Grouping with Bias Revisited

Richard Nock[*]
DSI/Univ. Antilles-Guyane
Campus de Schoelcher, B.P. 7209
97275 Schoelcher, France.

Frank Nielsen[†]
Sony CS Labs, FRL
3-14-13 Higashi Gotanda
Shinagawa-Ku, Tokyo 141-0022, Japan.

Abstract

In this paper, we improve and tailor a recent statistical region merging approach to biased (partially supervised) grouping. The approach appears to be attractive both for its theoretical benefits and its experimental results, as light bias brings dramatic improvements over unbiased approaches on difficult pictures. Comparisons with another biased grouping algorithm display very favorable results.

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 low level descriptions (*e.g.* pixels' color levels) and their spatial relationships, while a user always uses higher level of knowledge to cluster the image objects.

With the advent of media making it easier and cheaper to collect and store digital images, this challenge has been raised even further towards both computational efficiency and robust processing. Consider for example image `flower` in Figure 1. At the highest level, users would consider that this image contains two distinct objects (the bee and the flower) plus the background. Due to the distribution of colors, it is virtually impossible for a non-biased segmentation technique to make a single region out of the flower, whose colors are contrasted, *while* preventing the bee to be merged with the background. The right image is the result of our algorithm with a slight grouping bias. Regions found are delimited by white borders. This result is presented in greater details in the experimental Section (Figure 4).

Common grouping algorithms for image segmentation use a weighted neighborhood graph to formulate the spatial relationships among pixels [3, 2, 4, 5, 1], and then formulate the segmentation as a graph partitioning problem. An essential difference between these algorithms is the locality of the grouping process. [5, 1] solve it from a global standpoint, whereas [3, 2, 4] make greedy local decisions to merge the connex components of induced subgraphs.

Figure 1: Image `flower` (left), and `SRM.B`'s result (right).

Our approach to segmentation, which gives the results of Figure 1, is based on a framework previously studied by [1]. 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 [6], Human-computer interaction, spatial attention and others [1]. Grouping with bias is basically solved by pointing in the image some pixels that the user feel belong to identical/different objects, and then solving the segmentation as a constrained grouping problem. The solution for the global approach of [1, 7] is mathematically appealing, but it is computationally demanding, it cannot easily handle all constraints generated by bias, it works on a single (composite) channel, and it requires a significant bias to guide reliably the segmentation.

A different grouping approach appears to be conceptually appealing for an adaptation to biased segmentation [2, 4]. It considers that the observed image is the result of the sampling of a theoretical image, in which regions are statistical ("true") regions characterized by distributions ; the segmentation aims at approximating the statistical regions on the sole basis of the observed image. There is no distribution assumption on the statistical pixels of this theoretical image, apart from an homogeneity property which relies on the region's expectations. The biased grouping

[*]URL: `http://www.martinique.univ-ag.fr/~rnock`
[†]URL: `http://www.csl.sony.co.jp/person/nielsen/`

problem turns out to allow statistical regions to contain different statistical sub-regions, not necessarily connex, each satisfying independently the homogeneity property, *and* for which the user feels they all belong to the same perceptual object. Thereby, this relaxes even further the constraints on the theoretical image.

Our contribution in this paper is twofold. First, it consists of two modifications and improvements to the unbiased segmentation algorithm of [2, 4]. Second, we extend this algorithm to grouping with bias. Our extension keeps both fast processing and theoretical bounds on the segmentation's quality. Experimental results display very favorable results compared to [5, 1] for both the unbiased and the biased settings.

Section 2 summarizes the unbiased approach of [2, 4], and presents our modifications to their algorithm in the unbiased setting. Section 3 presents extensions to biased grouping. Section 4 presents experimental results and comparisons with [5, 1], and Section 5 concludes.

2. Improving grouping as in [2, 4]

We first recall the basic facts of the model of [2, 4]. Throughout this paper, "log" is the base-2 logarithm. The notation $|.|$ 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 $\{1, 2, ..., g\}$.

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.

From this model, [2, 4] 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 channel a in region R of I, and let $\mathcal{R}_l$ be the set of regions with l pixels. Let $b(R) = g\sqrt{(1/(2Q|R|))\ln(|\mathcal{R}_{|R|}|/\delta)}$, with $|\mathcal{R}_l| \leq (l+1)^g$. The merging predicate is [2]:

$$\mathcal{P}(R,R') = \begin{cases} \texttt{true} & \text{iff} \quad \forall a \in \{\mathbf{R},\mathbf{G},\mathbf{B}\}, \\ & \quad |\overline{R'}_a - \overline{R}_a| \leq b(R) + b(R') \\ \texttt{false} & \text{otherwise} \end{cases} \quad (1)$$

The description of the algorithm `PSIS` of [2] is straightforward:

1. it makes a sorting over the set S_I of the pairs of adjacent pixels of the image, according to the increasing values of $f(p,p') = \max_{a\in\{\mathbf{R},\mathbf{G},\mathbf{B}\}} |p'_a - p_a|$, with p_a color channel a of pixel p;
2. afterwards, 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 $R(p'_i)$, with the merging predicate $\mathcal{P}(R(p_i), R(p'_i))$.

This approach to segmentation is interesting from the algorithmic standpoint, because it is simple and fast: its complexity is $\mathcal{O}(|I|)$, both in time and space [2, 4], which is significantly better than [5]. We now propose two modifications and improvements of the approach of [2, 4].

The first is a modification of step [1.] of the algorithm, with a more reliable f. [4] have shown that f should theoretically be an estimator, as reliable as possible, of the local between-pixel gradients. We have chosen to extend Sobel convolution kernels classically used in edge detection for pixel-wise gradient estimation, and use in 4-connexity those estimating $\hat{\partial}_x$ or $\hat{\partial}_y$ between neighbor pixels, for each color channel $a \in \{\mathbf{R}, \mathbf{G}, \mathbf{B}\}$. 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's border) do we keep the usual f of [2, 4].

Our second improvement concerns step [2.] and the merging predicate $\mathcal{P}$. It relies on the following theorem whose proof follows from [8].

Theorem 1 *Consider a fixed couple (R, R') of regions of I. $\forall 0 < \delta \leq 1$, $\forall a \in \{\boldsymbol{R}, \boldsymbol{G}, \boldsymbol{B}\}$ the probability is no more than δ that $|(\overline{R}_a - \overline{R}'_a) - \boldsymbol{E}(\overline{R}_a - \overline{R}'_a)| \geq g\sqrt{(1/(2Q))\,((1/|R|) + (1/|R'|))\ln(2/\delta)}$.*

Theorem 1 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 and taking into account that in 4-connexity, there are $\mathcal{O}(|I|)$ couples of adjacent pixels.

Theorem 2 *$\forall 0 < \delta \leq 1$, there is probability at least $1 - (3|S_I|\delta)$ that all couples (R, R') tested shall verify $\forall a \in \{\boldsymbol{R}, \boldsymbol{G}, \boldsymbol{B}\}$, $|(\overline{R}_a - \overline{R}'_a) - \boldsymbol{E}(\overline{R}_a - \overline{R}'_a)| \leq b(R,R')$, with $b(R,R') = g\sqrt{(1/(2Q))(1/|R| + 1/|R'|)\ln(2/\delta)}$.*

$b(R,R')$ would lead to a very good theoretical merging predicate $\mathcal{P}$ instead of using the threshold $b(R) + b(R')$ in eq. 1, provided we pick a δ small enough. This predicate would be much better than [2, 4] 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 $\tilde{\mathcal{O}}(b(R,R'))$. Remark that provided regions R and R' are not empty, $b(R,R') \leq \sqrt{b^2(R)+b^2(R')} < b(R)+b(R')$. This right quantity is that of eq. 1. The center quantity (which is indeed $\tilde{\mathcal{O}}(b(R,R'))$ provided a good upperbound on $|\mathcal{R}_l|$ is used) is the one we use for our merging predicate. It is thus:

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

Let us now concentrate on $|\mathcal{R}_l|$. [2, 4] pick $|\mathcal{R}_l| \leq (l+1)^g$, considering that a region is an unordered bag of pixels (each color level is given $0, 1, ..., l$ pixels). This bound counts numerous duplicates for each region: *e.g.* at least $(l+1)^{g-l}$ when $l < g$. Thus, we consider $|\mathcal{R}_l| \leq (l+1)^{\min\{l,g\}}$.

Because our merging predicate is tighter than [2, 4], the segmentation enjoys both the qualitative and the quantitative error limitation bounds proven in [2, 4]. In the sequel, we shall refer to our modified version of PSIS as $\texttt{SRM}_{.\texttt{B}}$ (SRM stands for Statistical Region Merging).

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 [1], we define a *grouping bias* to be user-defined disjoint subsets of V: $\{V_1, V_2, ..., V_m\} = \mathcal{V}$. 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 $\mathcal{V}$, and

(ii) $\forall 1 \leq i \leq m$, 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 $\mathcal{V}$, and condition (ii) states that each pointed pixel (coming from $\mathcal{V}$) 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 $\mathcal{V}$. The term model makes statistical sense because any $V_i \in \mathcal{V}$ with $|V_i| > 1$ may represent a single object for the user, but composed of different statistical sub-regions ("models") of I^*.

There are two types of regions in a biased grouping's result. The first are usual regions, without models, which we call "model-free". Each other region ("model-based") corresponds to some element of $\mathcal{V}$. A region corresponding to some V_i with $|V_i| = 1$ may be reduced to a single model (model = region). A region corresponding to some V_i with $|V_i| > 1$ is the union of $|V_i|$ different models, each of which representing a different sub-part of the whole object represented by the region. Notice that grouping with bias authorizes occluded regions (*i.e.* non-connex).

Our region merging algorithm keeps the following invariant: merging a model-free region and a model-based region results in the merging of the first region into one model (sub-region) of the second. The modification of the approach of [2, 4] consists in first making each V_i defined by the user, and then, through the traversing of S_I (step [1.] above), replacing the merging stage 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 $\mathcal{P}(R(p), R(p'))$, and decide whether to merge or not;

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 $\mathcal{V}$ (and we obviously do not have to merge them), or they are defined by vertices of the same subset of $\mathcal{V}$. 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 $\mathcal{P}(M(p), R(p'))$, 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 $\mathcal{P}(M(p), R(p'))$ 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 $\mathcal{V}$ come from different statistical regions of I^*. This is sound with the goal of bias, which is precisely to make it possible for the observer to integrate in the same perceptual object different statistical (true) regions of I^*.

4. Experiments

We have run $\texttt{SRM}_{.\texttt{B}}$ on a benchmark of pictures of various domains and hardness levels, to test its ability to improve

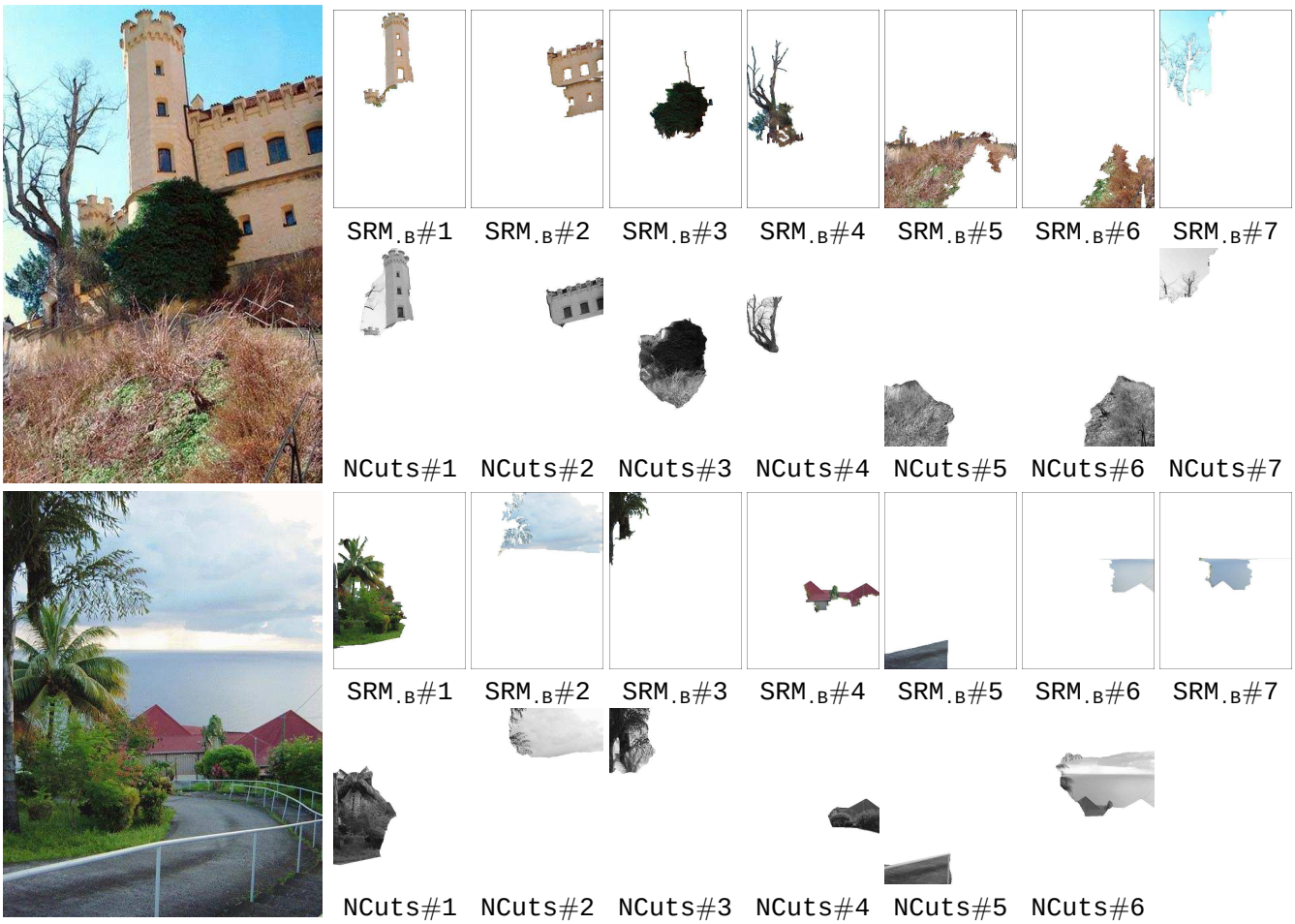

Figure 2: Results of SRM.B and NCuts without bias on images castle (top, 567×378) and saint-pierre (bottom, 640×480). For each image, the top row displays the largest regions found by SRM.B, and the bottom row those of NCuts.

the segmentation quality over the unbiased approach, while using a bias as light as possible. We have also compared SRM.B against normalized cuts (NCuts) approaches, with and without bias [5, 1].

While looking at the experiments of SRM.B, the reader may keep in mind that images are segmented as they are, *i.e.* without any preprocessing, and with the same parameters value following [2, 4]: $Q = 32, \delta = 1/(3|I|^2)$. Thus, the quality of the results may be attributed only to SRM.B, and not to any domain-specific tuning nor preprocessing optimization.

We split our results in three subsections. The first makes comparisons of SRM.B and NCuts without bias. The two others explore segmentation with bias: the first one compares both approaches for segmenting with a restricted focus of attention [1]. The last one explores grouping with bias without restricted focus of attention. In this last part, only results for SRM.B are provided, because it appears that the bias used can hardly be handled by NCuts.

4.1. SRM.B vs NCuts without bias

Figure 2 gives segmentation's results on two digital pictures. Both approaches manage to find regions that the user would consider as perceptually distinct: the castle's tower, the big bushy tree, the clouds, etc. . However, there are some notable differences between the algorithms. The regions found by SRM.B fit more precisely the object contours. Those of NCuts sometimes encompass more than a single object, even when these objects are clearly distinct, such as region #6 for NCuts on image saint-pierre. This advantage for SRM.B partially seemingly comes from the fact that its neighborhood graph is planar. The number of tunable parameters gives also advantage to SRM.B: NCuts integrate various parameters which we had to tune extensively to make its results' quality come close to SRM.B's (neighborhood radius, sample rate,

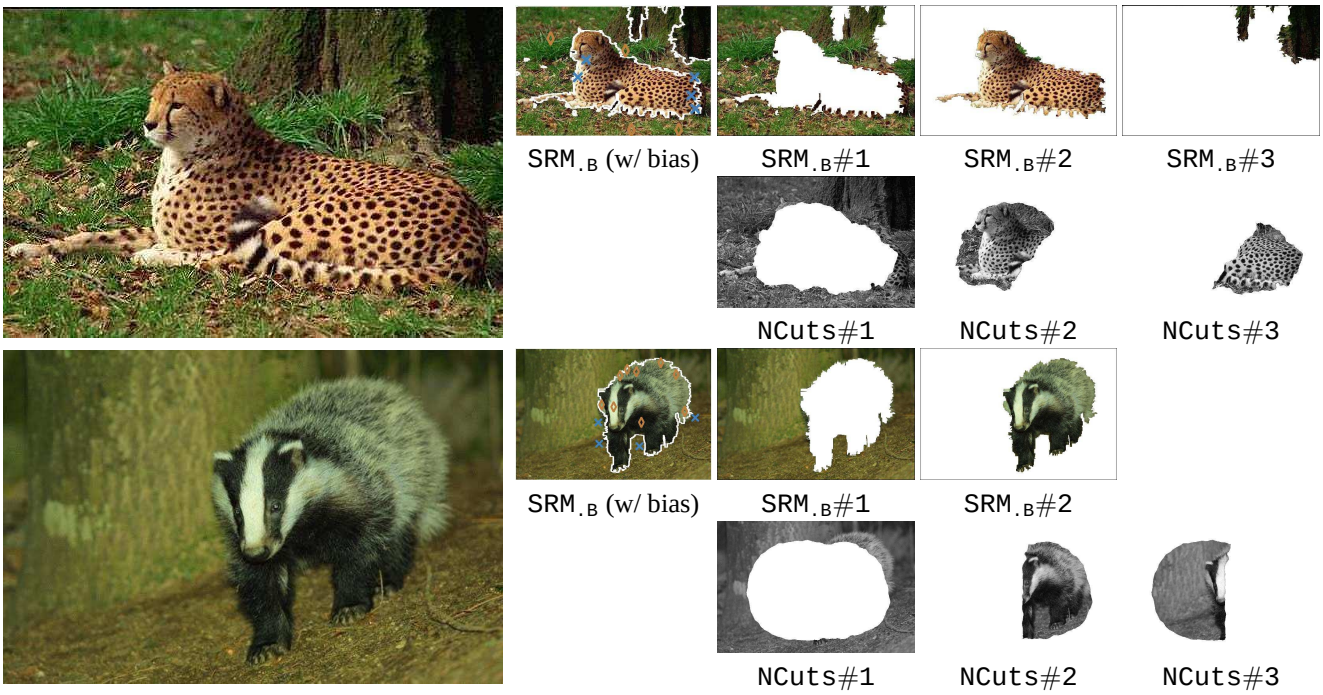

Figure 3: Results of SRM.B and NCuts without bias on images leopard (top, 331×500) and badger (bottom, 427×640). For each image, the partition of the image found by SRM.B is shown (top row, left image), with the user-defined bias. For image leopard, we have $m = 2, |V_1| = 4, |V_2| = 5$, and $m = 2, |V_1| = 10, |V_2| = 4$ for badger. The largest regions found follow the convention of Figure 2 (see text for details).

sigmas, number of regions, etc.). Finally, the results have also to be appreciated in the light of the execution time: both algorithms were ran on a Pentium IV 2 Ghz PC with 512 Mb ram. While SRM.B took less than one second on each image, NCuts took more than four *minutes* on the same images.

4.2. Bias as restricted focus of attention

We ran SRM.B and NCuts with a bias aimed at segregating an object from the background [1, 7]. The most convenient solution which led to the best results for NCuts was to define for the bias a frame of 10 pixels width, which is supposed to point for the background, and then try to separate this background from the object of attention. Executing SRM.B was by far simpler, since we only had to point exclusion constraints between two types of models: those for the background, and those for the object. Notice also the very light bias imposed for SRM.B, as no more than a total of 14 pixels are pointed for the models.

Image leopard is an example of what can be obtained for a particularly hard picture, due to the speckled animal's coat. Five models are pointed for the animal's coat to retrieve almost entirely the animal in a single region, out of the dozen regions obtained when segmenting the picture without bias. Both images show a limit of NCuts when it comes to segmenting images with deep localized contrasts inside objects: both animals are split into two regions each. Again, the execution time gives a significant advantage to SRM.B: both images were segmented in about a second with SRM.B, while NCuts took five minutes for leopard and nine minutes for badger. 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, an algorithm running in no time to get the results is clearly an advantage.

4.3. More grouping results for SRM.B

NCuts make it hard to define exclusion constraints (*i.e.* points which *must-not* belong to the same region), and it is even harder when constraints define more than two regions. On the other hand, SRM.B does not put any such restriction on the bias. Figures 4 and 5 show results on two more pictures (Figure 4 details the experiments of Figure 1). Figure 4 shows that a light bias for SRM.B brings a nice solution to the problem pointed out in the introduction, since we keep almost all the flower as a single region (region #1), while preventing the merge of the bee (region #3) with the background (region #2). On Figure 5, the segmentation of

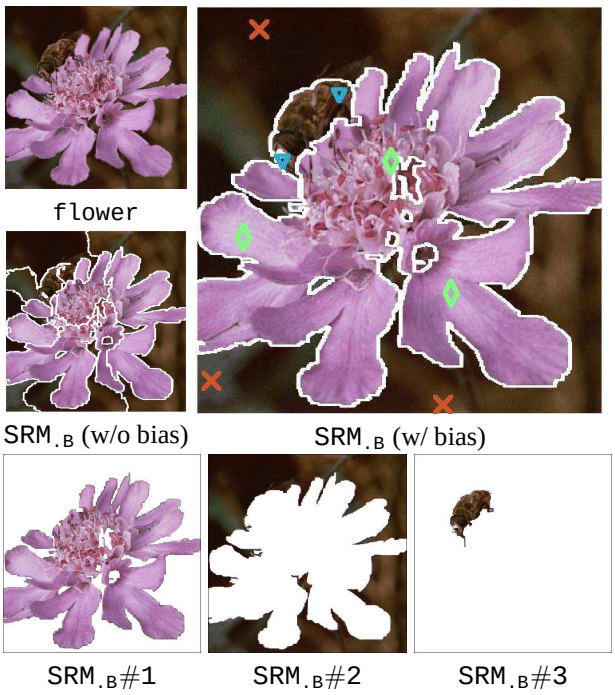

Figure 4: Image `flower` (512×512), and its segmentation by SRM.B, without bias, and with bias (upper right, *Cf* Figure 1). In the segmentations' results, regions found are delimited with white borders. Here, $m = 3$, $|V_1| = |V_2| = 3$ and $|V_3| = 2$. The bottom table shows the largest regions extracted from the segmentation.

SRM.B gives many regions due to the shadows on the facade. With a total of only twelve models for three regions, the biased segmentation yields accurate approximations to the conceptually distinct regions of the image: the first floor, the ground floor, the tower and the sky.

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. Our algorithm relies on an improvement of an unbiased segmentation algorithm [2, 4]. Unbiased and biased results compare all the more favorably with respect to normalized cuts approaches [5, 1] as we read the results in the light of the execution times.

Code availability SRM.B is freely available for Linux and Windows operating systems, from the authors webpages.

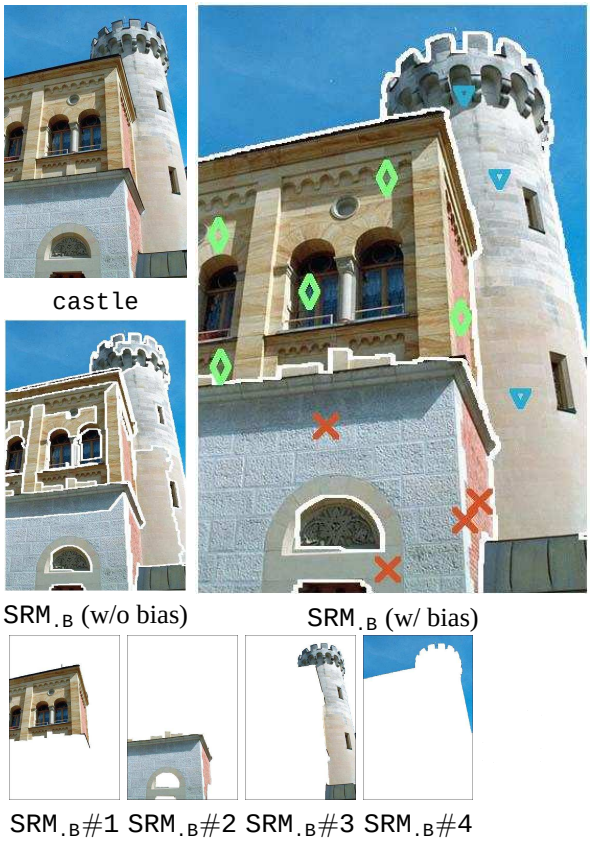

Figure 5: Results on `castle` (567× 378), with $m = 3$, $|V_1| = 5, |V_2| = 4, |V_3| = 3$. Conventions follow Figure 4.

References

[1] S.-X. Yu and J. Shi, "Grouping with Bias," in *NIPS*13*, 2001, pp. 1327–1334.

[2] R. Nock, "Fast and Reliable Color Region Merging inspired by Decision Tree Pruning," in *IEEE CVPR*. 2001, pp. 271–276, IEEE CS Press.

[3] P. F. Felzenszwalb and D. P. Huttenlocher, "Image segmentation using local variations," in *IEEE CVPR*. 1998, pp. 98–104, IEEE CS Press.

[4] F. Nielsen and R. Nock, "On region merging: The statistical soundness of fast sorting, with applications," in *IEEE CVPR*. 2003, pp. 19–27, IEEE CS Press.

[5] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE TPAMI*, vol. 22, pp. 888–905, 2000.

[6] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of Images," *IEEE TPAMI*, vol. 6, pp. 721–741, 1984.

[7] S.-X. Yu and J. Shi, "Segmentation given partial grouping constraints," *IEEE TPAMI*, pp. 173–183, 2004.

[8] C. McDiarmid, "Concentration," in *Probabilistic Methods for Algorithmic Discrete Mathematics*, M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, Eds. 1998, pp. 1–54, Springer Verlag.