

# Spatial Sampling for Image Segmentation

Mariano Rivera<sup>1</sup>, Oscar Dalmau<sup>1</sup> and Washington Mio<sup>2</sup>

<sup>1</sup> Centro de Investigacion en Matematicas A.C, Guanajuato GTO 36000, Mexico

<sup>2</sup> Florida State University, Tallahassee FL 32306, USA

**Abstract.** We present a framework for image segmentation based on the ML estimator. A common hypothesis for explaining the differences among image regions is that they are generated by sampling different Likelihood Functions. We adopt last hypothesis and, additionally, we assume that such samples are *i.i.d.* Thus, the probability of a model generates the observed pixel value is estimated by computing the likelihood of the sample composed with the surrounding pixels.

## 1 Introduction

Image segmentation consists in partitioning an image in regions with similar characteristics: color, texture, local orientation, *etc.* Although there are many strategies for segmenting images, we can classify such algorithms in *hard* and *soft* segmentation approaches. Hard segmentation methods try to directly estimate the label map, meanwhile soft approaches compute a membership map. Given that, in a Bayesian framework, the uncertainties can be represented by probabilities, then soft segmentation procedures are commonly named Probabilistic Segmentation (PS) approaches.

We use the following notation: we assume that the observed image  $g : \Omega \rightarrow \mathbb{R}^n$  ( $n = 1$  for gray scale images and  $n = 3$  for color images) is generated by sampling unknown probability density functions named models  $\mathcal{M} = \{M_k\}_{k=1}^K$  with parameters  $\theta_1, \theta_2, \dots, \theta_K$ ;  $\Omega$  denotes the set of all the pixels in a regular lattice and  $\mathcal{K} = \{1, 2, \dots, K\}$  the label set such that the label field  $c : \Omega \rightarrow \mathcal{K}$  indicates the source for each pixel. Then the task is to solve the inverse problem: to segment the image  $g$  (to estimate  $c$ ) into  $K$  classes. This task may require one of estimating possible unknown parameters  $\theta$ . We denote by

$$v_k(r) = P(g(r) | \theta_k, c(r) = k). \quad (1)$$

the likelihood of observing  $g(r)$  given the model  $k$ . This can be seen as the preference of the data,  $g(r)$ , for the model  $k$ . Thus the Maximum Likelihood (ML) estimator (classification) is given by  $c^{ML}(r) = \operatorname{argmax}_k v_k(r)$ . This Winner Takes All (WTA) assignment is used as estimate of the true label,  $c(r)$ . A disadvantage of the ML estimator is its sensibility to noisy data that results in noisy segmentations. For improving those segmentations of a noisy image one has two choices:

1. To increase the number of samples per pixel, *i.e.*, to acquire a set of independent observations  $\{g_i(r)\}_{i=1}^I$ . The samples' noise contributions are averaged and therefore improving the ML estimator; where the likelihood is given by

$$v_k(r) = \prod_i P(g_i(r) | \theta_k, c(r) = k). \quad (2)$$

2. To use prior knowledge that promotes smooth solutions. If this prior is coded as the probability  $P(c)$ , then, by using the Bayes' rule, the MAP estimator can be computed from the posterior probability:

$$c^{MAP} = \underset{c}{\operatorname{argmax}} \prod_r \prod_k v_k(r) P(c). \quad (3)$$

Both strategies allows one to estimate simultaneously the segmentation and the model parameter. This joint estimation can be implemented by an EM strategy [2]. Furthermore, note that both strategies are not mutual exclusive: the likelihood in (3) can be improved by the use of multiple samples. However from last two options, Bayesian regularization is the preferred strategy given that, in general, we are limited to a single image [3,7]. In such a case, one can find a vast literature for solving the optimization problem stated in (3). Such techniques can be classified as combinatorial optimization approaches (the ones that try to directly estimate the  $c^{MAP}$ ) [3,1,5,9,6] and probabilistic approaches (the ones for estimating a hidden real variable that represents the probability that  $c(r)$  takes a particular label, a PS) [8,11,4].

## 2 Spatial Sampling

We present a method for improving the likelihood in the lack of multiple observations. Our improved likelihoods can be used for directly compute the segmentation by means of the ML estimator or used as prime matter for a Bayesian segmentation method that solve (3). The general idea is simple, we assume that all the pixels are *i.i.d.* samples of generative models and the source is determined by the label map  $c$ . Since the image regions are relative large (this assumption is frequently codified as a prior in Bayesian regularization), then the pixels in a small neighborhood are very likely samples of a unique model. Thus the small pixel neighborhood can be assumed as multiple observations of the central pixel. This simple idea is actually the underlying idea of all spatial filtering techniques in image processing. Inspired on that, we purpose a novel and efficient framework for image probabilistic segmentation. Our approach, differently from those frameworks that regularize (smooths) the pixel values, regularizes the likelihood density functions.

Let  $\mathcal{N}_r$  be a neighborhood of pixels centered at  $r$  and  $\mathcal{G}_r = \{g(s) : s \in \mathcal{N}_r\}$  their corresponding pixels values. For the moment, we assume the simple neighborhood  $\mathcal{N}_r = \{s : \|r - s\|^2 \leq \rho\}$ , —in section 3 we discuss the neighborhood selection. We can note that  $\mathcal{G}_r$  is a sample with mixed likelihood:

$$P(\mathcal{G}_r | \theta, c, \pi) = \sum_k \pi_k(r) \left[ \prod_{s \in \mathcal{N}_r} P(g(s) | \theta_k, c(s) = k) \right] = (\pi^T \tilde{v})(r); \quad (4)$$

where we define the spatial likelihood

$$\tilde{v}_k(r) \stackrel{def}{=} \prod_{s \in \mathcal{N}_r} v_k(s) \quad (5)$$

and  $\pi(r) \in \mathbb{S}^K$  is a vector whose components are unknown mixture coefficient. Where we denote by  $\mathbb{S}^K$  the simplex with all the positive vector that sum one:  $z \in \mathbb{S}^K \iff$

$z \geq 0, \sum_k z_k = 1$  for  $k = 1, 2, \dots, K$ . Thus  $\pi_k(r)$  is the fraction of the sample  $\mathcal{G}_r$  generated with the  $k$ th model.

Then, the image segmentation can be estimated from an estimator of  $\pi$  if an appropriated  $\mathcal{N}_r$  is selected, subsection 3. Next we investigate two estimators of  $\pi$ :

1. A hard segmentation can be computed by the maximization of (4), a Linear Programming problem. It is easy to prove that the ML estimator is the indicator vector:  $p^{(1)}(r) = e_{k^*}$ , where  $e_k$  is the  $k$ th basis vector and  $k^* = \operatorname{argmax}_k \tilde{v}_k(r) = \operatorname{argmax}_k \hat{v}_k(r)$ ; where the normalized spatial likelihood is given by

$$\hat{v}_k(r) = \tilde{v}_k(r) / \sum_i \tilde{v}_i(r). \quad (6)$$

2. A soft estimation of  $\pi$  can be computed by the maximization:  $p^{(2)}(r) = \operatorname{argmax}_\pi (\pi^T \hat{v})(r) / \|\pi(r)\| \|\hat{v}(r)\|$ ; that results in  $p^{(2)}(r) = \hat{v}(r)$ .

Therefore, in any of the last two cases, the estimation of  $\pi$  is reduced to the computation of  $\hat{v}_k(r)$ : spatial products of individual likelihoods, or sums of log-likelihoods:

$$\hat{v}_k(r) \propto \prod_{s \in \mathcal{N}_r} v_k(s) = \exp \left( \sum_{s \in \mathcal{N}_r} \log v_k(s) \right). \quad (7)$$

Now we extend the above introduced probabilistic segmentation to the case of multiple sources, *i.e.*, to combine independent segmentation from different clues. After that, we will be in the capability of presenting the neighborhood choices and their algorithmic implications. First, we note that (7) can be written as

$$\hat{v}_k(r) \propto v_k(r) \bar{v}_k(r) \quad \text{with} \quad \bar{v}_k(r) = \prod_{s \in \mathcal{N}_r \setminus \{r\}} v_k(s). \quad (8)$$

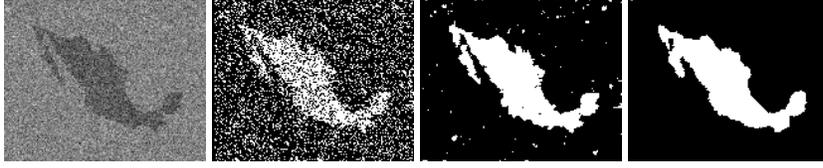
Eq. (8) can be understood as the combination of two independent sources: the likelihood estimated from the observed value,  $v_k(r)$ , and the likelihood,  $\bar{v}_k(r)$ , estimated with the neighbor pixels except  $r$ . We can give a further step by generalizing (8) to  $J$  independent sources and introducing their confidence factor. Then  $\alpha_j$  is our grade of confidence in the  $j$ th-source (the  $v^{(j)}$  likelihood) and it holds  $\alpha \in \mathbb{S}^J$ . So that  $\alpha_j = 1$  means that the  $j$ th-source has the largest possible confidence and it becomes irrelevant as  $\alpha_j \rightarrow 0$ . Thus:

$$\hat{v}_k(r) \propto \prod_{j=1}^J \left[ v_k^{(j)}(r) \right]^{\alpha_j} = \exp \left( \sum_{s=1}^J \alpha_j \log v_k^{(j)}(s) \right); \quad (9)$$

Eq. (9) is a simple form of combining a set of likelihoods (probabilistic segmentations). The different PS (sources) can result from the use of different clues as, for instance, color and local statistical descriptors for texture.

### 3 On the Neighborhood Selection

An accurate segmentation depends on selecting a pixel neighborhood such that its majority belongs to the correct class. Note that the right side of (9) defines a spatial filtering in the log-space of the likelihoods. Our derivation for distinct filters, both linear and



**Fig. 1.** ML segmentation with a squared neighborhood ( $\mathcal{N}_r = \{s : \|r - s\|_\infty \leq \rho\}$ ) for different values of  $\rho$ . The data corresponds to a binary map (left) corrupted with Gaussian noise [ $N(m = 0, \sigma = 0.7)$ ] and the segmentations to  $\rho = 0, 2$  and  $4$ .

nonlinear, constructs on the assumption that each neighbor pixel  $s \in \mathcal{N}_r$  is an independent source for estimating the likelihood at the pixel of interest,  $r$ .

**Homogeneous Windows (HW).** The simplest neighborhood is a regular window centered at the pixel  $r$ :  $\mathcal{N}_r = \{s : \|r - s\|_m \leq \rho\}$ , where the parameter  $\rho$  controls the sample size and  $\|\cdot\|_m$  is a given metric. For instance, the  $L_\infty$  norm leads to a square-shaped neighborhood (7) is reduced to a box-filter in the logarithmic space). The segmentations based on the ML estimator for different values of  $\rho$ . are shown in Fig. 1. Large  $\rho$ -values reduce the granularity but at the same time over-smoothed small details .

**General Homogeneous Windows (GHW).** The use of the previous HW is equivalent to apply a box-shape linear filter in the log-domain. This result can be generalized to any arbitrary linear filter if each neighbor pixel is considered an independent source [10,12,7]. . Then, similar to (9), the sources (neighbor pixels) are combined with a confidence factor that depends on its spatial distance to the central pixel  $r$ :

$$\hat{v}_k(r) \propto \prod_{s \in \mathcal{N}_r} [v_k(s)]^{\alpha(s)} = \exp \left( \sum_{s \in \mathcal{N}_r} \alpha(s) \log v_k(s) \right). \quad (10)$$

The Gaussian filtering results of choosing  $\alpha(s) \propto \exp \left( -(r - s)^T \Sigma^{-1} (r - s) \right)$ , where  $\Sigma$  is a covariance matrix and the simple homogeneous membrane is implemented as

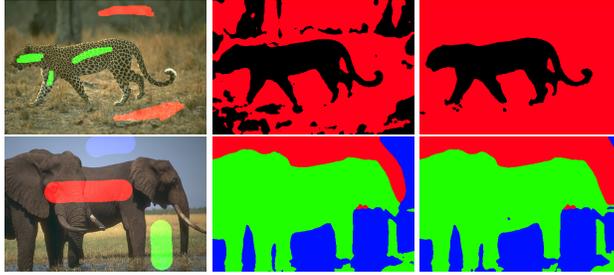
$$d^* = \underset{d: \Omega \rightarrow \mathbb{R}^K}{\operatorname{argmin}} \sum_{r: C(r)=0} \left\{ \|d(r) - \log v(r)\|^2 + \frac{\lambda}{2} \sum_{s \in \mathcal{N}_r} \|d(r) - d(s)\|^2 \right\}. \quad (11)$$

**Spatially Adapting Windows (SAW).** The shape neighborhood can be adapted by depending on the local properties of the image.

## 4 Experiments

We developed an interactive procedure for multiclass image segmentation based on ML segmentation. The models are empirically initialized from user marked data (scribbles) on the image. The purpose is to demonstrate that the final segmentation is improved by combining multiples sources (likelihood vectors) and the source combination is naturally implemented in our proposal.

In interactive image segmentation the user's scribbles define the multimap  $C : \Omega \rightarrow \{0\} \cup \mathcal{K}$  such that  $C(r) \in \mathcal{K}$  indicates that the pixel  $r$  is labelled as member of class



**Fig. 2.** Original image (first column). Segmentation using: just color distribution (second column) and with multiple clues (last column).

$k$  and  $C(r) = 0$  if such a pixel is unlabeled and hence its label needs be estimated. The segmentation procedure is as follows. Let  $g = \{g_i\}_{i=1}^3$  be the original image (in RGB space), then we computed for each pixel and color layer the local structure tensor  $\Sigma_g = \{\Sigma_1, \Sigma_2, \Sigma_3\}$ ; where

$$\Sigma_i(r) = \begin{bmatrix} g_{i11}(r) & g_{i12}(r) \\ g_{i12}(r) & g_{i22}(r) \end{bmatrix} \quad \text{for } i = 1, 2, 3. \quad (12)$$

is a symmetric semi-positive definite matrix: Then we group the original data and tensor's coefficients in four feature sets that will be consider four independent segmentation sources:  $g = \{g_i\}$ ,  $g_{11} = \{g_{i11}\}$ ,  $g_{12} = \{g_{i12}\}$ , and  $g_{22} = \{g_{i22}\}$ ; for  $i = 1, 2, 3$ . The additional feature sets [  $g_{11}$ ,  $g_{12}$  and  $g_{22}$  ] codify the local texture information. Afterward, we compute four Spatial Likelihoods  $\{v^{(j)}\}_{j=1}^4$ . The Likelihood Functions are estimated by histograms with  $64 \times 64 \times 64$  bins and the dynamic range of each feature image is linearly mapped into the interval [1, 64].

The confidence factor of a Likelihood Function set (says the  $j$ th set) is its capability for predict the correct pixel class. In our interactive scheme, such a confidence  $\alpha_j$  is large if the likelihoods of the hand-labeled pixels are large for their respective models (and small for the other ones). In particular the confidence of the  $j$ th source for of the  $k$ th class is  $\alpha_{jk} = \sum_r v_k^{(j)}(r) \delta(k - C(r)) / \sum_r \delta(k - C(r))$ . Figure 2 shows images segmented with the proposed procedure. The confidence scores are shown in Table 1. This Table shows in bold font the more confident source for each image. Note that if the color source is the one with largest confidence, then the color based segmentation is qualitatively as good as the one with four sources, these are the case of the *Elephant* image. However, the accuracy of the single source (color) segmentation is reduced as the confidence on such a source is reduced. Indeed, in all our experiments the best segmentation was computed with the proposed integration of all the sources. The filter used was the simple homogeneous membrane (11) with  $\lambda = 20$ .

## 5 Conclusions

The presented probabilistic segmentation strategy computes the uncertainty (probability) associated to each particular label. We start our development by noting that the

**Table 1.** Confidence factor to combine the probabilistic segmentations, see Fig. 2

Image	$g$	$g_{11}$	$g_{12}$	$g_{22}$
Cheetah	0.2284	<b>0.2714</b>	0.2411	0.2592
Elefant	<b>0.3280</b>	0.2270	0.2156	0.2294

spatial sampling is an alternative to the lack of multiple pixels' observations. Differently from multiple observations, the pixel neighborhood is a mixed sample and the estimated mixture coefficients can be used as the probability for the labels. We have noted that the neighborhood selection is an important issue to obtain a good segmentation. Our strategy allows us to combine multiple sources (probabilistic segmentations) in a natural way and is general enough to be applied in the development of algorithms for different computer vision applications.

**Acknowledgement.** This work was partially supported by Conacyt, Mexico: grant 61367, sabbatical stance for M. Rivera and scholarship for O. Dalmau.

## References

1. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Machine Intell.* 23(11), 1222–1239 (2001) 2
2. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B* 39, 1–38 (1977) 2
3. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE PAMI* 6(6), 721–741 (1984) 2
4. Kohli, P., Torr, P.H.S.: Dynamic graph cuts for efficient inference in Markov random fields. *IEEE Trans. Pattern Anal. Mach. Intell.* 29(12), 2079–2088 (2007) 2
5. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* 26(2), 147–159 (2004) 2
6. Komodakis, N., Tziritas, G., Paragios, N.: Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal–dual strategies. *Computer Vision and Image Understanding* 112, 14–29 (2008) 2
7. Li, S.Z.: *Markov Random Field Modeling in Image Analysis*. Springer-Verlag, Tokyo (2001) 2, 4
8. Marroquin, J.L., Velazco, F., Rivera, M., Nakamura, M.: Probabilistic solution of ill-posed problems in computational vision. *IEEE Trans. Pattern Anal. Machine Intell.* 23, 337–348 (2001) 2
9. Olsson, C., Eriksson, A.P., Kahl, F.: Improved spectral relaxation methods for binary quadratic optimization problems. *Computer Vision and Image Understanding* 112, 30–38 (2008) 2
10. Perona, P., Malik, J.: Scale-space and edge-detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 12(7), 629–639 (1990) 4
11. Rivera, M., Ocegueda, O., Marroquin, J.L.: Entropy-controlled quadratic Markov measure field models for efficient image segmentation. *IEEE Trans. Image Processing* 8(12), 3047–3057 (Dec 2007) 2
12. Terzopoulos, D.: Regularization of inverse visual problems involving discontinuities. *IEEE Trans. Pattern Anal. Mach. Intell.* 8(4), 413–424 (1986) 4