

A Concentration-Based Adaptive Approach to Region Merging of Optimal Time and Space Complexities

Christophe Fiorio
LIRMM
161, rue Ada
34392 Montpellier, France
fiorio@lirmm.fr

Richard Nock
Université des Antilles-Guyane
UFR SEN-Dept of Maths and CS
97159 Pointe-A-Pitre Cedex, France
rnock@univ-ag.fr

Abstract

In this paper, we investigate image segmentation as a statistical and computational problem. The observed image is sampled from a theoretical, unknown image, in which pixels are represented by distributions. Our objective is to approximate as best as possible the region segmentation in the ideal image, where each region has pixels with identical expectations, but adjacent regions have different pixel's expectations. From that model, a concentration-based statistical test for deciding region merging is built, limiting the risk of wrong merges. The analysis is carried out without any assumption on the distributions: we avoid in particular the classics of variance analysis, normality and homocedasticity. A practical approximation of the test is given, of constant time and space computation, which leads in turn to a segmentation algorithm of optimal complexity, easy to implement. Some experiments on various types of images shed light on the quality of the segmentations obtained.

1 Introduction

In the early stage of a vision process it is necessary to isolate “objects” in the scene before recognising them, a step called *image segmentation*. In this paper, we investigate image segmentation by region merging, which consists in building up regions by combining smaller regions, pixels being taken as primary regions. Our ultimate goal is to obtain a good partition of an image I into regions $\{R_1, R_2, \dots, R_n\} = P(I)$, each region being 4-connected: between any of its pixels, a path exists, lying in the region, with horizontal or vertical connections.

In the field of image segmentation, many algorithms use statistical or mathematical techniques, models, relying on hypotheses made *a priori* on the image [7, 1, 6]. Many of these hypotheses constrain the models by introducing distributional assumptions on the image. Also, the computational costs can be huge. Other approaches focus rather on topologic and algorithmic properties of the image [4, 3], striving to find fast segmentation algorithms. However, statistical confidence in the results is generally not ensured. In that paper, we are interested in a bridging technique between the two approaches, based on a weakly constrained statistical model of image generation, which ultimately leads to a

reliable segmentation algorithm of optimal complexity.

Our model supports the claim according to which an observed image I is principally a snapshot of an ideal object “scene” I^* , taken under particular conditions. Formally, in I^* , pixels are represented by distributions, used to generate the observed image. Ideal regions satisfy the homogeneity criterion that pixel’s expectations are equal, expectations that are in turn different between adjacent regions. This model gives an intuitive notion of what is the *optimal segmentation*. This could be obtained by infinite sampling of the distributions, thus under infinitely many “conditions of observations”. Averaging over the observed images would produce for each region its theoretical grey-level expectation, and we would observe the exact frontiers between regions. Optimising segmentation to recognise regions of I^* may be of great practical interest, in particular when processing numerous images of similar scenes, such as for real-time segmentation. In that case, we better look for fast results of overall quality, rather than single image optimisation at the possible expense of time complexity.

In the following section, we present our model of image generation. Then, we propose a statistical merging test to recognise if two observed, adjacent regions, belong actually to the same object in I^* . It has the major computational feature of being constant time and space approximable. It has the major statistical feature of being completely adaptive, and relies on a concentration bound whose proof sketch is given. Finally, we provide a segmentation algorithm of optimal complexity implementing this test, along with experiments that were conducted.

2 A Model for Image Generation in I^*

Let I denote the image observed, containing $|I|$ pixels ($|\cdot|$ denotes the cardinality). I is described over a maximum, theoretical number, of g grey levels (generally, $g = 256$), but effectively contains g_I observed grey-levels. The model easily endorses RGB coding, by considering three separate models in the same way. I is an observation of a perfect scene I^* we do not know of, in which pixels are perfectly represented by a family of *distributions*, from which the observed grey-level is sampled. In all that follows, the “*” superscript denotes objects taken from I^* , such as R^* for a region of I^* , to which corresponds a region R of I . The j^{th} pixel in the i^{th} region R_i of the image I , $p_{i,j}$, is obtained by the outcome of some random variable (r.v.) $p_{i,j}^* \in I^*$, itself sum of Q independent r.v. $p_{i,j,1}^*, p_{i,j,2}^*, \dots, p_{i,j,Q}^*$ of I^* (each of these take values in the set $[0, g/Q]$, see Figure 1). Also, pixels are supposed sampled independently from each other. The role of Q is mainly to ensure the practical tractability of our merging test. Actually, as we shall see, the bigger the image I and the regions tested, the smaller Q can be chosen. As outlined in the introduction, each region in I^* satisfies an homogeneity property, formalised as follows:

- First, $\forall i \in \{1, 2, \dots, |I|\}, \forall j \in \{1, 2, \dots, |R_i^*|\}, \mathbf{E}(p_{i,j}^*) = \mathbf{E}(R_i^*)$ (expectations of all pixels in R_i^* are equal).
- Second, any couple of adjacent regions (R_i^*, R_j^*) satisfies $\mathbf{E}(R_i^*) \neq \mathbf{E}(R_j^*)$.

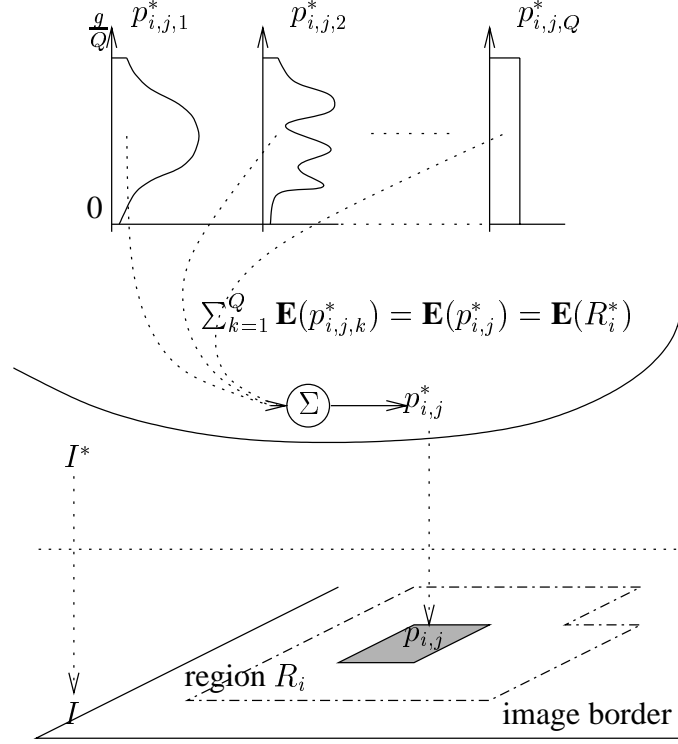


Figure 1: One pixel in I^* and I ($\mathbf{E}(\cdot)$ denotes expectation).

3 A statistical merging test in I

The test relies on some definitions related to ordinary variance analysis. In contrast however, our study does not rely on assumptions such as normality or homocedasticity, *i.e.* equal standard deviation between pixels in I^* . Though we do not observe it directly in I , we denote as $p_{i,j,k}$ the outcome of $p_{i,j,k}^*$, *i.e.* $p_{i,j} = \sum_{k=1}^Q p_{i,j,k}$. In all that follows, the bar notation " \bar{R} " denotes the *observed grey-level average of some region*, that is, $(\sum_{j,k} p_{i,j,k})/|R_i|$. Similarly, the average grey-level of the observed image I is denoted \bar{I} .

Definition 1 Let $\{R_1, R_2, \dots, R_n\} = P(I)$ be a partition of I .

- The *Intra-Pixel Variability (IPV)* in I , is

$$V_{\mathbf{p}}(I, P(I)) = \sum_{i,j,k} \left(p_{i,j,k} - \frac{p_{i,j}}{Q} \right)^2$$

- The *Intra-Region Variability (IRV)* in I , is

$$V_{\mathbf{i}}(I, P(I)) = \frac{1}{Q} \sum_{i,j} (p_{i,j} - \bar{R}_i)^2$$

- The Extra-Region Variability (ERV) in I , is

$$V_e(I, P(I)) = \frac{1}{Q} \sum_i |R_i| (\bar{R}_i - \bar{I})^2$$

- The Overall Image Variability of I is

$$V(I) = \sum_{i,j,k} \left(p_{i,j,k} - \frac{\bar{I}}{Q} \right)^2$$

Some of the formulae presented before might appear misleading with respect to classical notations of variance analysis, but we have given them in the most compact form. For example, in order to follow the usual conventions, the IRV should be stated

$$V_i(I, P(I)) = \sum_{i,j} Q \left(\frac{\sum_k p_{i,j,k}}{Q} - \frac{\sum_{j,k} p_{i,j,k}}{Q|R_i|} \right)^2$$

This strictly represents the same formula as in definition 1. Also, classical result of variance analysis is preserved since we have

$$V(I) = V_p(I, P(I)) + V_i(I, P(I)) + V_e(I, P(I)) \quad (1)$$

(proof straightforward). The membership to some region $R^* \subseteq I^*$ of two adjacent regions R and R' of I is checked by a measure of their “proximity”, the increase of IRV when merging them, $\Delta_i(R, R')$ (In RGB coding, it is the sum of three IRV increases in each of the R, G and B coordinates, and each pixel results from the outcomes of $3Q$ distributions). Though we do not have access to any $p_{i,j,k}$, the fact that pixels are themselves “regions” of fixed size Q makes that $V_p(I, P(I))$ is constant, and thus does not influence $\Delta_i(R, R')$. The following lemma shows that the computation of $\Delta_i(R, R')$ is indeed easy (proof straightforward).

Lemma 1 *The increase of IRV when merging R and R' satisfies*

$$\Delta_i(R, R') = \frac{|R| \times |R'|}{Q(|R| + |R'|)} (\bar{R} - \bar{R}')^2 \quad (2)$$

Therefore, as long as we merge regions in the image, $V_i(I, P(I))$ increases by *local* amounts, which can be computed in constant time, as well as for the parameter’s update (means and sizes). The deviations of $\Delta_i(R, R')$, when R and R' come from the same region in I^* , can be quantified as follows ($\#R$ is the number of distinct regions of size $|R|$, having grey levels included into those of R):

Theorem 1 *Let R, R' be two adjacent regions in I , belonging to the same region R^* in I^* . Fix*

$$m = \frac{6\sqrt{2}g^2 \sqrt{\max\{|R|, |R'|\} (|R|(|R| + 2|R'|)^2 + |R'|^3)}}{Q^3 |R| |R'| (|R| + |R'|)}$$

Suppose that $g, Q, |R|, |R'|$ are such that $m < 1/2$. Fix some $\tau > 0$. Then,

$$\Pr[\Delta_i(R, R') \geq \tau] < \frac{8\#R\#R'}{1 - 2m} \times e^{-\frac{2Q^2 (|R| + |R'|)^2 \tau}{g^2 |R| |R'|}} \quad (3)$$

Proof sketch: Though concentration bounds are not so hard to obtain provided unrealistic hypotheses are assumed, the necessity for us to stay as close as possible of a practical model of image generation led us to build a rather long proof for our bound. Moreover, ordinary concentration bounds, in that case, generally lead to too large bounds for practical use: in practice indeed, and we have observed it, this can lead to an over merged image with too few regions. That is why we have developed a bound tailor-made for our model, whose proof steps are the following ones. First, lemma 1 is used to shift from variability analyses to the deviation analysis of $|\overline{R} - \overline{R}'|$ beyond some τ' , for some *fixed* regions R and R' . This is of better easiness to upperbound. Second, we remark that

$$(\overline{R \cup R'} - \overline{R}^*)^2 = (\overline{R} - \overline{R}')^2 + Z \quad (4)$$

with R^* being the region of I^* to which R and R' belong, and Z defining the outcome of a r.v. . In other words,

$$\begin{aligned} \Pr[|\overline{R} - \overline{R}'| \geq \sqrt{\tau'}] &< \Pr[(\overline{R \cup R'} - \overline{R}^*)^2 \geq \tau'] / \Pr[Z \geq 0] \quad (5) \\ &= \Pr[\mathbf{B}] / \Pr[\mathbf{C}] \quad (6) \end{aligned}$$

Then, a lowerbound $\Pr[\mathbf{C}] \geq \frac{1}{2} - m$ is obtained, where m is the median of Z (Z has 0 expectation). A concentration upperbound for the median m is obtained by combining the use of the Independent Bounded Difference Inequality (IBDI, [5]) and a derived version of Talagrand's theorem [5]. Finally, using the IBDI on $\Pr[\mathbf{B}]$ gives the final result of theorem 1. Since the result holds for two *fixed* regions R and R' , the multiplication of the bound by the number of possible choices for R and R' gives the final, desired result. \square

Theorem 1 can be used to build a merging threshold τ for $\Delta_i(R, R')$, under which we can expect R and R' to come from the same region in I^* . This is done by solving the right inequality of theorem 1 for some particular probability. The merging test for two adjacent regions R, R' is then straightforward: "if $\Delta_i(R, R')$ does not exceed τ , then merge R and R' ".

4 Practical approximations and experimental results

Theorem 1 gives a way to practically quantify the risk we make of rejecting the merging of two regions R and R' that would actually be elements of the same region R^* in I^* . Of course, when executing n' rejections $(R_1, R'_1), (R_2, R'_2), \dots, (R_{n'}, R'_{n'})$, the probability that the event of theorem 1 occurs can be upperbounded by n' times the upperbound given. Solving for τ the equality between the right member of Ineq. 3 and δ/n' gives the theoretical testing threshold, which is conveniently approximated for practical computability as follows. First, the segmentation algorithm we use is optimal but precludes the knowledge of n' in advance. We replace therefore n' by an upperbound for our segmentation algorithm, $2|I|$. n' appearing in a log factor for τ , this gives no visual difference in the results. Let

$$\#R = g_I^{|R|} / U_R \quad (7)$$

where U_R is a correction factor:

- $U_R = 0.5 * \frac{|R|!}{\lceil |R|/2 \rceil!}$ if $|R| < g_I$,
- $U_R = 0.5 * \frac{|R|!}{\lceil |R| - (g_I/2) \rceil!}$ otherwise (proof omitted due to the lack of space)

Factorials are approximated using Stirling's formula: $x! \approx \sqrt{2\pi x}(x/e)^x$. All tests are processed with reasonable values of Q , m , δ :

$$Q = \sqrt{g} \quad (8)$$

$$m = 1/4 \quad (9)$$

$$\delta = 1/|I| \quad (10)$$

While Q remains sufficiently small, this limits the risk of overmerging for significant, "visible" regions. The threshold τ , computable in constant time, is finally:

$$\tau(R, R') = \frac{g|R||R'|}{2(|R| + |R'|)^2} \log \left(\frac{32|I|^2 g_I^{(|R|+|R'|)}}{U_R U_{R'}} \right) \quad (11)$$

Finally, each image is preprocessed in linear time and space to merge adjacent pixels having the same grey-level: this does not degrade optimal solutions, but increases the average region's size, and thus the efficiency of the merging test. The segmentation algorithm consists in making a single pass of the image, and testing (if possible) the merging of the current pixel (or region) with the upside pixel's region, and then the left one. As showed in [3], a particular data structure allows to obtain optimal (linear) time and space complexities. The running time for the C implementation over a Pentium II PC never exceeded the second for every experiment.

Figures 2, 3 and 4 present some experiments on various images. We emphasise that *all* segmentations were obtained using the *same* tuning for all parameters, even if the images are much different from each other (synthetic, medical, aerial images, portraits, etc.). In particular, the quality of the results does not stem from the optimisation of the algorithm on each image. Also, images are segmented without any preprocessing (e.g. noise filtering). Only *street* was Gaussian smoothed according to [2]'s experimental setup, to make accurate comparisons.

In *cornouaille*, the algorithm manages a great reduction in the boat's noise, particularly on its left part, whereas it keeps almost intact the overall shape and the name of the boat (note also the segmentation of the man in the bottom-right part). In image *synth*, which mainly features a ramp gradient on the left, and a noisy rectangle area on the right, the algorithm detects exactly three regions, approximating those most people would consider as perceptually distinct. The *street* image shows that our algorithm outperforms two other approaches on two aspects: the number of regions selected compared to [2] (e.g. the van), and the accuracy of the segmentation compared to [3] (e.g. the grass, highly noisy).

References

- [1] S. Dugelay, C. Graffigne, and J.-M. Augustin. Segmentation of multibeam acoustic imagery in the exploration of the deep-sea bottom. In *Int. Conf. on Pattern Recognition*, pages 437–446, 1996.

- [2] P. F. Felzenszwalb and D. P. Huttenlocher. Image segmentation using local variations. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 98–104, 1998.
- [3] Christophe Fiorio and Jens Gustedt. Two linear time Union-Find strategies for image processing. *Theoretical Computer Science*, 154:165–181, 1996.
- [4] W. G. Kropatsch and S. B. Yacoub. A revision of pyramid segmentation. In *Int. Conf. on Pattern Recognition*, pages 477–481, 1996.
- [5] C. McDiarmid. Concentration. In M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, pages 1–54. Springer Verlag, 1998.
- [6] R. Watzel, K. Braun, A. Hess, W. Zuschratter, and H. Scheich. Restoration of dendrites and spines with the objective of Topologically Correct Segmentation. In *Int. Conf. on Pattern Recognition*, pages 472–476, 1996.
- [7] S. C. Zhu and A. Yuille. Region competition : Unifying snakes, region growing, and bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 884–900, 1996.

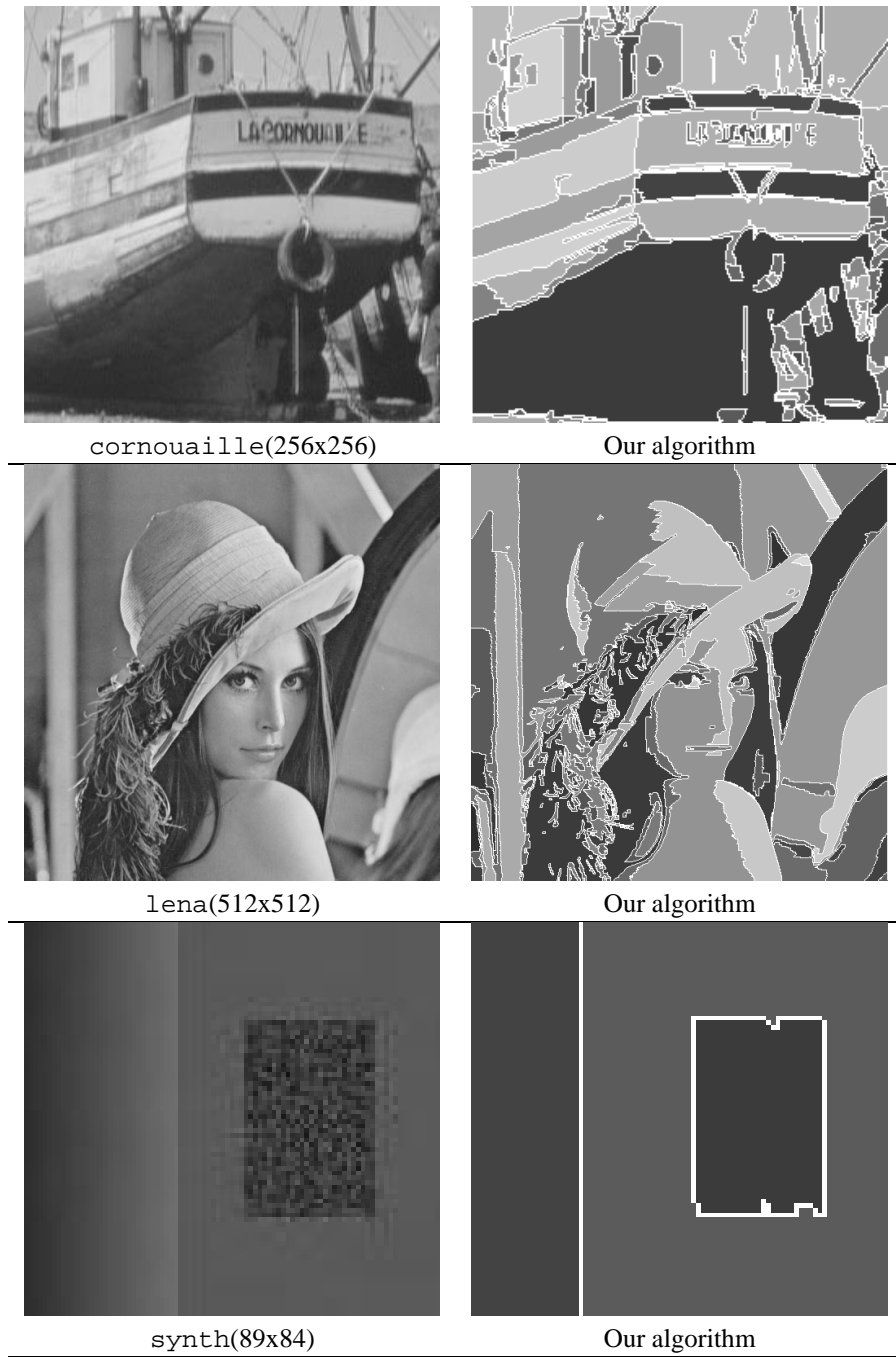


Figure 2: Some experiments conducted (regions are white bordered, and averaged inside).

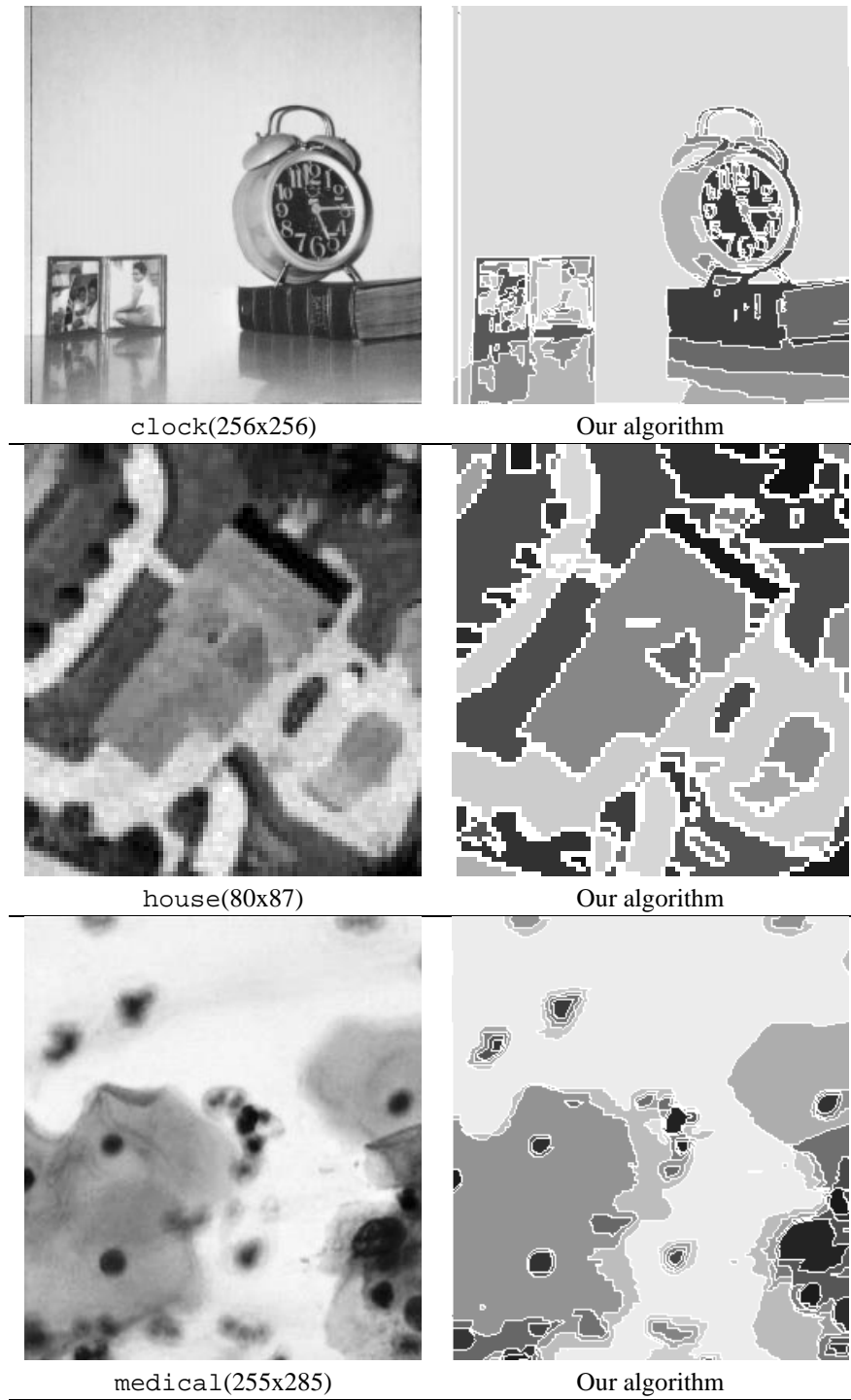


Figure 3: More experiments (regions are white bordered, and averaged inside).

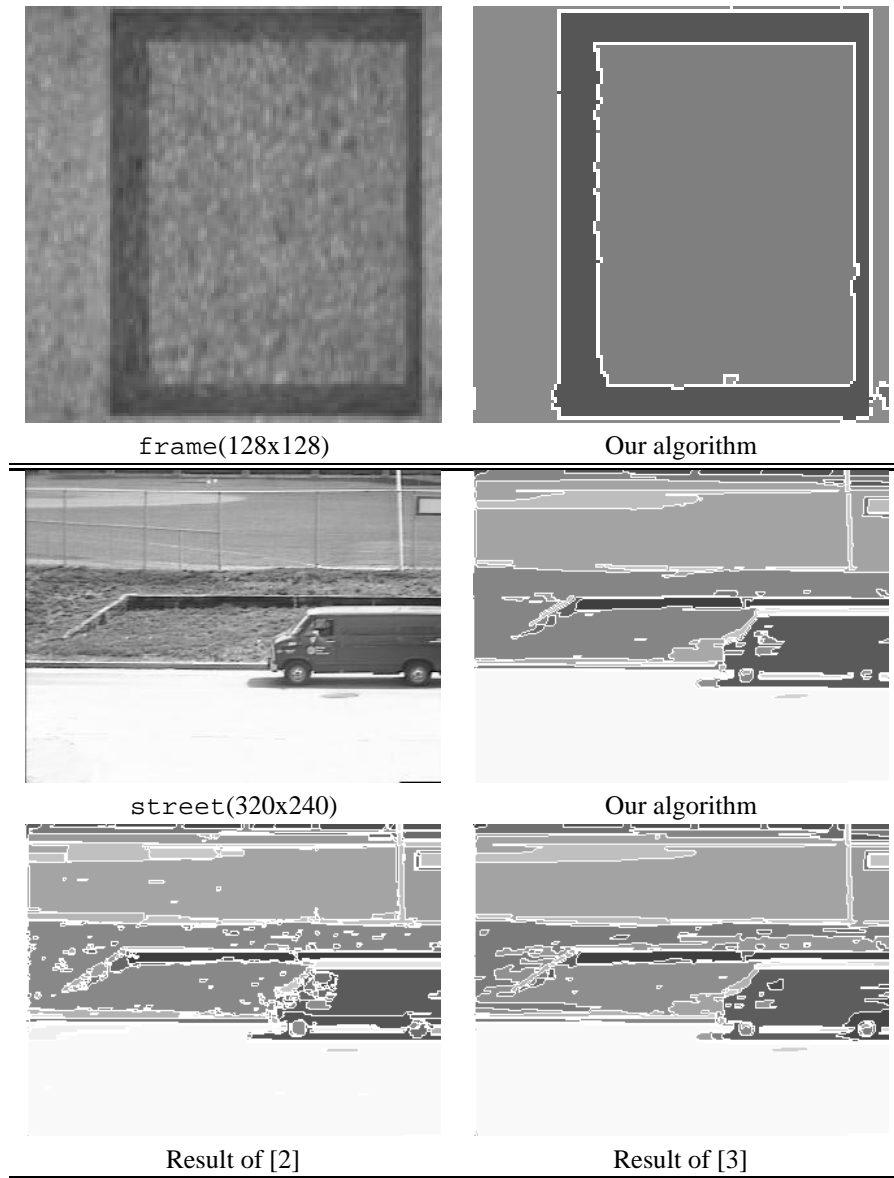


Figure 4: More experiments (regions are white bordered, and averaged inside).