

Beta-measure for Probabilistic Segmentation

Oscar Dalmau¹ and Mariano Rivera¹

Centro de Investigación en Matemáticas, A.C. Jalisco S/N, Colonia Valenciana. C.P. 36240,
Guanajuato, Gto. México. {dalmau, mrivera}@cimat.mx.

Abstract. We propose a new model for probabilistic image segmentation with spatial coherence through a Markov Random Field prior. Our model is based on a generalized information measure between discrete probability distribution (β -Measure). This model generalizes the quadratic Markov measure field models (QMMF). In our proposal, the entropy control is achieved through the likelihood energy. This entropy control mechanism makes appropriate our method for being used in tasks that require of the simultaneous estimation of the segmentation and the model parameters.

1 Introduction

Image segmentation consists in partitioning the image into non-overlapping meaningful homogenous regions *i.e.* flat regions, movement (stereo, optical flow), model-based, texture, color, ... etc. Image segmentation has been widely used in different applications, for example, in medical image [1,2,3], robot vision [4,5], image colorization [6] and image editing [7]. Several techniques have been used to solve this problem. Among them one can find variational approach [8,9], clustering techniques [10], the fuzzy c-means (FCM) methods [11,12], graph theory [13,14] and Bayesian approach [15,16,2,17]. Recently Dalmau and Rivera presented a general framework for probabilistic image segmentation [17]. Using this framework one can obtain new probabilistic segmentation models by selecting a metric-divergence between discrete probability distributions [18]. Based on the previous framework, in this work we present a new model β -MMF (β -Markov measure field) for probabilistic image segmentation. This model has theoretical implications. First, it could be considered as a generalization of the *quadratic Markov measure field* (QMMF) models [2]. Second, the model presented here is related with fuzzy c-means [19,20] based-methods for image segmentation. Our model relies on a generalized measure between discrete probability distributions [21]. We are going to study the particular case of the β -MMF that produces a family of Half-quadratic Markov measure field (HQMMF) models.

This paper is organized as follows. In Section 2 we make a brief review of the Bayesian formulation for probabilistic segmentation. Also, we introduce the β -MMF model and a half-quadratic version that allows us to solve this model efficiently. Section 3 shows some experimental results and finally, in the last section, we present our conclusions.

2 Mathematical Formulation

2.1 Review of Bayesian formulation for Markov Measure Field (MMF).

In general, the probabilistic image segmentation problem can be described as the computation of the preference of a pixel to certain models or classes. Given K models $\mathcal{M} = \{m_1, m_2, \dots, m_K\}$, an observable vector measure random field \mathbf{Y} , *i.e.* at each pixel one observe a vector $\mathbf{Y}(r) = [Y_1(r), Y_2(r), \dots, Y_K(r)]^T$ such that $Y_k(r) \geq 0$ and $\sum_k^K Y_k(r) = 1$, and a generative model

$$\Psi(\mathbf{Y}(r), \mathbf{X}(r), \mathbf{N}(r)) = \mathbf{0}, \forall r \in \mathcal{L}, \quad (1)$$

where the function Ψ should satisfy the hypothesis of the *implicit function theorem*, *i.e.* there exists a function Γ such that $\mathbf{N}(r) = \Gamma(\mathbf{Y}(r), \mathbf{X}(r))$, $\mathbf{N}(r)$ is a random variable with known distribution and \mathbf{X} is a hidden measure Markov random field. If \mathbf{x}, \mathbf{y} are realizations of \mathbf{X}, \mathbf{Y} respectively, the problem consists in finding an estimator \mathbf{x}^* . Using a Bayesian formulation, the MAP estimator \mathbf{x}^* can be computed as

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} U(\mathbf{x}; \mathbf{y}) \stackrel{def}{=} D(\mathbf{x}; \mathbf{y}) + \lambda R(\mathbf{x}), \quad (2)$$

$$\text{s.t. } \sum_k x_k(r) = 1, x_k(r) \geq 0, \forall r \in \mathcal{L}, \quad (3)$$

where $D(\mathbf{x}; \mathbf{y})$ is the likelihood energy, $R(\mathbf{x})$ is the prior energy, λ is a regularization hyper-parameter and \mathcal{L} is the regular lattice that corresponds to the pixel sites. According to [17] $D(\cdot; \cdot)$ and $R(\cdot)$ are based on metric-divergence between distributions [21, 18]. The variational model (2) and (3) is a generalization [17] of the particular cases proposed in Refs. [15, 16, 2],

2.2 β -MMF

Based on the general formulation Eqs. (2)-(3) we present a new model for probabilistic segmentation. Then, we need to define the likelihood and the prior energies. First, we start with the prior energy. Here we use a very popular energy which relies on the Euclidean distance, that is

$$R(\mathbf{x}) = \sum_{r \in \mathcal{L}} \sum_{s \in \mathcal{N}_r} \omega_{rs} \|\mathbf{x}(r) - \mathbf{x}(s)\|^2, \quad (4)$$

where ω_{rs} is a weight function *i.e.* $\omega_{rs} = \frac{\gamma}{\gamma + \|\mathbf{G}(r) - \mathbf{G}(s)\|^2}$, $\gamma = 10^{-3}$ and \mathcal{N}_r is the set of first neighboring pixels of r . The energy (4) promotes piecewise-smooth spatial changes of the vector measure field \mathbf{x} . This spatial smoothness is controlled by the parameter $\lambda > 0$ in the functional (2). Therefore, if λ is very large then the segmented regions tend to be large, and *vice-versa*, the smaller the value of λ the more granular is the solution. We note that the prior energy (4) corresponds to the functional of the Random Walker algorithm [22]. However, this energy has largely been used in different image segmentation models, see Refs [15, 16, 23, 2, 17].

Second, we obtain the β -Measure as a particular case of the $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ -measure between two discrete probability distributions \mathbf{f}, \mathbf{h} , i.e. $\sum_k f_k = 1, f_k \geq 0$ and $\sum_k h_k = 1, h_k \geq 0$, see Ref. [21]. This measure is defined as

$$I_{(\gamma, \delta)}^{(\alpha, \beta)}(\mathbf{f}, \mathbf{h}) = (2^{-\beta} - 2^{-\delta})^{-1} \sum_k (f_k^\alpha h_k^\beta - f_k^\gamma h_k^\delta), \quad (5)$$

where $\alpha, \beta, \gamma, \delta > 0$. The β -Measure is defined as the particular case

$$I^\beta(\mathbf{f}, \mathbf{h}) \stackrel{\text{def}}{=} \lim_{\alpha \rightarrow 0} I_{(\beta, 0)}^{(\beta, \alpha)}(\mathbf{f}, \mathbf{h}), \quad (6)$$

and therefore

$$I^\beta(\mathbf{f}, \mathbf{h}) = -\frac{1}{\log 2} \sum_k f_k^\beta \log h_k. \quad (7)$$

Using the previous measure in the likelihood term of the functional (2), we obtain the β -MMF model:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} U^\beta(\mathbf{x}; \mathbf{y}), \quad (8)$$

$$\text{s.t. } \sum_k x_k(r) = 1, \forall r \in \mathcal{L}, \quad (9)$$

where the functional $U^\beta(\cdot; \cdot)$ is defined in the following way

$$U^\beta(\mathbf{x}; \mathbf{y}) \stackrel{\text{def}}{=} \sum_{r \in \mathcal{L}} \left(\sum_k -\frac{1}{\beta} x_k^\beta(r) \log y_k(r) + \frac{\lambda}{2} \sum_{s \in \mathcal{N}_r} \omega_{rs} \|\mathbf{x}(r) - \mathbf{x}(s)\|^2 \right) \quad (10)$$

The parameters β, γ, λ could be trained, see Experiment Section, or can be manually tuned. In interactive experiments we use $\beta = 1.5, \gamma = 10^{-3}$ and $\lambda = 10^6$. Note that for $\beta \geq 1$ the previous functional is convex, so the constraint optimization problem (8)-(9) has an unique solution, and for $0 < \beta < 1$ the optimization problem is not convex.

Two interesting particular cases of the β -MMF models are obtained when $\beta = 1$ and $\beta = 2$. If $\beta = 1$ then β -measure becomes the Kerridge's measure [24], also known as the *Cross-entropy*, and the likelihood energy in (10) is linear. If $\beta = 2$ then the β -MMF models becomes the QMMF model [2].

2.3 Half-Quadratic MMF

The optimization problem β -MMF could be solved using gradient descent method [25]. However, here we present a more efficient alternative. When $\beta \in (0, 2]$ we can minimize (8)-(9) by using the half-quadratic approach introduced by Geman and Reynolds [26]. In Appendix A we present an outline of the proof, see Ref. [27] for details about half-quadratic minimization technique. Therefore, the solution is computed by iterating until convergence the following two steps (assuming a initial point for x):

1. Update $d_k(r) \stackrel{def}{=} -\frac{2}{\beta} x_k^{\beta-2}(r) \log y_k(r)$
2. Solve, approximately, the quadratic programming problem:

$$\min_x \frac{1}{2} \sum_{r \in \mathcal{L}} \left(\sum_k x_k^2(r) d_k(r) + \lambda \sum_{s \in \mathcal{N}_r} \omega_{rs} \|\mathbf{x}(r) - \mathbf{x}(s)\|^2 \right), \quad (11)$$

subject to $\sum_k x_k(r) = 1$, $x_k(r) \geq 0$, $\forall r$.

This half-quadratic technique produces an interactive scheme and the solution to (11), at iteration $t + 1$, can be obtained using the Lagrange multipliers method, see [28]. Here we present two algorithms: one for the multi-class segmentation and the second one for the binary segmentation problem.

- For multi-class segmentation we obtain the following iterative scheme

$$x_k(r) = \frac{\pi(r) + \lambda \sum_{s \in \mathcal{N}_r} \omega_{rs} x_k(s)}{d_k(r) + \lambda \sum_{s \in \mathcal{N}_r} \omega_{rs}}, \quad (12)$$

where the Lagrange multipliers are approximated with

$$\pi(r) = \frac{1}{K} \sum_k d_k(r) x_k(r) = E_N(d_N), \quad (13)$$

that is, $\pi(r)$ is the expected value of $d_N \in \{d_1(r), d_2(r), \dots, d_K(r)\}$ with respect to $N \in \{1, 2, \dots, K\}$, with discrete probability distribution $\mathbf{x}(r)$.

Note that if the previous solution at time t is positive then the obtained vector at time $t + 1$ is also positive, see expression (12).

- For binary segmentation: The kind of segmentation is very useful in many applications, for instance: organ segmentation, object tracking, foreground/background or object/no-object segmentation. Although, we can use two classes in the multi-class scheme, a more efficient approach can be developed. First, the functional is rewritten as follows

$$U^\beta(\mathbf{x}_1; \mathbf{y}_1, \mathbf{y}_2) = \frac{1}{2} \sum_{r \in \mathcal{L}} x_1^2(r) d_1(r) + (1 - x_1(r))^2 d_2(r) + \lambda \sum_{s \in \mathcal{N}_r} \omega_{rs} (x_1(r) - x_1(s))^2,$$

Again, the solution can be obtained using the Lagrange multipliers method. This produces the iterative formulas:

$$x_1(r) = \frac{d_2(r) + \lambda \sum_{s \in \mathcal{N}_r} \omega_{rs} x_1(s)}{d_1(r) + d_2(r) + \lambda \sum_{s \in \mathcal{N}_r} \omega_{rs}}, \quad (14)$$

where

$$d_1(r) = -\frac{2}{\beta} x_1^{\beta-2}(r) \log y_1(r), \quad (15)$$

$$d_2(r) = -\frac{2}{\beta} x_2^{\beta-2}(r) \log y_2(r), \quad (16)$$

$$x_2(r) = 1 - x_1(r). \quad (17)$$

We remark that for $\beta \in (0, 1)$ the functional is non-convex and in general the solution is a local minimum. To obtain a ‘good’ local minimum we can apply some kind of Graduated Non-Convexity method [29]. If $\beta \in (1, 2]$ the problem is convex and the global minimum is guaranteed.

2.4 Observation modeling

The observation modeling depends strongly on the problem we are dealing with. Here we present two examples. First, for the very popular interactive image segmentation task and second, for model parameter estimation problem.

Interactive Segmentation In the interactive segmentation problem the observation models can be represented by intensity or color histograms of user-marked regions [14,30]. This seed regions are known in the case of foreground/background segmentation as *trimap* [30], and in the case of multi-objects segmentation as *multimap*. Consider that some pixels in certain region of interest, Ω , are interactively labeled by a user. If \mathcal{K} is the class label set, we define the pixels set (region) that belongs to the class k as $R_k = \{r : \mathcal{R}(r) = k\}$, and

$$\mathcal{R}(r) \in \{0\} \cup \mathcal{K}, \forall r \in \Omega, \quad (18)$$

is the label field (class map or *multimap*) where $\mathcal{R}(r) = k > 0$ indicates that the pixel r is assigned to the class k and $\mathcal{R}(r) = 0$ if the pixel class is unknown and needs to be estimated. Let g be an image such that $g(r) \in t$, where $t = \{t_1, t_2, \dots, t_T\}$ with $t_i \in \mathbb{R}^n$, $n = 1, 3$. $n = 1$ in the case of gray level images (t_i represents an intensity value) and $n = 3$ in the case of color level images (t_i represents a RGB color). Let $h_k(t) : \mathbb{R}^n \rightarrow \mathbb{R}$ be the empirical histogram on the marked pixels which belong to class k , i.e. the ratio between the number of pixels in R_k whose intensity (or RGB color) is t and the total number of pixels in the region R_k . We denote as $\hat{h}_k(t)$ the smoothed normalized histograms (i.e. $\sum_t \hat{h}_k(t) = 1$), then the preference (observation) of the pixel r to a given class k is computed with:

$$y_k(r) = \frac{\hat{h}_k(g(r)) + \epsilon}{\sum_{j=1}^K [\hat{h}_j(g(r)) + \epsilon]}, \epsilon = 10^{-3}, \forall k > 0. \quad (19)$$

One can use more complex statistical models for defining the preference functions, for instance parametric models such as Gaussian Mixture Models. However, in the experiments we work with low dimension feature space (1 and 3 dimensions for gray and color images respectively), so the smoothed normalized histograms are computationally more efficient, i.e. they are implemented as look up tables. For higher dimension of the feature space, parametric models are in general more appropriate.

Model Estimation In this case we consider that the preference measure of a pixel r to some model k , or the preference to belonging to the region R_k , could be described

through a parametric model, *i.e.* $y_k(r) = f_k(r, g, \theta_k)$, $k \in \{1, 2, \dots, K\}$. The problem consists in computing simultaneously the vector measure field \mathbf{x} and the set of parameters $\theta = \{\theta_1, \theta_2, \dots, \theta_K\}$.

$$(\mathbf{x}^*, \theta^*) = \arg \min_{(\mathbf{x}, \theta)} U^\beta(\mathbf{x}; \mathbf{y}(\theta)), \quad (20)$$

$$\text{s.t. } \sum_k x_k(r) = 1, \forall r \in \mathcal{L}, \quad (21)$$

In general, this is a very hard problem and commonly highly non-linear. One alternative to face these drawbacks is to use the two step Segmentation/Model estimation (SM) algorithm [31,32]. First, we minimize with respect to \mathbf{x} fixing the parameters θ (Segmentation step). Second, we minimize with respect to θ fixing the parameters \mathbf{x} (Model estimation step). These two steps are repeated until a convergence criteria.

We illustrate this through an example. Consider the observations are given by Gaussian functions: $y_k(r) = \exp(-\|g(r) - \theta_k\|^2)$. In this case, for the *segmentation step* we use the half-quadratic technique explained in Section 2.3. For the *model estimation step* we obtain the following closed formula:

$$\theta_k = \frac{\sum_r x_k^\beta(r) g(r)}{\sum_r x_k^\beta(r)}.$$

3 Experiments

3.1 Image Binary Segmentation

We evaluate the performance of the β -MMF model using the binary segmentation problem. First, we have 4 data-sets, three of them composed by the letters of English alphabet in three fonts. The last one is composed by 65 maps of different countries. The images are normalize into the interval $[0, 1]$, and we add, to each data-set, 5 levels of Gaussian noise with zero mean and standard deviation in $\{0.2, 0.4, 0.6, 0.8, 1.0\}$, see Fig 1. Finally we renormalize each data-set into $[0, 1]$. In summary, we obtained 20 data-sets with a total of 715 noisy images. The original data-sets, without noise, is the groundtruth. To measure the segmentation quality we use *mean square error* (MSE) between the groundtruth (t) and the segmented image (s) using the β -MMF model, that is:

$$MSE(s, t) = \frac{1}{|\mathcal{L}|} \sum_{r \in \mathcal{L}} (s(r) - t(r))^2 \quad (22)$$

The observations are modeled using Gaussian functions, see Section 2.4. We made two kind of experiments. One with fixed models between $[0, 1]$, in the experiment we use $\theta_1 = \frac{1}{16}, \theta_2 = \frac{15}{16}$ as the mean of the models. And the other using parameter model estimation. Then, we have two set of parameters: the parameters of the models θ and the hyper-parameters of the algorithm $\Theta = [\beta, \lambda, \gamma]^T$. In order to obtain the best results of the algorithm, *i.e.* the ones with less MSE error, we train hyper-parameter set

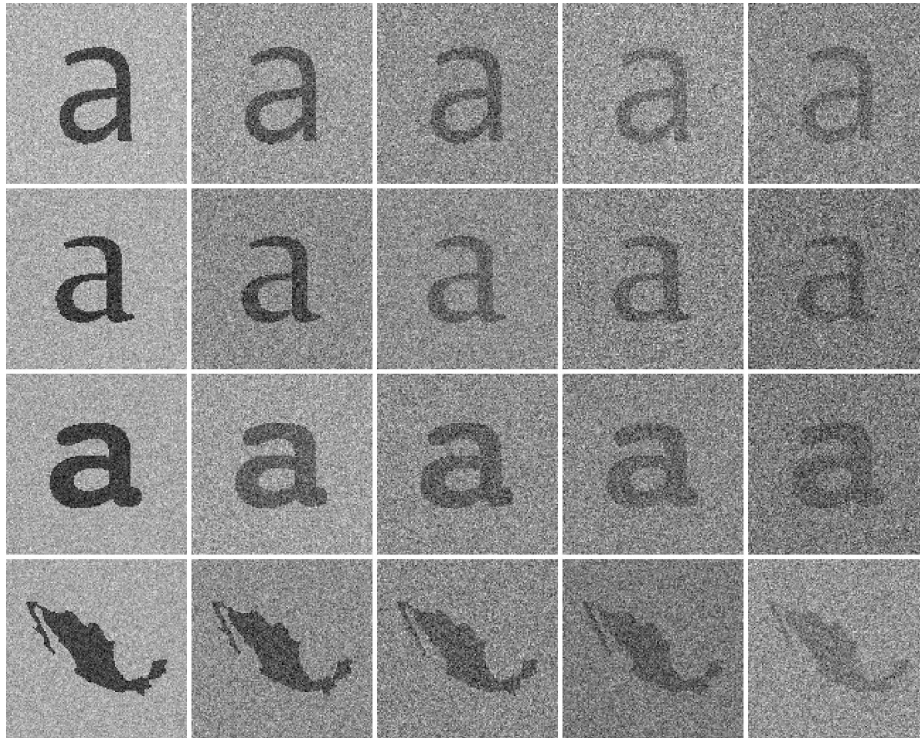


Fig. 1. Representative image of each noisy data-set. First row: Andale mono font, Second row: Constantia font, Third row: Courier bold font and Fourth row: Maps. From left to right: images with different levels of Gaussian noise with zero mean and standard deviations in $\{0.2, 0.4, 0.6, 0.8, 1.0\}$.

Table 1. The best parameters obtained after training the β -MMF with Andale Mono Dataset.

noise std dev	Fix models				Model estimation			
	λ	γ	β	training error	λ	γ	β	training error
0.2	0.52	0.08	0.64	0.01	0.54	0.02	0.90	0.01
0.4	0.52	0.12	0.81	0.20	0.22	0.06	1.08	0.21
0.6	0.57	0.08	0.97	1.41	0.47	0.07	1.48	8.01
0.8	0.51	0.14	0.97	11.54	0.47	0.09	1.52	16.20
1.0	1.06	0.11	0.97	18.36	0.53	0.08	1.39	22.18

Table 2. The best parameters obtained after training the β -MMF with Constantia Dataset.

noise std dev	Fix models				Model estimation			
	λ	γ	β	training error	λ	γ	β	training error
0.2	0.52	0.10	0.62	0.01	0.53	0.02	0.92	0.01
0.4	0.52	0.09	0.89	0.32	0.52	0.02	0.94	0.34
0.6	0.50	0.12	0.91	3.53	0.47	0.04	1.56	8.30
0.8	0.52	0.13	0.97	11.25	0.55	0.05	1.48	14.47
1	0.51	0.12	0.96	19.50	0.31	0.20	1.33	18.10

Θ using the Nelder–Mead method [33,34]. In particular, we use the implementation in the Numerical Recipes [34]. The results of the training, for all data-sets, are shown in Tables 1, 2, 3, 4. We note that while the level noise increases the training error also increases. Obviously, this is what one expects. For noise standard deviation less than 0.6 the results are in general good, see Fig 2. However, for noise standard deviation greater than 0.6 the result is poor, see Fig. 3. This could be explained through the Fig 4. This figure shows the histograms of the images of first row in Fig 1. When the standard deviation is 0.2 we can distinguish two models. However, when the standard deviation increases the parameter of the models collapse.

Table 3. The best parameters obtained after training the β -MMF with Courier Bold Dataset.

noise std dev	Fix models				Model estimation			
	λ	γ	β	training error	λ	γ	β	training error
0.2	0.53	0.10	0.57	0.01	0.54	0.01	1.08	0.01
0.4	0.52	0.10	0.86	0.24	0.53	0.01	0.90	0.24
0.6	0.51	0.11	0.93	0.71	0.51	0.01	1.05	0.67
0.8	0.48	0.12	0.97	8.96	0.48	0.08	1.63	8.16
1	0.46	0.16	0.95	13.99	0.47	0.02	1.61	10.11

Table 4. The best parameters obtained after training the β -MMF with Maps Dataset.

noise std dev	Fix models				Model estimation			
	λ	γ	β	training error	λ	γ	β	training error
0.2	0.60	0.02	1.08	0.03	0.56	0.01	0.63	0.03
0.4	0.54	0.05	0.90	0.36	0.57	0.01	0.92	0.37
0.6	0.52	0.09	0.97	0.87	0.19	0.05	1.02	0.71
0.8	0.55	0.13	0.97	4.40	0.22	0.35	0.88	1.81
1	0.55	0.11	0.97	10.73	0.48	0.08	1.45	6.48

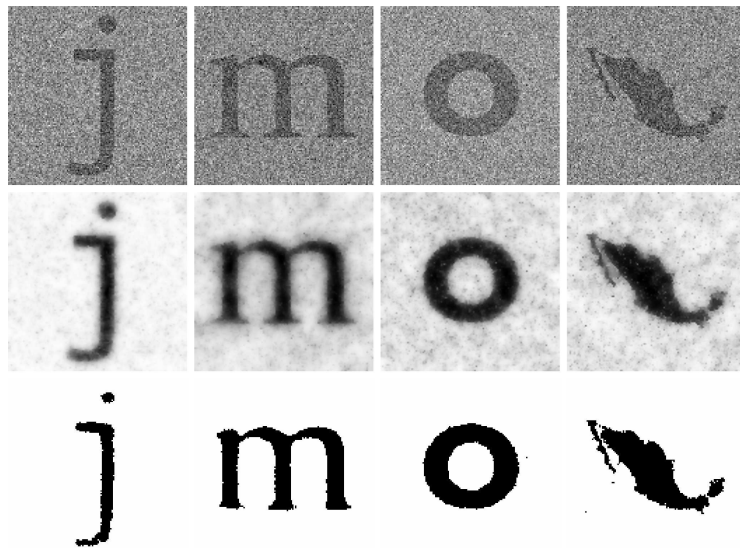


Fig. 2. Segmentation of selected images with noise standard deviation 0.6. First row: noisy images to be segmented, second row: soft segmentation, second row: hard segmentation.

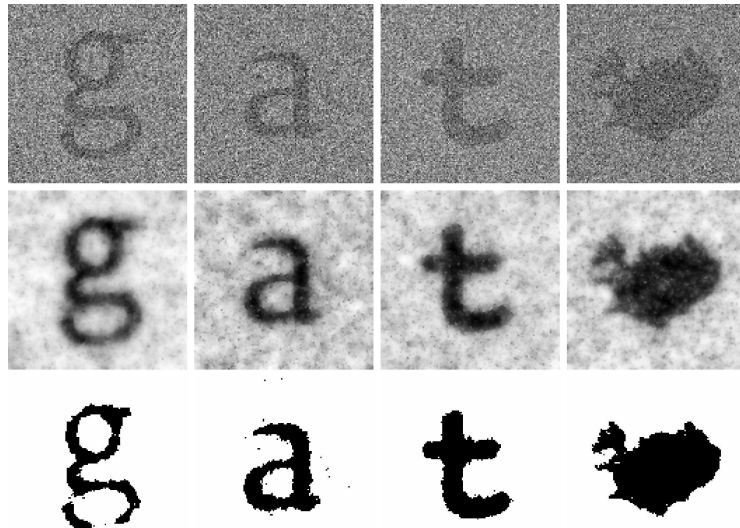


Fig. 3. Segmentation of selected images with noise standard deviation 1.0. First row: noisy images to be segmented, second row: soft segmentation, second row: hard segmentation.

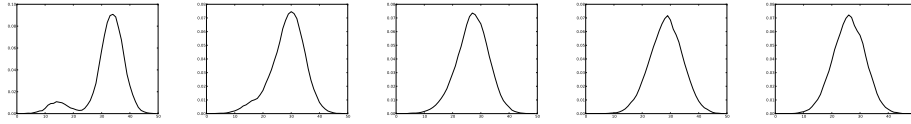


Fig. 4. From left to right: Image histograms from first row of Fig 1 with Gaussian noise with zero mean and standard deviations in $\{0.2, 0.4, 0.6, 0.8, 1.0\}$.

3.2 Interactive Segmentation Application

Fig. 5 depicts an application of the probabilistic segmentation methods for interactive image editing task. In this example we use the editing scheme proposed in Ref. [7] and the β -MMF model, Sections 2.2, 2.3. The likelihood is computed using the scribbles provided by a user, first column of Fig. 5, and following Section 2.4. The first column shows scribbles on the image to be edited, the second column shows the segmented image, and the third and fourth columns show edited images. In the horses image, first row, we blend three images, one source image per class, see [7] for details. The used parameters for this experiment are $\beta = 1.5$, $\gamma = 10^{-3}$ and $\lambda = 10^6$.

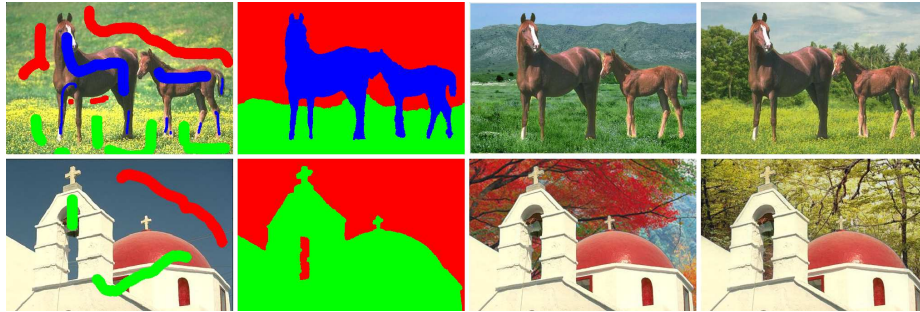


Fig. 5. Interactive image editing using the models β -MMF and the probabilistic editing framework proposed in Ref. [7]. First column: Scribbles made on the original image, Second column: segmented images, Third and fourth columns: edited images using several source images (one per class).

4 Conclusions

We presented a new model for probabilistic segmentation (β -MMF). This model generalizes the *Quadratic Markov Measure Field* model (QMMF) and can be seen as a Half-Quadratic variant that keeps convex the energy for low entropy promotion. As it

is demonstrated in our experiments, this is an important characteristic of the β -MMF models for the task of simultaneous model estimation and image segmentation. In particular, this model presents good results for binary image segmentation. The β -MMF models can be used for other applications, as for instance: image editing. In future work we will focus in improving our method for model parameter estimation in the case of corrupted images with high variance noise.

Acknowledges. This work was supported by the Consejo Nacional de Ciencia y Tecnología, Mexico: DSc. Scholarship to O. Dalmau and Grant 61367-Y to M Rivera.

Appendix A.

Here we show that the optimization problem (8)-(9) can be solved using the half-quadratic [26,27] technique. Here we follow the notation and methodology described in Ref. [27]. Then, one needs to find the conditions that the function $\rho(x) = x^\beta$ should satisfy to be used in a half-quadratic regularization. According to [27]

$$\rho(x) = \min_z x^2 z + \Psi(z), \quad (23)$$

where

$$\Psi(z) = \phi((\phi')^{-1}(z)) - z(\phi')^{-1}(z), \quad \phi(x^2) \stackrel{def}{=} \rho(x).$$

Based on the previous definition, one obtains that

$$\psi(z) = \frac{2 - \beta}{2} \left(\frac{2}{\beta} z \right)^{\frac{\beta}{\beta-2}}.$$

As the optimization problem (23) should have a minimum at $z = \phi'(x^2) = \frac{\rho'(x)}{2x} = \frac{\beta}{2} x^{\beta-2}$, see Ref. [27] for details¹, then the coefficient of $\psi(z)$, i.e. $\frac{2-\beta}{2}$, should be positive, and therefore $\beta \leq 2$. Another way to obtain the previous result is using the condition that the function $\phi(x)$ should be convex, that is $\phi''(x) \leq 0$.

$$\phi''(x) = \frac{\beta}{2} \frac{\beta-2}{2} x^{\frac{\beta-2}{2}}$$

then, $\frac{\beta}{2} \frac{\beta-2}{2} \leq 0$ and again we conclude that $\beta \leq 2$.

Observe also that for $1 \leq \beta \leq 2$ the function $\phi(x)$ satisfies

$$\lim_{x \rightarrow 0} \phi'(x) = \infty, \quad \lim_{x \rightarrow \infty} \phi'(x) = 0, \quad (24)$$

that is, the function $\phi(x)$ does not satisfy the condition $\lim_{x \rightarrow 0} \phi'(x) = 1$, which is very important to define a bounded weight function $z = \frac{\beta}{2} x^{\beta-2}$ for $x > 0$. Note, however, that the ϕ -function of Total Variation (TV), i.e. $\rho(x) = |x|$, satisfies the

¹ The same result is obtained if one compute the derivative with respect to z of the function in Eq. (23) and set it equal to zero.

conditions (24). So, similar to TV [35] we can redefine the ρ -function in the following way $\tilde{\rho}(x) = \frac{2}{\beta} \epsilon^{1-\frac{\beta}{2}} (x^2 + \epsilon)^{\frac{\beta}{2}}$, where $\epsilon > 0$ is a small real value, for instance $\epsilon = 10^{-6}$. For this ρ -function, the corresponding weight function is $z = \phi'(x^2) = \frac{\tilde{\rho}'(x)}{2x} = \epsilon^{1-\frac{\beta}{2}} (x^2 + \epsilon)^{\frac{\beta}{2}-1}$. Now the function $\phi(x)$ satisfies:

$$\lim_{x \rightarrow 0} \phi'(x) = 1, \quad \lim_{x \rightarrow \infty} \phi'(x) = 0, \quad (25)$$

References

1. Kaus, M.R., Warfield, S.K., Nabavi, A., Black, P.M., Jolesz, F.A., Kikinis, R.: Automated Segmentation of MR Images of Brain Tumors. *Radiology* **218** (2001) 586–591 [1](#)
2. Rivera, M., Ocegueda, O., Marroquín, J.L.: Entropy-controlled quadratic markov measure field models for efficient image segmentation. *IEEE Transactions on Image Processing* **16** (2007) 3047–3057 [1](#), [2](#), [3](#)
3. Hower, D., Singh, V., Johnson, S.: Label set perturbation for mrf based neuroimaging segmentation. In: *IEEE International Conference on Computer Vision ICCV09*. (2009) 849–856 [1](#)
4. Chamorro-Martinez, J., Sanchez, D., Prados-Suarez, B.: A Fuzzy Colour Image Segmentation Applied to Robot Vision. In: *In Advances in Soft Computing Engineering, Design and Manufacturing*, Springer (2002) [1](#)
5. Mishra, A., Aloimonos, Y.: Active segmentation for robots. In: *International Conference on Intelligent ROBots and Systems*. (2009) [1](#)
6. Dalmau, O., Rivera, M., Mayorga, P.P.: Computing the alpha-channel with probabilistic segmentation for image colorization. In: *IEEE Proc. Workshop in Interactive Computer Vision (ICV'07)*. (2007) 1–7 [1](#)
7. Dalmau, O., Rivera, M., Alarcon, T.: Bayesian Scheme for Interactive Colourization, Recolourization and Image/Video Editing. to appear in *Computer Graphics Forum* (2010) [1](#), [10](#)
8. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problem. *Commun. Pure Appl. Math.* (1989) 577–685 [1](#)
9. Hewer, G.A., Kenney, C., Manjunath, B.S.: Variational image segmentation using boundary functions. *IEEE Transactions on Image Processing* **7** (1998) 1269–1282 [1](#)
10. Weiss, Y.: Segmentation using eigenvectors: A unifying view. In: *ICCV* (2). (1999) 975–982 [1](#)
11. Ahmed, M.N., Yamany, S.M., Mohamed, N., Farag, A.A., Moriarty, T.: A modified fuzzy c-means algorithm for bias field estimation and segmentation of mri data. *IEEE Trans. Med. Imaging* **21**(3) (2002) 193–199 [1](#)
12. Chuang, K.S., Tzeng, H.L., Chen, S., Wu, J., Chen, T.J.: Fuzzy c-means clustering with spatial information for image segmentation. *Computerized Medical Imaging and Graphics* **30** (2006) 9–15 [1](#)
13. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(8) (2000) 888–905 [1](#)
14. Boykov, Y., Jolly, M.P.: Interactive organ segmentation using graph cuts. In: *MICCAI, LNCS* 1935. (2000) 276–286 [1](#), [5](#)
15. Marroquin, J.L., Velazco, F., Rivera, M., Nakamura, M.: Gauss-markov measure field models for low-level vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23** (2001) 337–348 [1](#), [2](#)

16. Marroquin, J.L., Arce, E., Botello, S.: Hidden Markov measure field models for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25** (Nov 2003) 1380–1387 [1](#), [2](#)
17. Dalmau, O., Rivera, M.: A general bayesian markov random field model for probabilistic image segmentation. In: LNCS 5852, IWCI. (2009) 149–161 [1](#), [2](#)
18. Cha, S.H.: Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. *International Journal of Mathematical Models and Methods in Applied Sciences* **1** (2007) 300–307 [1](#), [2](#)
19. Bezdek, J.C., Coray, C., Gunderson, R., Watson, J.: Detection and characterization of cluster substructure i. linear structure: Fuzzy c-lines. *SIAM Journal on Applied Mathematics* **40**(2) (Apr. 1981) 339–357 [1](#)
20. Bezdek, J.C., Coray, C., Gunderson, R., Watson, J.: Detection and characterization of cluster substructure ii. fuzzy c- varieties and convex combinations thereof. *SIAM Journal on Applied Mathematics* **40**(2) (Apr. 1981) 358–372 [1](#)
21. Taneja, I.J., Gupta, H.: On generalized measures of relative information and inaccuracy. *Applications of Mathematics* **23** (1978) 317–333 [1](#), [2](#), [3](#)
22. Grady, L., Schiwietz, T., Aharon, S., Westermann, R.: Random Walks for interactive organ segmentation in two and three dimensions: Implementation and validation. In: MICCAI (2), LNCS 3750. (2005) 773–780 [2](#)
23. Grady, L.: Multilabel random walker image segmentation using prior models. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 1 of CVPR., San Diego, IEEE (June 2005) 763–770 [2](#)
24. Kerridge, D.: Inaccuracy and inference. *Journal of the Royal Statistical, Series B* (1961) 184–194 [3](#)
25. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer Series in Operation Research (2000) [3](#)
26. Geman, D., Reynolds, G.: Constrained restoration and the recovery of discontinuities. *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(3) (1992) 367–383 [3](#), [11](#)
27. Black, M., Rangarajan, A.: On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *Int'l J. Computer Vision* **19**(1) (1996) 57–92 [3](#), [11](#)
28. Rivera, M., Dalmau, O., Tago, J.: Image segmentation by convex quadratic programming. In: ICPR, IEEE (2008) 1–5 [4](#)
29. Blake, A., Zisserman, A.: Visual Reconstruction. MIT Press (1987) [5](#)
30. Juan, O., Keriven, R.: Trimap segmentation for fast and user-friendly alpha matting. In: VLISM, LNCS 3752. (2005) 186–197 [5](#)
31. Neal, R.M., Hinton, G.E.: A view of the EM algorithm that justifies incremental, sparse, and other variants. In Jordan, M.I., ed.: *Learning in Graphical Models*, Kluwer Academic Publishers, Boston MA. (1998) 355–368 [6](#)
32. Marroquín, J.L., Santana, E.A., Botello, S.: Hidden Markov measure field models for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25** (Nov 2003) 1380–1387 [6](#)
33. Nelder, J.A., Mead, R.: A simplex method for function minimization. *Comput. J.* **7** (1965) 308–313 [8](#)
34. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical recipes in C (2nd ed.): the art of scientific computing. Cambridge University Press, New York, NY, USA (1992) [8](#)
35. Vogel, C.R., Oman, M.E.: Iterative methods for total variation denoising. *SIAM J. Sci. Comput* **17** (1996) 227–238 [12](#)