Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

PERGAMON

Pattern Recognition 38 (2005) 835–846

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Semi-supervised statistical region refinement for color image segmentation

Richard Nock[a,*], Frank Nielsen[b]

[a]GRIMAAG-*Département Scientifique Interfacultaire, Université des Antilles-Guyane, Campus de Schoelcher, BP 7209, 97275 Schoelcher, Martinique, France*

[b]*Sony Computer Science Laboratories, Inc., 3-14-13 Higashi Gotanda, Shinagawa-Ku, Tokyo 141-0022, Japan*

Received 9 August 2004

Abstract

Some authors have recently devised adaptations of spectral grouping algorithms to integrate prior knowledge, as constrained eigenvalues problems. In this paper, we improve and adapt a recent statistical region merging approach to this task, as a non-parametric mixture model estimation problem. The approach appears to be attractive both for its theoretical benefits and its experimental results, as slight bias brings dramatic improvements over unbiased approaches on challenging digital pictures.

Keywords: Image segmentation; Semi-supervised grouping

1. Introduction

Grouping is the discovery of intrinsic clusters in data [1]. Image segmentation is a particular kind of grouping in which data consists of an image, and the task is to extract as regions the objects a user may find conceptually distinct from each other. The automation and optimization of this task face computational issues [2] and an important conceptual issue: basically, segmentation has access only to the descriptions of pixels (e.g. color levels) and their spatial relationships, while a user always uses higher level of knowledge to cluster the image objects. Without such a significant prior world knowledge, the accuracy of grouping is not meant to be optimality or even near-optimality, but rather accurate candidacy, as segmentation should come up with partition(s) from a candidate segmentation set [3].

With the advent of media making it easier and cheaper to collect and store digital images, unconstrained digital photographic images have raised this challenge even further towards both computational efficiency and robust processing. Consider for example the well-known benchmark image `lena` in Fig. 1. Users would certainly consider the hat of the girl as an object different from the blurred background, and most would consider her shoulder as different from her face. Nevertheless, due to the distribution of colors, it is virtually impossible for segmentation techniques based solely on low-level cues, such as the colors, to make a clean separation of these regions. The right image displays the result of our algorithm. The regions found have white borders. This result is presented more in depth in the experimental section (Fig. 2). Notice from the result the segmentation of the hat, cleanly separated from the background, and also the segmentation of the girl's chin, which is separated from her shoulder.

* Corresponding author. Tel.: +596 72 74 24; fax: +596 72 73 62.
E-mail addresses: richard.nock@martinique.univ-ag.fr (R. Nock), frank.nielsen@acm.org (F. Nielsen)
URLs: http://www.univ-ag.fr/∼rnock, http://www.csl.sony.co.jp/person/nielsen/.

0031-3203/$30.00
doi:10.1016/j.patcog.2004.11.009

Fig. 1. Image lena (left), and our segmentation (right). In the segmentation's result, regions found are white bordered (see text for details).

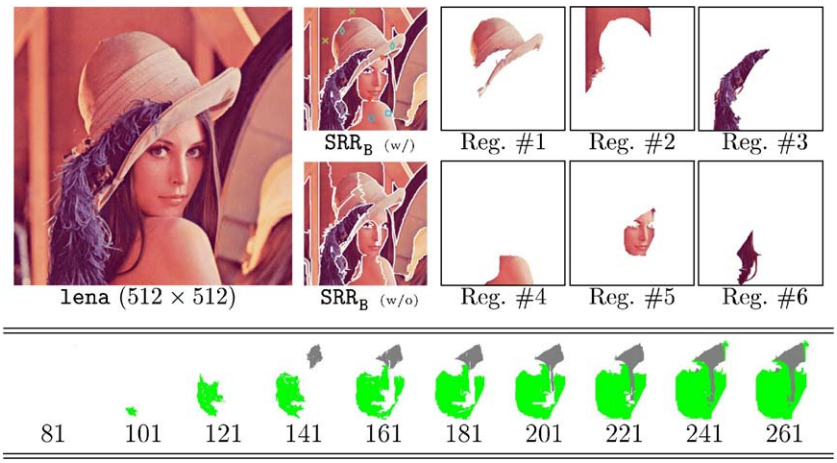

Fig. 2. Image lena (upper left), and its segmentation by SRR$_B$, without bias (w/o), and with bias (w/, see Fig. 1). In the segmentations' results, regions found are delimited with white borders. We have $m = 4$, $|V_1| = 2$, $|V_2| = 2$, $|V_3| = 2$, $|V_4| = 2$. The upper right table displays some of the largest regions extracted from the segmentation (Reg. #X). The bottom table shows how lena's face is built from the two models whose pixels, denoted by red triangles on the upper right image, have been pointed by the user (each model's color is random); the number below $\times 2000$ is SRR$_B$'s iteration number.

Common grouping algorithms for image segmentation use a weighted neighborhood graph to formulate the spatial relationships among pixels [1–6] and then formulate the segmentation as a graph partitioning problem. An essential difference between these algorithms is the locality of the grouping process. Shi and Malik [3] and Yu and Shi [1,7] solve it from a global standpoint, whereas Felzenszwalb and Huttenlocher [4], Nock [2] and Nielsen and Nock [5] make greedy local decisions to merge the connex components of induced subgraphs. Since segmentation is a global optimization process, the former approach is a priori a good candidate to tackle the problem, even when it faces computational complexity issues [3]. However, strong global properties can be obtained for the latter approaches, such as qualitative bounds on the overall segmentation error [4,2], or even quantitative bounds [5].

Our approach to segmentation, which gives the results of Fig. 1, is based on a segmentation framework previously studied by Yu and Shi [1,7]: grouping with bias. It is particularly useful for domains in which the user may interact with the segmentation, by inputting constraints to bias its result: sensor models in MRF [8], Human–computer interaction, spatial attention and others [1]. Grouping with bias is basically one step further towards the integration of the user in the loop, compared to the method of general expectations of a good segmentation integrated in non-purposive grouping [9]; it is solved by pointing in the image some pixels (the *bias*) that the user feel belong to identical/different objects, and then solving the segmentation as a constrained grouping problem: pixels with identical labels *must* belong to the same region in the segmentation's result, while pixels with different labels *must not* belong to the same region. The

solution for the global approach of Yu and Shi [1,7] is mathematically appealing, but it is computationally demanding, and it requires quite an extensive bias for good experimental results on small images. Furthermore, it makes it difficult to handle the constraints that some pixels must not belong to the same regions; that is why this technique is mainly used for the particular biased segmentation problem in which one wants to segregate some objects from their background.

In this paper, we propose a general solution to biased grouping, based on a local approach to image segmentation [2,5], which basically consists in using the bias for the estimation of non-parametric mixture models. Distribution-free processing techniques are useful, if not necessary, in grouping [10,2]. However, estimating clusters in data is already far from being trivial even when significative distribution assumptions are assumed [11]. This is where the bias is of huge interest when it comes to grouping, as bias defines partially labeled regions, with the constraint that different labels belong to different regions. The approach of Nock [2] and Nielsen and Nock [5] is also conceptually appealing for an adaptation to biased segmentation, because it considers that the observed image is the result of the sampling of a theoretical image, in which regions are statistical regions characterized by distributions. There is no distribution assumption on the statistical pixels of this theoretical image. The only assumption made is an homogeneity property, according to which the expectation of a single color is the same inside a statistical region, and it is different between two adjacent statistical regions. The segmentation problem is thus the problem of recognizing the partition of the statistical regions on the basis of the observed image. The biased grouping problem turns out to allow statistical regions to contain different statistical sub-regions, not necessarily connected, each satisfying independently the homogeneity property, *and* for which the user feels they all belong to the same perceptual object. Thus, it yields a significant generalization of the theoretical framework of Nock [2] and Nielsen and Nock [5].

Our contribution in this paper is twofold. First, it consists of two modifications and improvements to the unbiased segmentation algorithm of Nock [2] and Nielsen and Nock [5]. Their algorithm contains two stages. Informally, its first part is a procedure which orders a set of pairs of adjacent pixels, according to the increasing values of some real-valued function f. Its second part consists of a single pass on this order, in which it tests the merging of the regions to which the pixels belong, using a so-called merging predicate $\mathscr{P}$. f and $\mathscr{P}$ are the cornerstones of the approach of Nock [2] and Nielsen and Nock [5]. We propose in this paper a better f, and an improved $\mathscr{P}$ relying on a slightly more sophisticated statistical analysis. Our second contribution is the extension of this algorithm to grouping with bias. Our extension keeps both fast processing and the theoretical bounds on the quality of the segmentation. Experimentally, the results appear to be very favorable when comparing them to those of the approach of Yu and Shi [1,7]. They also appear to lead to dramatic improvements over unbiased grouping, when comparing our mostly automated biased grouping process to the human segmentations obtained on images of the Berkeley segmentation data set and benchmark images [12].

Section 2 summarizes the unbiased approach of Nock [2] and Nielsen and Nock [5], and presents our modification to their algorithm in the unbiased setting. Section 3 presents our extension to biased grouping and some theoretical results. Section 4 presents experimental results, and Section 5 concludes the paper.

2. Grouping exploiting concentration phenomena

We recall here the basic facts of the model of Nock [2] and Nielsen and Nock [5]. Throughout this paper, "log" is the base-2 logarithm. The notation $|\cdot|$ denotes the number of pixels (cardinality) when applied to a region R, or to the observed image I. Each pixel of I contains three color levels (**R**, **G**, **B**), each of the three belonging to the set $\{1, 2, \ldots, g\}$. The **RGB** setting is used to cast our results directly on the same setting as Nock [2] and Nielsen and Nock [5]; however, the versatile technique of Nock [2] and Nielsen and Nock [5] can be tailored to other numerical feature description spaces, and handles more complex formulations of the color gamuts, such as CIE, $L*u*v*$, **HSI**, etc. as well as channel sampling rates.

2.1. The model and algorithm of Nock [2] and Nielsen and Nock [5]

The image I is an observation of a perfect object (or "true region", or statistical region) scene I^* we do not know of, and which we try to approximate through the observation of I. It is I^* which captures the global properties of the scene: theoretical (or statistical) pixels are each represented by a set of Q *distributions* for each color level, from which each of the observed color level is sampled. The statistical regions of I^* satisfy a 4-connexity constraint, and the simple homogeneity constraint that the **R** (resp., **G**, **B**) expectation is the same inside a statistical region. In order to discriminate regions, we assume that between any two adjacent regions of I^*, at least one of the three expectations is different. It is important that, apart from an additional independency, no more assumptions are put on I^*: for instance, the distributions can *all* be different for all statistical pixels, thereby contrasting with usual statistical models used in image segmentation, involving hypotheses that can be quite restrictive [10]. Q is a parameter which quantifies the complexity of the scene, the generality of the model, and the statistical hardness of the task as well. If Q is small, the model gains in generality ($Q = 1$ brings the most general model), but segmenting the image is more difficult from a statistical standpoint. Experimentally, Q turns out to be a tunable parameter which controls the coarseness of the segmentation, even if one value in [2,5] ($Q = 32$) appears to be sufficient to obtain nice segmentations for a large body of images.

From this model, Nock [2] and Nielsen and Nock [5] obtain a merging predicate $\mathcal{P}(R, R')$ based on concentration inequalities, to decide whether two observed regions R and R' belong to the same statistical region of I^*, and thus have to be merged. Let $\overline{R}_a$ denote the observed average for color a in region R of I, and let $\mathcal{R}_l$ be the set of regions with l pixels. Let

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

Ref. [2] pick $|\mathcal{R}_l| = (l+1)^g$. The merging predicate is [2]

$$\mathcal{P}(R, R') = \begin{cases} \texttt{true} & \text{iff } \forall a \in \{\mathbf{R}, \mathbf{G}, \mathbf{B}\}, |\overline{R'}_a - \overline{R}_a| \\ & \leqslant b(R) + b(R'), \\ \texttt{false} & \text{otherwise.} \end{cases} \quad (2)$$

The description of the algorithm probabilistic-sorted image segmentation (PSIS) of Nock [2] is straightforward, as exposed in Algorithm 1. It basically consists in making a preliminary sorting over the set S_I of the pairs of adjacent pixels of the image, according to the increasing values of a real-valued function $f(p, p')$. $f(.,.)$ takes a pair of pixels p and p' as input, and returns the maximum of the three color differences (**R**, **G** and **B**) in absolute value between p and p':

$$f(p, p') = \max_{a\in\{\mathbf{R},\mathbf{G},\mathbf{B}\}} |p'_a - p_a|. \quad (3)$$

In 4-connexity, the number of such pairs is $|S_I| = 2r_I c_I - r_I - c_I = \mathcal{O}(|I|)$ if I has r_I rows and c_I columns. After ordering, the algorithm traverses this order only once, and test for any pair (p_i, p'_i) the merging of the two regions to which they currently belong, $R(p_i)$ and $R(p'_i)$ (here, the subscript i denotes the rank of the couple in the order). Such an algorithm is called region-merging, since it gradually merges regions, pixels being taken as elementary regions.

Algorithm 1: PSIS(I)

Input: an image I
$S'_I = \texttt{Order\_increasing}(S_I, f)$;
for $i = 1$ ***to*** $|S'_I|$ **do**
  if $R(p_i) \neq R(p'_i)$ ***and*** $\mathcal{P}(R(p_i), R(p'_i)) =$ ***true*** **then**
    Union $(R(p_i), R(p'_i))$;

This approach to segmentation is interesting from the algorithmic standpoint, because it is both simple and fast. The eigenvalue approach of Shi and Malik [3] admits a naive solution whose time complexity is $\mathcal{O}(|I|^3)$—impractical even for moderate-sized images—. Mathematical tricks can scale down the time complexity at the expense of greater implementation efforts, but it still lies somewhere in between $\mathcal{O}(|I|\sqrt{|I|})$ and $\mathcal{O}(|I|^3)$. In the case of Algorithm 1, the **for...to** complexity is $\mathcal{O}(|I|)$ (linear). Fortunately, the order is also cheap as radixsorting with $f(.,.)$ values as the keys brings a time complexity $\mathcal{O}(|I|\log g)$. Since g can be considered constant, we get an overall approximate linear-time complexity for the whole algorithm using an Union-Find implementation [13].

We now concentrate on our modifications to f and the merging predicate $\mathcal{P}$.

2.2. An improved f

Nielsen and Nock [5] have shown that f should theoretically be an estimator, as reliable as possible, of the local between-pixel gradients. Eq. (3) is the simplest way to compute these estimators, but there is another choice with which we have obtained yet better visual results. It consists in extending convolution kernels classically used in edge detection for pixel-wise gradient estimation. In 4-connexity, neighbor pixels are either horizontal or vertical. Thus, we only need $\hat{\partial}_x$ or $\hat{\partial}_y$ between neighbor pixels p and p', for each color channel $a \in \{\mathbf{R}, \mathbf{G}, \mathbf{B}\}$. A natural choice is to extend the Sobel convolution filter to the following kernels:

$$\hat{\partial}_x : \begin{bmatrix} -1 & 0 & 0 & 1 \\ -2 & 0 & 0 & 2 \\ -1 & 0 & 0 & 1 \end{bmatrix},$$

$$\hat{\partial}_y : \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}.$$

For example, whenever neighbor pixels p and p' are horizontal, $\hat{\partial}_x$ is used and the two pixels in the convolution filter are located in the second row, and the second and third columns. Only for the pixels for which the estimations with $\hat{\partial}_x$ and $\hat{\partial}_y$ cannot be done (i.e. those of the image border) do we keep the estimation of Eq. (3).

2.3. An improved merging predicate $\mathcal{P}$, and associated theoretical results

Our first result is based on the following theorem:

Theorem 1 (*The independent bounded difference inequality, McDiarmid [14]*). *Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a family of independent r.v. with X_k taking values in a set A_k for each k. Suppose that the real-valued function h defined on $\prod_k A_k$ satisfies $|h(\boldsymbol{x}) - h(\boldsymbol{x}')| \leqslant c_k$ whenever vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ differ only in the k-th coordinate. Let μ be the expected value of the r.v. $h(\boldsymbol{X})$. Then for any $\tau \geqslant 0$,*

$$\boldsymbol{Pr}(h(\boldsymbol{X}) - \mu \geqslant \tau) \leqslant \exp\left(-2\tau^2 \Big/ \sum_k (c_k)^2\right). \quad (4)$$

From this theorem, we obtain the following result on the deviation of observed differences between regions of I. Recall that $\overline{R}_a$ denotes the observed average of color a for region R.

Theorem 2. *Consider a fixed couple (R, R') of regions of I. $\forall 0 < \delta \leqslant 1$, $\forall a \in \{\boldsymbol{R}, \boldsymbol{G}, \boldsymbol{B}\}$ the probability is no more*

than δ *that*

$$|(\overline{R}_a - \overline{R}'_a) - \boldsymbol{E}(\overline{R}_a - \overline{R}'_a)| \geqslant g\sqrt{\frac{1}{2Q}\left(\frac{1}{|R|}+\frac{1}{|R'|}\right)\ln\frac{2}{\delta}}.$$

Proof. Suppose we shift the value of the outcome of one r.v. among the $Q(|R|+|R'|)$ possible r.v. of the couple (R, R'). $|\overline{R}_a - \overline{R}'_a|$ is subject to a variation of at most $c_R = g/(Q|R|)$ when this modification affects region R (among $Q|R|$ possible r.v.), and at most $c_{R'}=g/(Q|R'|)$ for a change inside R' (among $Q|R'|$ possible r.v.). We get $\sum_k (c_k)^2 = Q(|R|(c_R)^2 + |R'|(c_{R'})^2) = (g^2/Q)((1/|R|)+(1/|R'|))$. Using the fact that the deviation with the absolute value is at most twice that without, and using Theorem 1 (solving for τ) brings our result. This ends the proof of Theorem 2. □

Fix some $a \in \{\mathbf{R}, \mathbf{G}, \mathbf{B}\}$. Provided we have some way to evaluate the theoretical deviation of $|\overline{R}_a - \overline{R}'_a|$, a candidate merging predicate is straightforward. Nock [2] and Nielsen and Nock [5] use a concentration inequality on the deviation of the average color level around its expectation for any single arbitrary region R. Thus, they get for each region R probabilistic bounds on the deviation of $|\overline{R}_a - \mathbf{E}(\overline{R}_a)|$. When two adjacent regions R and R' come from the same true region in I^*, we have $\mathbf{E}(\overline{R}_a) = \mathbf{E}(\overline{R}'_a)$; thus the triangular inequality makes that the deviation of $|\overline{R}_a - \overline{R}'_a|$ is no more than the sum of deviations for each region's color around its expectation, and we get the merging predicate of Nock [2] and Nielsen and Nock [5] in Eq. (2). However, the use of the triangular inequality weakens the concentration bound. Notice that when R and R' come from the same true region in I^*, Theorem 2 directly yields a probabilistic bound on the deviation of $|\overline{R}_a - \overline{R}'_a|$ (since $\mathbf{E}(\overline{R}_a - \overline{R}'_a) = 0$), without a similar weakening.

However, Theorem 2 is a single event's concentration in what it considers a single couple of regions (R, R'), and one should extend this to the whole image, in order to obtain a convenient merging predicate. Fortunately, one can easily upperbound the probability that such a large deviation occurs in the observed image I, using the union bound. Nock [2] remark that this is no more than the probability to occur in the set of all regions (whether present *or* absent from I). The cardinal of each subset containing fixed-size regions can be upperbounded by a degree-$\mathcal{O}(g)$ polynomial [2], but it yields to a pretty large upperbound.

Another counting argument is possible, if we remark that the occurrence probability on I is also no more than that measured on the set of couples of I whose merging is tested, S'_I (See Algorithm 1). Its cardinal, $|S'_I|$, is comparatively small: for a single-pass algorithm, $|S'_I| < |I|^2$, and even $|S'_I| = \theta(|I|)$ in 4-connexity. Thus, we get the following theorem. □

Theorem 3. $\forall 0 < \delta \leqslant 1$, *there is probability at least* $1 - (3|S'_I|\delta)$ *that all couples* (R, R') *tested shall verify* $\forall a \in \{\boldsymbol{R}, \boldsymbol{G}, \boldsymbol{B}\}$, $|(\overline{R}_a - \overline{R}'_a) - \boldsymbol{E}(\overline{R}_a - \overline{R}'_a)| \leqslant b(R, R')$, *with*

$$b(R, R') = g\sqrt{\frac{1}{2Q}\left(\frac{1}{|R|}+\frac{1}{|R'|}\right)\ln\frac{2}{\delta}}. \tag{5}$$

$b(R, R')$ would lead to a very good theoretical merging predicate $\mathscr{P}$ instead of using the threshold $b(R) + b(R')$ in Eq. (2), provided we pick a δ small enough. This predicate would be much better than [2,5] from the theoretical standpoint, but slightly larger thresholds are possible that keep all the desirable theoretical properties we look for, and give much better visual results. Our merging predicate uses one such threshold, which turns out to be $\tilde{\mathcal{O}}(b(R, R'))$ (the tilde upon the big-Oh notation authorizes to remove constants and log-terms). Remark that provided regions R and R' are not empty, $b(R, R') \leqslant \sqrt{b^2(R) + b^2(R')} < b(R) + b(R')$. This right quantity is that of Eq. (2). The center quantity is the one we use for our merging predicate. Notice that it is indeed $\tilde{\mathcal{O}}(b(R, R'))$ provided a good upperbound on $|\mathscr{R}_l|$ is used. Our merging predicate is thus:

$$\mathscr{P}(R, R') = \begin{cases} \texttt{true} & \text{iff } \forall a \in \{\mathbf{R}, \mathbf{G}, \mathbf{B}\}, |\overline{R'}_a - \overline{R}_a| \\ & \quad \leqslant \sqrt{b^2(R) + b^2(R')}, \\ \texttt{false} & \text{otherwise.} \end{cases} \tag{6}$$

Let us now concentrate on $|\mathscr{R}_l|$. Nock [2] and Nielsen and Nock [5] pick $|\mathscr{R}_l| = (l+1)^g$, considering that a region is an unordered bag of pixels (each color level is given $0, 1, \ldots, l$ pixels). This bound counts numerous duplicates for each region: e.g. at least $(l+1)^{g-l}$ when $l < g$. Thus, we fix $|\mathscr{R}_l| = (l+1)^{\min\{l,g\}}$.

As advocated before, the merging predicate in Eq. (6) has proven experimentally to be more interesting than that of Eq. (2). Let us briefly illustrate its theoretical interests, that encompass those of the merging predicate of Nock [2] and Nielsen and Nock [5] (due to the fact that it is tighter).

Suppose that we are able to make the merging tests through f 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. Let us name **A** this assumption. Informally, **A** stresses the need to make f as accurate as possible (and it has motivated our study of Section 2.2). Notice also that **A** does not postulate *at all* that we know where the statistical regions of I^* are in I. Under assumption **A**, Nielsen and Nock [5] have shown that with high probability, the error of the segmentation is limited from both the qualitative *and* the quantitative standpoint. There are basically three kind of errors a segmentation algorithm can suffer with respect to the optimal segmentation. First, under-merging represents the case where one or more regions obtained are strict subparts of statistical regions. Second, over-merging represents the case where some regions obtained strictly contain more than one statistical region. Third, there is the "hybrid" (and most probable) case where some regions obtained contain more than one strict subpart of true regions. Name for short $s^*(I)$ as the set of regions of the ideal segmentation of

I following the statistical regions of I^*, and $s(I)$ the set of regions in the segmentation we get. Then we have the following qualitative bound on the error.

Theorem 4. *With probability* $\geqslant 1-\theta(|I|\delta)$, *the segmentation on* I *satisfying* **A** *is an over-merging of* I^*, *that is*: $\forall O \in s^*(I), \exists R \in s(I) : O \subseteq R$.

Proof. From Theorem 3, with probability $>1-(3|S'_I|\delta)$ (thus $>1-\theta(|I|\delta)$ in 4-connexity), all regions R and R' coming from the same statistical region of I^*, whose merging is tested, satisfy $\forall a \in \{\mathbf{R},\mathbf{G},\mathbf{B}\}$, $|\overline{R}_a - \overline{R}'_a| \leqslant b(R,R') \leqslant \sqrt{b^2(R)+b^2(R')}$ (thus, $\mathscr{P}(R,R') =$ `true`). Since **A** holds, the segmentation obtained is an over-merging of I^*. □

Because of the choice of our $\mathscr{P}$, it satisfies also the quantitative bounds of Nielsen and Nock [5], i.e. we can show that with high probability, the over-segmentation of Theorem 4 shall not hopefully result in a too large error. We refer the reader to Nielsen and Nock [5] for the explicit values of this error bound, not needed here. In the sequel, we shall refer to our modified version of PSIS as $\mathsf{SRR_B}$, which stands for Statistical Region Refinement (with Bias). We have chosen to use refinement better than merging, because it shall be clear from the experiments that the bias may help to improve both the merging and the splitting of regions, even when $\mathsf{SRR_B}$ formally belongs to region merging algorithms.

3. Grouping with bias

It shall be useful in this section to think of I as containing vertices instead of pixels, and the (4-)connexity as defining edges, so that I can be represented by a simple graph (V, E). Following Yu and Shi [1], we define a *grouping bias* to be user-defined disjoint subsets of V: $\{V_1, V_2, \ldots, V_m\} = \mathscr{V}$. Any feasible solution to the constrained grouping problem is a partition of V into connex components (thus, a partition of the pixels of I into regions), such that

(i) any such connex component intersects at most one element of $\mathscr{V}$, and
(ii) $\forall 1 \leqslant i \leqslant m$, any element of V_i is included into one connex component.

The first condition states that no region in the segmentation of I may contain elements from two distinct subsets in $\mathscr{V}$, and condition (ii) states that each vertex in $\mathscr{V}$ belongs to a region.

For any region R of I, we define a *model* for the region to be a subset of R, without connexity constraints on its elements, containing *one* vertex of some element of $\mathscr{V}$. The term model makes statistical sense because any $V_i \in \mathscr{V}$ with $|V_i| > 1$ may represent a single object for the user, but composed of different statistical regions ("models") of I^*. Each element of V_i can thus represent one theoretical pixel of each different statistical sub-regions, whose union makes the object perceived by the user. For example, the hat of `lena` in Fig. 1 visually contains two parts belonging to the same conceptual object (a bandeau and the hat itself). There is a significant gradient between these two parts of the hat. So, one may imagine to create some $V_i \in \mathscr{V}$ by pointing two pixels, one on the top of the hat, and one somewhere on the bandeau (or on the bright hat's border whose color is similar to the bandeau). This indicates that these two parts with different colors define two models that belong to the same object, and should thus be considered as a single region in the segmentation's result (Notice that this is indeed the case from our biased segmentation's result in Fig. 1.)

Grouping with bias has another advantage, since it naturally handles occlusions, as the user may specify that two (or more) models not connected represent the same region. This could be the case in Fig. 1 for `lena`'s hair for example.

Any region R defined by some $V_i \in \mathscr{V}$ is a partition of models, each represented by one element of V_i. We name regions without vertices from $\mathscr{V}$ as "model-free". There are therefore two types of regions in our segmentation: those model-free, and those being a partition of sub-regions, each defined by the elements of one subset in $\mathscr{V}$. Our region merging algorithm keeps this as an invariant: thus, merging a model-free region and a model-based region results in the merging of the first region into one model of the second. The modification of the approach in Refs. [2,5] consists in first making each V_i defined by the user, and then, through the traversing of S'_I, replacing the merging stage (the **if** condition in the **for...to** of Algorithm 1) by the following new **if** condition ($\forall (p,p') \in S'_I$):

if If $R(p)$ and $R(p')$ are model-free, then we compute $\mathscr{P}(R(p),R(p'))$ as in Algorithm 1, and eventually merge them;
else if both contain models, then we do not merge them; Indeed, in that case, either the models are defined by vertices of different subsets in $\mathscr{V}$ (and we obviously do not have to merge them), or they are defined by vertices of the same subset of $\mathscr{V}$. However, in that case, they have been defined by the used as different sub-regions of the same object, so we keep these sub-regions distinct until the end of the algorithm.
else consider without loss of generality that $R(p)$ contains models and $R(p')$ does not. We first compute $\mathscr{P}(M(p),R(p'))$, with M the model of $R(p)$ adjacent to $R(p')$ (notice that $p \in M(p)$):
  if it returns `true`, then a merge is done: we fold $R(p')$ into $M(p)$; thus, $M(p)$ grows (and not the other models of $R(p)$), as after this merging it integrates $R(p')$.
  else we search for the best matching model $M(p)$ of $R(p)$ w.r.t. $R(p')$ (the one minimizing $\max_{a\in\mathbf{R},\mathbf{G},\mathbf{B}} |\overline{M(p)}_a - \overline{R(p')}_a|$), and eventually fold $R(p')$ into $M(p)$ iff $\mathscr{P}(M(p),R(p'))$ returns `true`.

At the end of the algorithm, all models of each V_i are merged altogether in the segmentation's output.

The theoretical properties enjoyed by this extension to grouping with bias are the same as the unbiased approach given in Section 2, provided we make the assumption that all the vertices of the subsets in $\mathscr{V}$ come from different statistical regions of I^*. Recall that this is sound with the fact that the aim of bias is precisely to make it possible for the observer to integrate in the same perceptual object different statistical (true) regions of I^*. For example, we have:

Theorem 5. *Suppose that the vertices of $\mathscr{V}$ come from $\sum_{i=1}^{m}|V_i|$ different statistical regions of I^*. Then, with probability $\geqslant 1-\theta(|I|\delta)$, the segmentation on I satisfying $\boldsymbol{A}$ is an over-merging of I^*.*

Quantitative bounds on the are also possible, following Nielsen and Nock [5].

4. Experiments

We have run SRR$_\mathsf{B}$ on a benchmark of digital pictures of various contents and difficulty levels, to test its ability to improve the segmentation quality over the unbiased approach, while using a bias as slight as possible. While looking at the experiments performed on Figs. 2–7 , the reader may keep in mind that images are segmented as they are, i.e. without any preprocessing, and with the same value for the parameters, following Nock [2] and Nielsen and Nock [5]:

$$Q = 32, \tag{7}$$

$$\delta = \frac{1}{3|I|^2}. \tag{8}$$

Thus, the quality of the results may be attributed only to SRR$_\mathsf{B}$, and not to any content-specific tuning or preprocessing optimization.

4.1. SRR$_\mathsf{B}$'s model choice rules, and first experiments

Fig. 2 shows detailed results on the image `lena`, that have been previously outlined in the introduction. We have chosen this image because it is one of the mostly used benchmark in image processing, and it has features making it difficult to segment, such as its blurred regions, its thin differences between perceptually distinct regions (e.g. her face and her shoulder), its single majority tone (reddish), its strong gradients (e.g. her hat). The result of SRR$_\mathsf{B}$ with bias displays four model-based regions. Each such region is defined only by two models. In the result of SRR$_\mathsf{B}$ with bias, the symbols denote the pixels that have been pointed by the user to define each element of $\mathscr{V}$. Different symbols denote different elements of $\mathscr{V}$, and thus different perceptual regions for the user.

The result of SRR$_\mathsf{B}$ with bias displays a very good segmentation of the image. The girl's hat and her shoulder, two very difficult regions to segment, are almost perfectly segmented. The quality of the segmentation is to be evaluated in the light of the bias given. The user has specified only eight pixels for the whole image. Obviously, these pixels have not been chosen at random.

Two simple "rules of thumb" appear to be enough for a limited processing of most of the images, while obtaining the significant improvements over unbiased segmentation observed in `lena`, and over images as well such as those of Fig. 3.

The first and most important is a *gradient rule*: the subregions of smoothest gradients between two perceptually distinct regions are good places for specifying models. This prevents the order to merge distinct perceptual sub-regions in the early steps of SRR$_\mathsf{B}$, during which the statistical accuracy of the merging predicate is the smallest. In image `lena`, the two red triangles defining her face are approximately located in smooth gradient parts between the face and the hat, and between the face and the shoulder respectively.

The second is a *size rule*: during the merging steps, no two models mix altogether into a single model, even when they belong to the same region. Furthermore, the merging of a model-free region into a model of another region does not necessarily imply that they are connected, as connexity is ensured only with the region containing the model. Thus, if the user specifies very small models in perceptually distinct regions, this may yield a significant over-merging for the regions to which such model belongs. For example, if we had put a red triangle in `lena`'s right eye, her hair would have been merged with her face. Such a mistake does not really represent a real additional interaction burden, as putting a single different additional model in the hair would solve the problem. However, this simple rule of leaving model-free too small regions may save a significant number of models, and thus reduce the interaction time for the user.

For relatively "easier" pictures, such as those of Fig. 3, sparse and simple choices yield dramatic improvements over unbiased grouping: the stairs and the tower of `castle-1` are almost perfectly segmented, and the castle of `castle-2` is almost perfectly extracted as a whole (notice the model in the drainpipe, which prevents it to be merged with the bushy tree).

In the next subsection, we make further comparisons of SRR$_\mathsf{B}$ with biased normalized cuts (`NCuts`).

4.2. SRR$_\mathsf{B}$ vs. `NCuts`

Yu and Shi [1,7] have proposed a mathematically appealing extension of the original unbiased normalized cuts problem of Shi and Malik [3]. In the original problem, the image is transformed into a weighted graph, and the objective is to make a partition of this graph into a fixed number of connex components, so as to minimize the cut between the components and maximize an association (within-component)

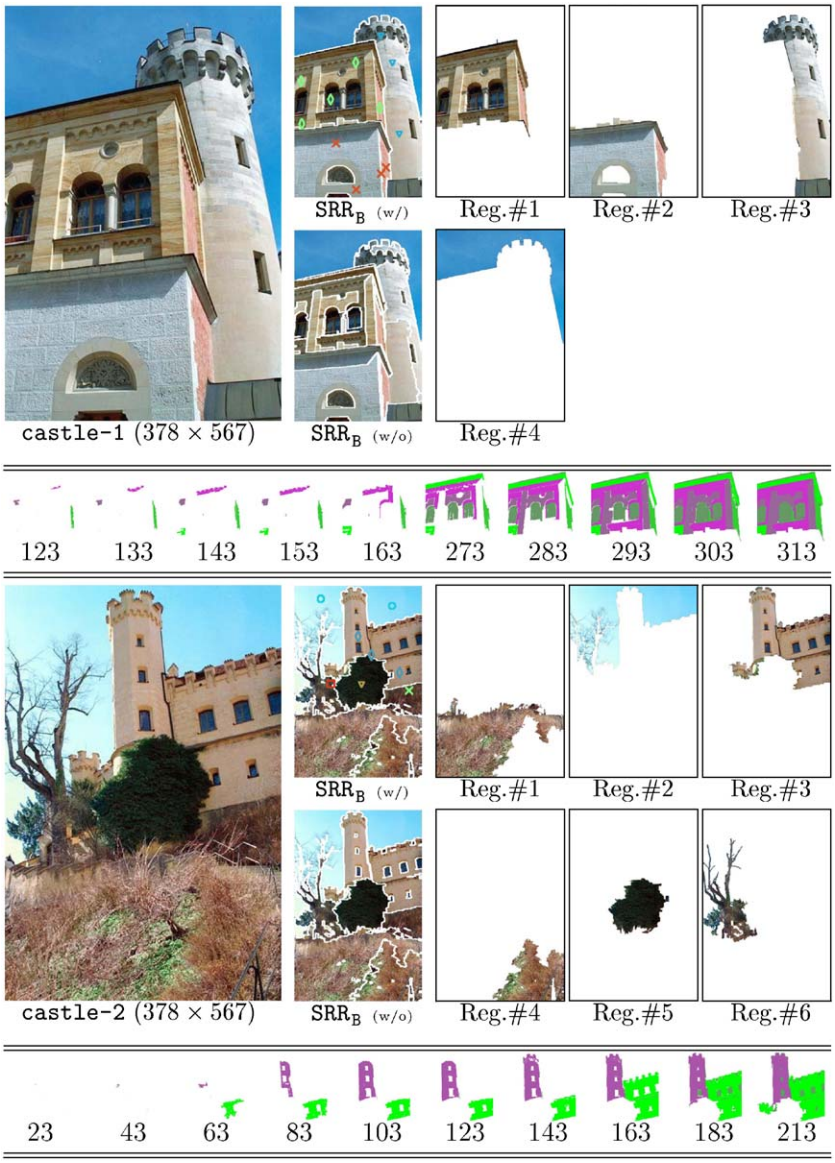

Fig. 3. Segmentation results on images `castle-1` ($m = 3, |V_1| = 5, |V_2| = 4, |V_3| = 3$) and `castle-2` ($m = 5$, $|V_1| = 3, |V_2| = 1, |V_3| = 1, |V_4| = 1, |V_5| = 2$). Conventions follow Fig. 2.

measure. The solution of this problem, even when computationally intractable [3], can be fairly well approximated by the eigen decomposition of a stochastic matrix. The extension of this technique to biased grouping involves constrained eigenvalue problems. These problems make it difficult to handle non-transitive constraints, such as the constraint "must not belong to the same region", which belongs to the core of the biased grouping problem (see Section 3). In the Refs. [1,7], the bias takes the form of transitive constraints, such as "must belong to the same region". In that case, nothing is assumed for the pixels with different labels. In order to make a fair comparison with `NCuts`, we have decided to study a particular biased grouping problem whose constraints fit in this category, and for which the `NCuts` give some of their best results [7]: the segregation of the foreground from the background of an image. This problem is addressed on `NCuts` by making a frame (10 pixels width in most of our experiments) on an image, and then constrain the grouping to treat this frame as a single region.

Fig. 4 presents some results that have been obtained for three animal pictures, in which we want to segregate the animal from its background. Fig. 5 presents additional re-

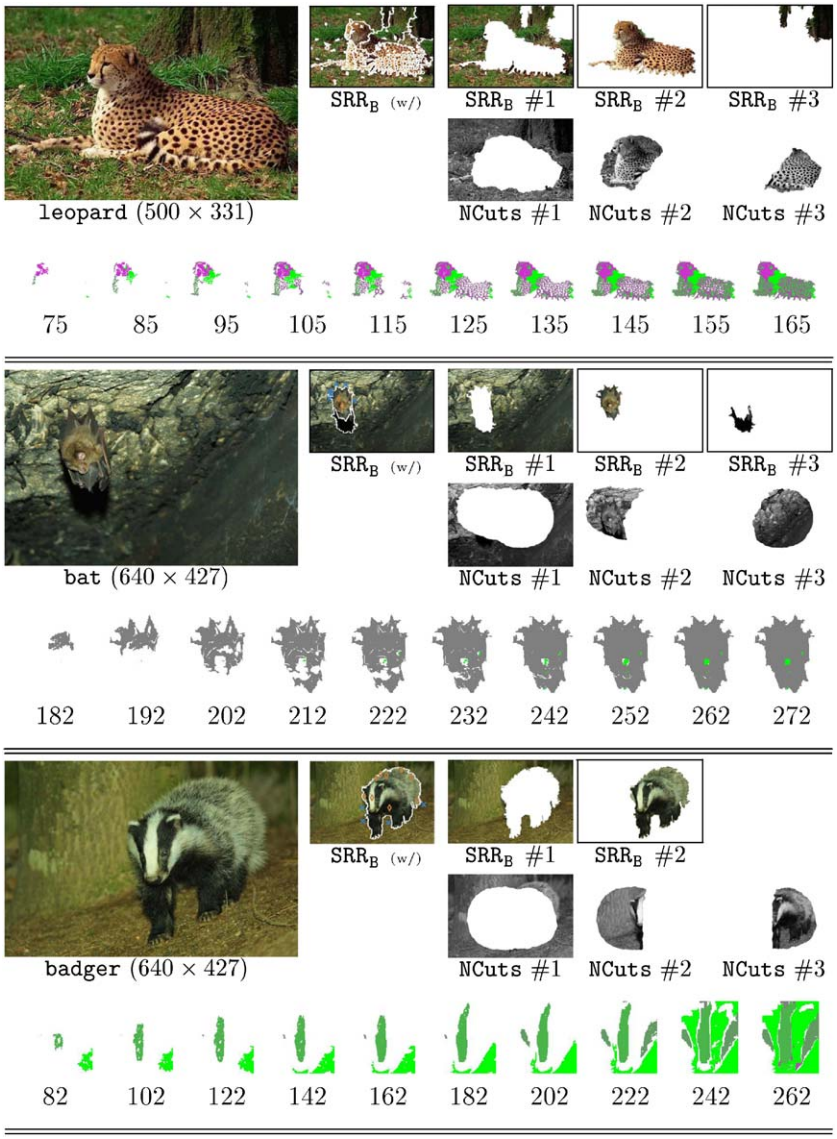

Fig. 4. $\mathtt{SRR_B}$ vs. NCuts on images leopard (m=2, $|V_1|$=4, $|V_2|$=5), bat (m=2, $|V_1|$=2, $|V_2|$=7) and badger (m=2, $|V_1|$=10, $|V_2|$=4). For each image, the result of $\mathtt{SRR_B}$ with bias and the largest regions found by $\mathtt{SRR_B}$ and NCuts are shown. The remaining regions follow the convention of Fig. 2.

sults on an animal, and on a flower. The results display that the NCuts perform reasonably well, but fail to make accurate segmentations of regions with deep localized contrasts, such as the head of the badger, the speckled coat of the leopard, or the flower. Because these contrasts make very small local cut values for NCuts, they tend to be selected as region frontiers, and transitive constraints do not seem to be enough for preventing their split. This is clearly a drawback that $\mathtt{SRR_B}$ does not suffer, as it manages very accurate segmentations of all animals. Even the bee of the flower image receives an accurate segmentation, which segregates the insect from both the background and the flower, while NCuts essentially succeed only in making an accurate segregation of the insect from the background. This flower image is an example of a difficult image for grouping algorithms based only on low-level cues: due to the distribution of colors, it is virtually impossible to make a single region out of the flower, whose colors are contrasted, *while* preventing the bee to be merged with the background. Eight models pointed are enough to solve this problem in $\mathtt{SRR_B}$.

Since our merging predicate relies on comparing observed averages, making segmentations without bias in $\mathtt{SRR_B}$ faces the problem of under merging for regions with strong

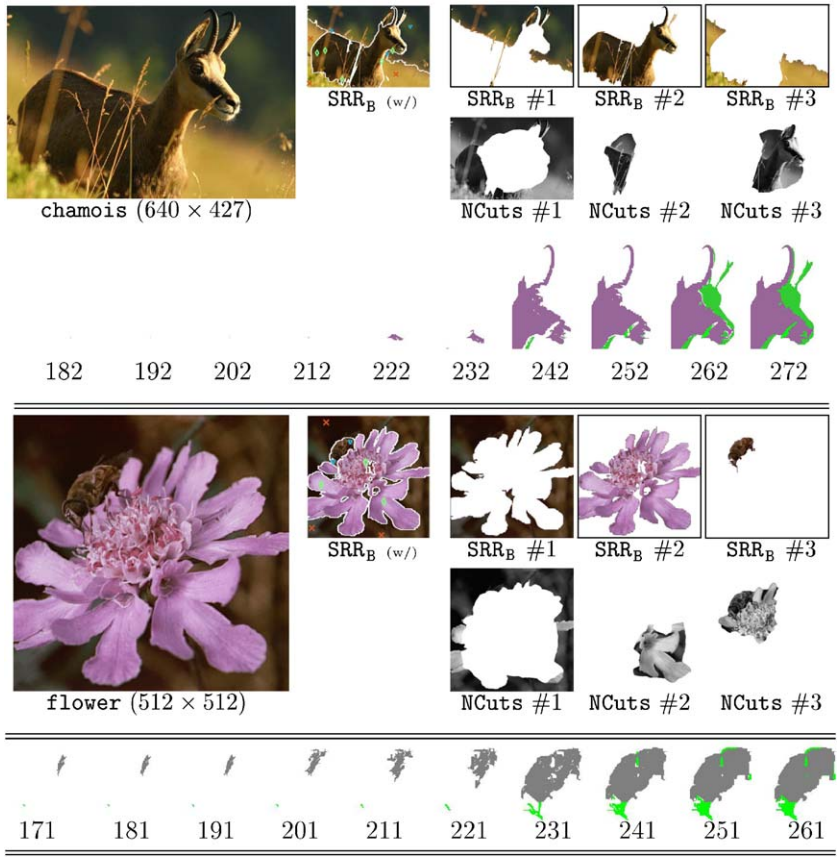

Fig. 5. More results on images `chamois` ($m=3$, $|V_1|=5$, $|V_2|=3$, $|V_3|=3$) and `flower` ($m=3$, $|V_1|=3$, $|V_2|=3$, $|V_3|=2$). Conventions follow Fig. 4.

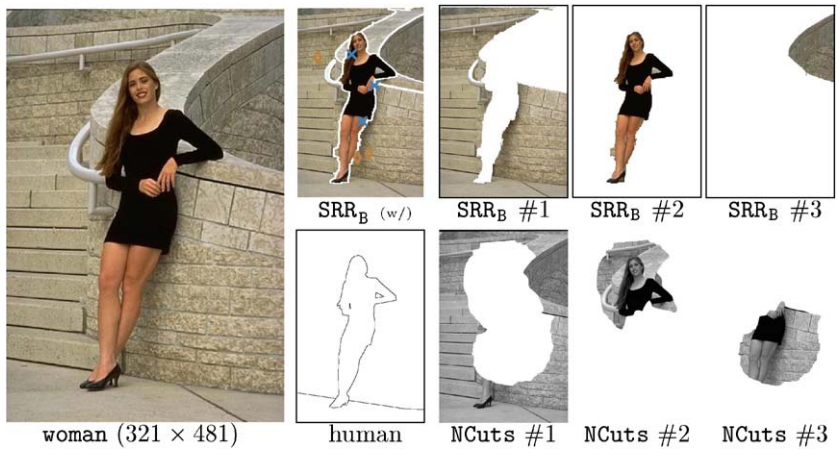

Fig. 6. Comparison of $\mathtt{SRR_B}$ and `NCuts` on the BSDB image `woman` ($m=2$, $|V_1|=3$, $|V_2|=3$). The "human" result is a segmentation from a human, taken from the BSDB (regions are displayed white with black borders). Other conventions follow Fig. 4.

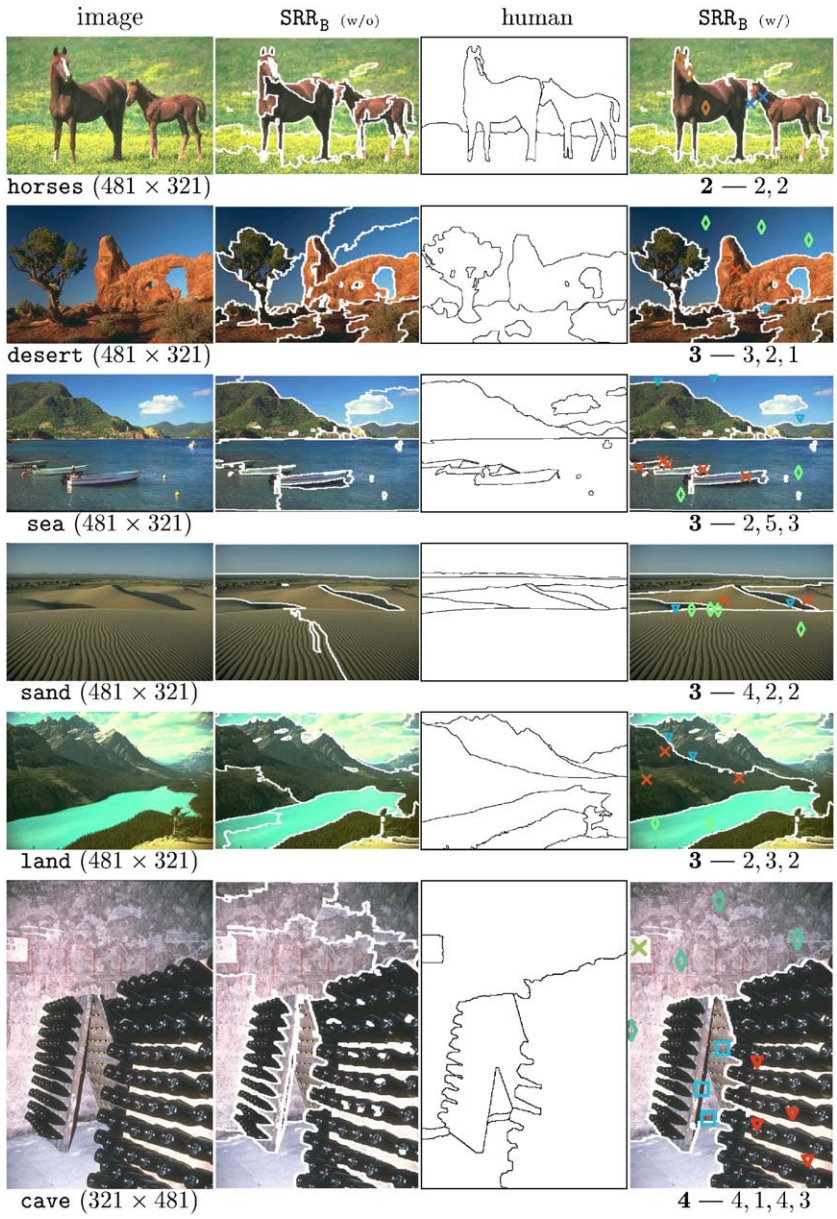

Fig. 7. Experiments on the BSDB. Segmentation's conventions follow Fig. 6. The numbers below the results of $\mathtt{SRR}_{\mathtt{B}}$ (w/) are m (bold), and the cardinal of the V_is.

gradients, such as the ceiling of the cave in the `bat` image of Fig. 4. Fortunately, the bias makes it possible to handle gradients in quite an efficient way. The strong gradient of the `bat` picture comes from the flash of the camera and the rocky irregular ceiling of the cave. The background contains very bright parts (upper left) up to very dark parts (lower right); in that case, only seven models have been necessary to segregate the bat.

All these result have to be appreciated in the light of the amount of bias imposed, and the execution times. $\mathtt{SRR}_{\mathtt{B}}$ has required no more than a total of 14 pixels pointed in each image. Furthermore, the execution times give a significant advantage to $\mathtt{SRR}_{\mathtt{B}}$: each image was segmented in about a second with $\mathtt{SRR}_{\mathtt{B}}$, while it took between 5 and 9 min with `NCuts`. The experiments were ran on a Pentium IV 2 GHz PC with 512 MB ram. Grouping with bias involves an interaction with the user to define the constraints, and a loop between the user and the machine for their optimization: in that case, a program running in no time to get the results is clearly an advantage.

Fig. 6 shows a first experiment on the Berkeley Segmentation Data set and Benchmark (BSDB, [12]). This data base

presents, on a large amount of images, the results of human segmentations. This is particularly relevant for biased grouping, which precisely involves the human in the loop. Again, the task of segregating the woman from her background is much more accurate for $\mathtt{SRR_B}$ when compared to NCuts: $\mathtt{SRR_B}$ fits very well the human segmentation, with only six pixels pointed in the image.

4.3. $\mathtt{SRR_B}$ vs. human on the BSDB images

We have performed more experiments on the BSDB, to compare $\mathtt{SRR_B}$ against human segmentations. The problems considered are more general than the segregation background/foreground, which has led us to leave somewhat NCuts for these experiments (whose best results were above average), and focus on the way the biased segmentation in $\mathtt{SRR_B}$ could come close to that of an human. The results are presented in Fig. 7. Apart from image cave, we have not necessarily tried to fit the human segmentations observed in the BSDB. Rather, we have tried to fit the way we would segment the images, and then we have chosen in the BSDB an human segmentation close to the result obtained. Notice that the images chosen are, on average, quite difficult. A typically difficult example is the sand image, in which very little can be obtained from the colors only. In this image, the NCuts without bias have obtained almost exactly the same result as $\mathtt{SRR_B}$ without bias. The biased segmentation, which makes extensive use of the gradient rule to choose the models (Section 4.1), has obtained a very good segmentation of the picture with few models, and the result closely follows the human segmentation. On image cave, we have tried to fit as exactly as possible an human segmentation of the BSDB (shown), using the fewest number of models. Without an extensive tuning (this took only few tries), we have almost exactly matched the human segmentation while using only 12 models.

5. Conclusion

In this paper, we have proposed a novel method for segmenting an image with a user-defined bias. The bias takes the form of pixels pointed by the user on the image, to define regions with distinctive sub-parts. The algorithm proposed rely on an unbiased segmentation algorithm [2,5], which we first modify and improve. The extension to biased segmentation keeps both the fast processing time and the theoretical properties of the former unbiased approaches onto which it is based. Experimental results show dramatic improvements of the biased segmentations over the unbiased results, and the biased results compare very favourably to previous approaches [1,7]. All these benefits come at a negligible additional computational cost when compared to unbiased grouping, while biased eigenvalue-based segmentation approaches suffer from slow-downs of magnitude orders [1]. From an experimental standpoint, the running time of our biased approach is also a magnitude order smaller than those of Yu and Shi [1,7], and it is compatible with the constraint that grouping with bias involves close interactions with the user. We also emphasize the slight bias we use, with which the quality improvements it brings make it a valuable companion well worth the try for the segmentation of complex images, such as those obtained with digital media.

Code availability: $\mathtt{SRR_B}$ can be obtained from the author's webpages.

References

[1] S.-X. Yu, J. Shi, Grouping with bias, in: Advances in Neural Information Processing Systems, vol. 14, 2001, pp. 1327–1334.
[2] R. Nock, Fast and reliable color region merging inspired by decision tree pruning, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, IEEE Computer Society Press, Silver Spring, MD, 2001, pp. 271–276.
[3] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 888–905.
[4] P.F. Felzenszwalb, D.P. Huttenlocher, Image segmentation using local variations, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, IEEE Computer Society Press, Silver Spring, MD, 1998, pp. 98–104.
[5] F. Nielsen, R. Nock, On region merging: the statistical soundness of fast sorting, with applications, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, IEEE Computer Society Press, Silver Spring, MD, 2003, pp. 19–27.
[6] C.T. Zahn, Graph-theoretic methods for detecting and describing gestalt clusters, IEEE Trans. Comput. 20 (1971).
[7] S.-X. Yu, J. Shi, Segmentation given partial grouping constraints, IEEE Trans. Pattern Anal. Mach. Intell. 26 (2004) 173–183.
[8] S. Geman, D. Geman, Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of Images, IEEE Trans. Pattern Anal. Mach. Intell. 6 (1984) 721–741.
[9] J. Luo, C. Guo, Perceptual grouping of segmented regions in color images, Pattern Recognition 36 (2003) 2781–2792.
[10] D. Comaniciu, P. Meer, Robust analysis of feature spaces: color image segmentation, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, IEEE Computer Society Press, Silver Spring, MD, 1997, pp. 750–755.
[11] S. DasGupta, Learning mixtures of gaussians, in: Proceedings of the 40th IEEE Symposium on the Foundations of Computer Science, 1999, pp. 634–644.
[12] UC Berkeley vision group, The Berkeley segmentation dataset and benchmark, 2004, http://www.cs.berkeley.edu/projects/vision/grouping/segbench/.
[13] C. Fiorio, J. Gustedt, Two linear time Union-Find strategies for image processing, Theoret. Comput. Sci. 154 (1996) 165–181.
[14] C. McDiarmid, Concentration, in: M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, B. Reed (Eds.), Probabilistic Methods for Algorithmic Discrete Mathematics, Springer, Berlin, 1998, pp. 1–54.