac{|O_1|}{|I|}|\overline{R} - \overline{O_1}| + \frac{|O_2|}{|I|}|\overline{R} - \overline{O_2}| \\
&\leq \quad \frac{|O_1\cup O_2|}{|I|}|\overline{R} - \overline{O_1\cup O_2}| \\
&\qquad + \frac{2|O_1||O_2|}{|I||O_1\cup O_2|}(b(O_1) + b(O_2)) \qquad (5) \\
&= \quad \frac{|N_1|}{|I|}|\overline{R} - \overline{N_1}| \\
&\qquad + \frac{2|O_1||O_2|}{|I||O_1\cup O_2|}(b(O_1) + b(O_2)) \ , \qquad (6)
\end{aligned}$$

where N_1 is the direct common ancestor (the parent) of O_1 and O_2 (see also Figure 2). Therefore, if we except the penalizing term depending on $b(O_1)+b(O_2)$, the contribution of R to the error can be upperbounded by the error of the leaves of the tree representation of R (the ideal regions) in which we replace the two terms due to O_1 and O_2 by a new term due to the parent of O_1 and O_2, thus producing a new tree in which we still have to measure the contribution to the error. Repeating recursively this procedure consisting in measuring the error of a tree which is stepwise reduced by replacing the contribution of the currently merged regions by their parent's, plus a penalizing term whose value is known and depends on $b(.)$, we easily obtain the following relationship:

$$\sum_{\substack{O\in s^*(I)\\ O\subset R}} \frac{|O|}{|I|}|\overline{R} - \overline{O}| \leq \sum_{N\in\text{i-node}(R)} \frac{2|C_1||C_2|}{|I||N|}(b(C_1) + b(C_2)),$$

where C_1 and C_2 are the two children of N in the tree representation of R, themselves internal nodes ($\in$ i-node(R))

or leaves ($\in \{O_1, O_2, ...\}$). Now, note that

$$\begin{aligned}
&\frac{2|C_1||C_2|}{|I||N|}\left(b(C_1)+b(C_2)\right)\\
&= \frac{\sqrt{2}g}{|I|\sqrt{Q}}\left(\frac{|C_2|}{|N|}\sqrt{|C_1|\left(\ln\frac{2}{\delta'}+g\ln(|C_1|+1)\right)}\right.\\
&\quad\left.+\frac{|C_1|}{|N|}\sqrt{|C_2|\left(\ln\frac{2}{\delta'}+g\ln(|C_2|+1)\right)}\right) .
\end{aligned}$$

Function $f(x)=\sqrt{x(a+b\ln(x+1))}$ for constants $a>0$, $b>1$ (an integer) is concave in $x\geq 1$, which means by Jensen's inequality that $\alpha_1 f(x)+\alpha_2 f(y)\leq f(\alpha_1 x+\alpha_2 y)$ whenever $\alpha_1+\alpha_2=1$, $\alpha_1,\alpha_2\geq 0$. If we fix $x=|C_1|$, $y=|C_2|$, $\alpha_1=|C_2|/|N|$ and $\alpha_2=|C_1|/|N|$, then we obtain the upperbound:

$$\begin{aligned}
&\frac{2|C_1||C_2|}{|I||N|}\left(b(C_1)+b(C_2)\right)\\
&\leq \frac{\sqrt{2}g}{|I|\sqrt{Q}}\sqrt{\frac{2|C_1||C_2|}{|N|}\left(\ln\frac{2}{\delta'}+g\ln\left(\frac{2|C_1||C_2|}{|N|}+1\right)\right)} .
\end{aligned}$$

Now, fix $a_N=\frac{2|C_1||C_2|}{|N|}$ and $b_N=\ln\frac{2}{\delta'}+g\ln(a_N+1)$. We have

$$\begin{aligned}
&\sum_{O\in s^*(I),O\neq I(O)}\frac{|O|}{|I|}|\overline{I(O)}-\overline{O}|\\
&\leq \frac{\sqrt{2}g}{|I|\sqrt{Q}}\sum_{(R\in s(I),n(R)>1)}\ \sum_{N\in\text{i-node}(R)}\sqrt{a_N}\sqrt{b_N}\\
&\leq \frac{\sqrt{2}g}{|I|\sqrt{Q}}\sqrt{A_N B_N} , \qquad (7)
\end{aligned}$$

with

$$A_N = \sum_{(R\in s(I),n(R)>1)}\ \sum_{N\in\text{i-node}(R)} a_N , \qquad (8)$$

$$B_N = \sum_{(R\in s(I),n(R)>1)}\ \sum_{N\in\text{i-node}(R)} b_N . \qquad (9)$$

The two next Lemmas bound respectively A_N and B_N.

Lemma 3 *For any region $R\in s(I)$ for which $n(R)>1$,*

$$\sum_{N\in\textit{i-node}(R)} a_N \leq |R|\log(n(R)) . \qquad (10)$$

(proof omitted due to the lack of space; it relies on an induction on the depth of R's tree shown in Figure 2, that is, on the maximum number of edges to go from the root to a leaf of the tree.) Because of the fact that $n(R)\leq s_{opt}$ for any region R in $s(I)$, we have by Lemma 3 and eq. (8):

$$A_N \leq \log(s_{opt})|I| . \qquad (11)$$

Lemma 4 *For any region $R\in s(I)$ for which $n(R)>1$,*

$$\sum_{N\in\textit{i-node}(R)} b_N \leq n(R)\left(\ln\frac{2}{\delta'}+2g\ln|I|\right) . \qquad (12)$$

Proof: We have $\sum_{N\in\text{i-node}(R)} b_N = \sum_{N\in\text{i-node}(R)}\left(\ln\frac{2}{\delta'}+g\ln(a_N+1)\right)$. Therefore

$$\begin{aligned}
&\sum_{N\in\text{i-node}(R)} b_N\\
&\leq n(R)\ln\frac{2}{\delta'}+g\ln\left(\prod_{R\in s(I),n(R)>1}(a_N+1)\right) . \qquad (13)
\end{aligned}$$

Note that $a_N+1=\frac{2|C_1||C_2|+|C_1|+|C_2|}{|C_1|+|C_2|}\leq(|C_1|+|C_2|)^2$, where C_1 and C_2 are the two children of N in the tree representation of R. Note also that $|C_1|+|C_2|\leq|I|$, and $|\text{i-node}(R)|=n(R)-1$. Putting this altogether with ineq. (13), we get $\sum_{N\in\text{i-node}(R)} b_N\leq n(R)\ln\frac{2}{\delta'}+2g\times n(R)\ln|I|$, as claimed. □

Using the notation of eq. (9), we get

$$B_N \leq s_{opt}\left(\ln\frac{2}{\delta'}+2g\ln|I|\right) . \qquad (14)$$

Putting altogether ineqs (7), (8), (9), (11) and (14), we get the statement of Lemma 2. □

The next Lemma upperbounds the last term in Lemma 1.

Lemma 5

$$\sum_{O\in\mathcal{S}}\frac{|O|}{|I|}(b(I(O))+b(O))\leq g\sqrt{\frac{2s_{opt}}{|I|Q}\left(\ln\frac{2}{\delta'}+g\ln|I|\right)} .$$

Proof: For the sake of clarity, fix $A=\sum_{O\in s^*(I),O\neq I(O)}\frac{|O|}{|I|}(b(I(O))+b(O))$. Remark that $O\subset I(O)$ implies $b(I(O))<b(O)$, and we get

$$A\leq\sum_{O\in s^*(I),O\neq I(O)}\frac{\sqrt{2}g}{|I|\sqrt{Q}}\sqrt{|O|\left(\ln\frac{2}{\delta'}+g\ln(|O|+1)\right)} .$$

Let us fix for short $A_1=\sum_{O\in s^*(I),O\neq I(O)}|O|$, $A_2=\sum_{O\in s^*(I),O\neq I(O)}\left(\ln\frac{2}{\delta'}+g\ln(|O|+1)\right)$. We get $A\leq\frac{\sqrt{2}g}{|I|\sqrt{Q}}\sqrt{A_1A_2}$, and few more inequations leads to the statement of the Lemma. □

Lemmas 1, 2, 5 put altogether end the proof of Theorem 3; they also show that the constant hidden in the $\mathcal{O}$ notation of Theorem 3 is small (actually $<\sqrt{6}$). Now, consider the result of Theorem 3. If we ignore logarithmic factors, then the upperbound on the error is driven by $g\sqrt{s_{opt}/(|I|Q)}$, a close order approximation to the optimal error [5].

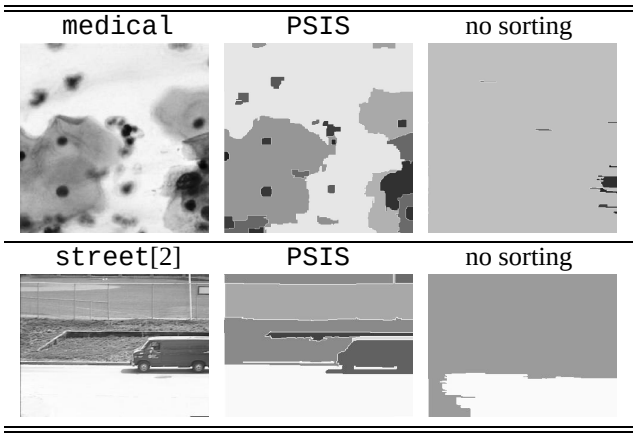

Figure 3. Sorting's importance. Regions obtained in the segmentations are grey-level averaged with white borders.

5. Color Images and Ordering

The extension of our analysis to **RGB** images is straightforward. The single color band of a pixel is replaced by a triple of values, each in the same range: $\{1, 2, ..., g\}$. The model of Section 2 is therefore replaced by a tri-model, each true pixel being described by a triple of Q r.v., for each of the **R**, **G** and **B** color levels. The homogeneity property is the same as the single color setting for each level $a \in \{\mathbf{R}, \mathbf{G}, \mathbf{B}\}$ taken apart. Therefore, the separability property postulates that at least one sum of expectations in $\{\mathbf{R}, \mathbf{G}, \mathbf{B}\}$ is different for adjacent true regions. The merging predicate for the **RGB** setting is [4]:

$$\mathcal{P}(R, R') = \begin{cases} \texttt{true} & \text{if} \quad \forall a \in \{\mathbf{R}, \mathbf{G}, \mathbf{B}\}, \\ & |\overline{R'}_a - \overline{R}_a| \leq b(R) + b(R') \\ \texttt{false} & otherwise \end{cases} .$$

Here, $\overline{R}_a$ denotes the observed average for color level a in region R. Provided the same assumption is made on algorithm $\mathcal{A}$ as in Section 4, our predicate preserves over-merging, and the same bound as that of Theorem 3 holds on the error if we measure it as the sum of errors over the three color levels. Only the constant multiplicative factor in the $\mathcal{O}$ notation slightly increases.

From an experimental point of view, following exactly $\mathcal{A}$ is impossible (we do not have access to I^*), but we can approximate it by a simple and fast algorithm. Suppose that I contains r rows and c columns. Then we have $2rc - r - c$ couples of adjacent pixels in 4-connexity, and as many merging tests to do in the worst case. We order these couples in increasing order of their absolute grey-level difference (for grey-level images), or in increasing order of their maximal **R**, **G**, **B** difference (for **RGB** images). Then, we test the merging between the corresponding regions of the pixels inside each couple, following this order, and using our merging predicate $\mathcal{P}$. Let us name PSIS this algorithm (after [4]). Since we do not update the order in PSIS when two regions are merged, the algorithm obtained is algorithmically very fast, and can be implemented without much programming tricks : for example, using radix sorting with color differences as the keys brings time complexity $\mathcal{O}(|I| \log(g))$ — linear in $|I|$ —. However, this choice which emphasizes complexity might seem to be quite a loose approximation for the order in $\mathcal{A}$ proned in Section 4. A simple experiment seems to display nice evidences that it is, fortunately, not the case.

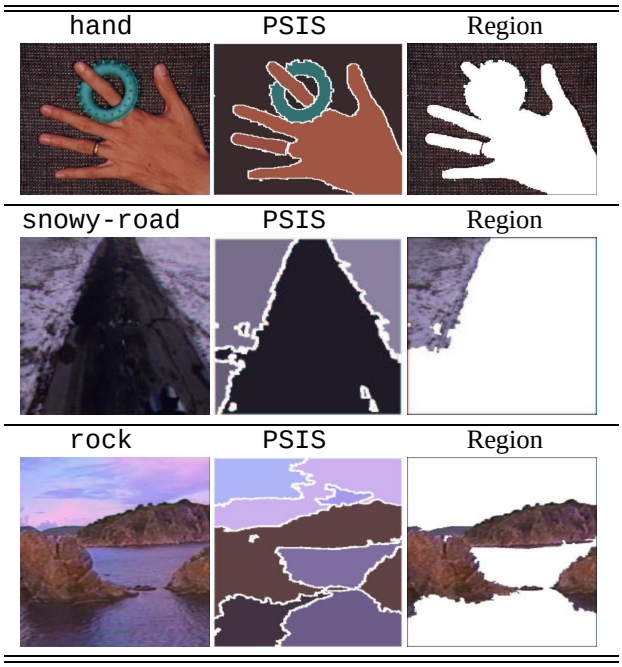

Figure 4. Results of PSIS on color images. The left column displays the original images, the center column is PSIS's results (regions are color averaged with white borders). The right column displays an example of region found by PSIS on the images.

It consists in comparing PSIS to the algorithm in which everything but the order is the same : the order in PSIS is replaced by a conventional scanning of the image covering the pixels from the top to the bottom of the image, and from the left to the right of the image (and we test the merging of the current pixel's region with those of its eventual left and up neighbors). Due to the lack of space, Figure 3 shows few eloquent experiments on grey-level images, in which PSIS clearly brings an advantage over conventional scanning (which, in addition, scarcely detects any region). Experiments were carried out with $Q = 32$, $\delta' = 1/|I|^2$, and without any pre-processing of the images.

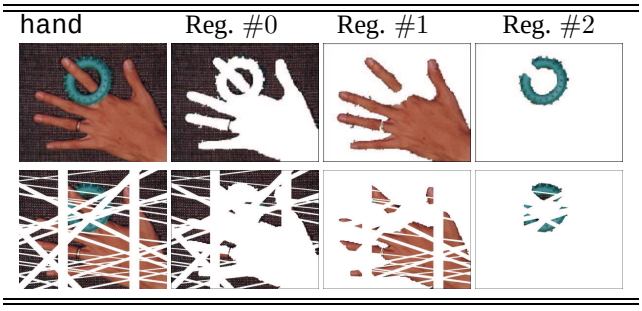

Figure 5. Occlusions handling on image `hand` (first row), and with many artificial random occlusions (second row). The largest regions found are shown (see text for details).

6. Experimental Results

Our implementation of `PSIS` to segment color images is a few kilobytes C code. All experiments are carried out without any pre-processing. Furthermore, we pick $Q = 32$ for all images (as proned in [4]), and $\delta' = 1/(3|I|^2)$ for **RGB** images. Therefore, there is no extensive domain- or image-dependent tuning of the parameters.

6.1. General Results

Results on the initial approach of `PSIS` can be found in [4]. Figure 4 presents few more results (see also Figure 1). Notice that on the `road` image, `PSIS` manages to segment two parts of what seems to be a vehicle, while it does not undermerge the snowy regions, with high variability.

6.2. Handling Occlusions

Suppose we make the hypothesis that the true regions of I^* can be occluded, that is, we relax the 4-connexity constraint. No part of our theoretical analysis relies on connexity: its removal keeps our properties on an extended model of image generation. This time, for each triple of expectations in (**R**, **G**, **B**) and set of pixels in I^* that match these expectations, the homogeneity constraint implies that the set is part of a true region of I^*, regardless of the position of its pixels in I^*. The modification of `PSIS` to cope with occlusions is simple: we first find the connex components of regions using `PSIS`, and then test merging between these components, regardless of connexity. In the next step, we run again `PSIS`, but with the couples of adjacent pixels replaced by all couples of regions found after the first step. In the order, the pixels color levels are replaced by the averages of color levels (of regions).

Figure 5, first row, shows image `hand` [1], with the three regions found by `PSIS` with occlusion handling. They approximate exactly the three perceptual regions of the image (hand, donut, and background). Following [6], we have tested occlusion handling on artificially occluded images, consisting in generating strips with variable orientation, width, and position in the image. However, where [6] generate a single fixed strip per image, we choose a heavier corruption by thirty random strips. Obviously, this hardens the task: 45.31% of the pixels in the image `hand` in Figure 5 (second row) are occluded. `PSIS` has managed to find an accurate segmentation of the `hand`, as four regions are found this time: the three original regions, plus a region containing all occlusions merged together (not shown in Figure 5).

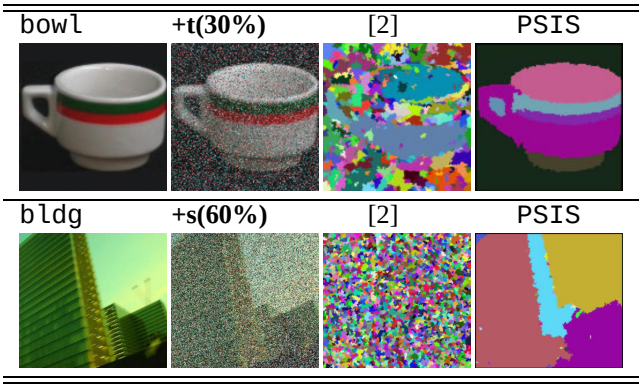

Figure 6. Sample results comparing the performances of the approach of [2] and `PSIS` on noisified images (left: images prior to noisification). Segmentation conventions are the same (region colors are chosen at random).

6.3. Handling Noise

Occlusions are sometimes presented as particular forms of noise generation [6]. We have chosen to study two types of hard noise corruption. Each is parameterized by a real $q \in [0, 1]$. Each color level $\in \{\mathbf{R}, \mathbf{G}, \mathbf{B}\}$ of each pixel $\in I$ is transformed with probability q into a new value, (i) chosen uniformly in $\{1, 2, ..., g\}$ for transmission noise (noted $\mathbf{t}(q)$), (ii) chosen uniformly in $\{1, g\}$ (the extremes) for salt and pepper noise (noted $\mathbf{s}(q)$). We have experienced that `PSIS` already handles significant amount of noise. Moreover, noisification of the image tends to increase dramatically its advantage over approaches such as [2], a phenomenon not previously reported. Two representative results are displayed in Figure 6. Notice the result of `PSIS` on the `bldg` corrupted by 60% of salt and pepper noise on each pixel, each channel. In the corrupted image, only roughly 6% of the pixels remain uncorrupted.

However, on some images, `PSIS` obtained a degradation of its performances for significant noise levels. Fortunately,

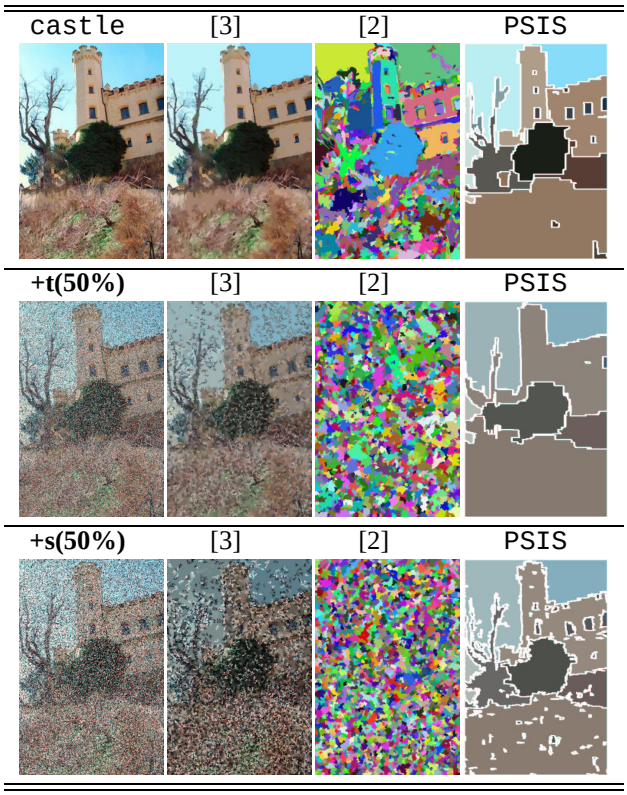

Figure 7. Results of [3, 2] and PSIS with our order modification. Left: original images, eventually corrupted by noise.

a simple modification of the algorithm makes it possible to increase experimentally its performances, by making the ordering prior to merges more robust. More precisely, the new ordering is not based on pixel color levels, but on averaged color levels, computed on a neighborhood of width Δ around each pixel (using the Manhattan distance).

Figure 7 reports results on the `castle` of Figure 1. Conventions for the segmentations results are as follows: [3]'s regions are averaged with the original colors, [2]'s are averaged with random colors, and `PSIS`'s follow [3]'s (with white bordered regions). Notice that the number of regions found by [3, 2] explode with corruption, a phenomenon which does not appear for `PSIS` modified. The segmentation time for the three algorithms gives a clear advantage to [2] and `PSIS` modified (non-optimized). Figure 8 shows more results on images taken from photographs or video, with larger noise corruption. In these cases, `PSIS` is the clear winner ([3, 2] obtain random results).

7. Acknowledgements

R. Nock would like to thank Sony CSL Tokyo for a generous invitation grant during which this work was finalized.

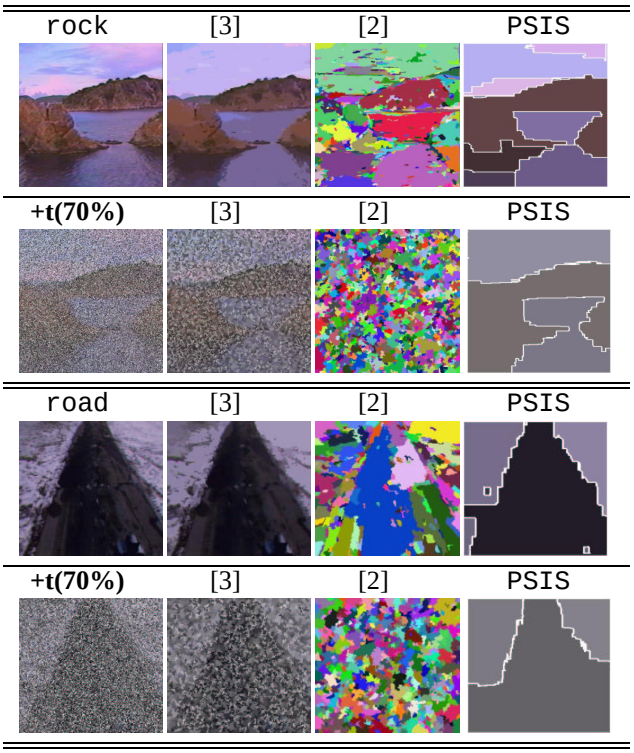

Figure 8. More noise results (*Cf* Fig. 7).

`PSIS` (Linux/Win) is available on the authors web pages.

References

[1] S.-C. Zhu and A. Yuille, "Region competition: unifying snakes, region growing, and bayes/MDL for multiband image segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 884–900, 1996.

[2] P. F. Felzenszwalb and D. P. Huttenlocher, "Image segmentation using local variations," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*. 1998, pp. 98–104, IEEE CS Press.

[3] D. Comaniciu and P. Meer, "Robust analysis of feature spaces: Color image segmentation," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*. 1997, pp. 750–755, IEEE CS Press.

[4] R. Nock, "Fast and Reliable Color Region Merging inspired by Decision Tree Pruning," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*. 2001, pp. 271–276, IEEE CS Press.

[5] M. J. Kearns and Y. Mansour, "A Fast, Bottom-up Decision Tree Pruning algorithm with Near-Optimal generalization," in *Proc. of the 15 th International Conference on Machine Learning*, 1998, pp. 269–277.

[6] D. Roth, M.-H. Yang, and N. Ahuja, "Learning to recognize objects," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*. 2000, pp. 724–731, IEEE CS Press.