A General Bayesian Markov Random Field Model for Probabilistic Image Segmentation.

Oscar Dalmau[1] and Mariano Rivera[1]

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 general Bayesian model for image segmentation with
spatial coherence through a Markov Random Field prior. We also study variants
of the model and their relationship. In this work we use the Matusita Distance,
although our formulation admits other metric-divergences. Our main contribu-
tions in this work are the following. We propose a general MRF-based model for
image segmentation. We study a model based on the Matusita Distance, whose
solution is found directly in the discrete space with the advantage of working in
a continuous space. We show experimentally that this model is competitive with
other models of the state of the art. We propose a novel way to deal with non-
linearities (irrational) related with the Matusita Distance. Finally, we propose an
optimization method that allows us to obtain a hard image segmentation almost
in real time and also prove its convergence.

1   Introduction

Image segmentation is an important field in computer vision. It has been one of the most
studied tasks in image processing and is considered to be a bridge between low and high
level image processing tasks. The image segmentation problem consists of obtaining a
partition of the image according to a homogeneous predicate. Depending on the image
modeling viewpoint, many strategies have been proposed. If the image is modeled from
a deterministic point of view, we can find different image segmentation approaches
based on the Mumford-Shah functional [1,2]. When it is modeled as a graph [3] we
find Graph Cut [4] and Normalized Cut [5]. In the context of data clustering, the fuzzy
c-means (FCM) methods are widely used [6,7] and if the image is modeled as a Markov
Random Field (MRF) several approaches have been reported [8,9,10]. Among them,
Bayesian formulation, including spatial coherence through a MRF prior, has shown to
be a powerful framework to design efficient and robust models for image segmentation.

In this work, we formulate the segmentation problem as the minimization of an en-
ergy functional in terms of the likelihood using a Bayesian regularization formulation
based on MRF. Our strategy reduces the probabilistic image segmentation modeling to
the choice of a convenient metric-divergence or a measure between distributions. In this
work we study the Matusita Distance (MD) and the possibility of using it for both hard
and soft segmentation. We also propose a way to deal with the non-linearities involved
with MD. We propose an algorithm for solving a particular case of a quadratic opti-
mization problem through a sequence of linear programming subproblems and prove
its convergence.

2 Mathematical Formulation

2.1 Notation

In general, the segmentation problem can be written in terms of a set of sites and a set of labels. Let $\mathcal{L} = \{r = (n, m) : n = 1, 2, \ldots, N \,; m = 1, 2, \ldots, M\}$ be a 2D regular lattice where $|\mathcal{L}| = NM$ represents the number of sites (or pixels) in the image. The neighborhood of the pixel r will be denoted as $\mathcal{N}_r$. Conveniently, we will also denote (or reindex) pixels (n, m) in the image as the subset of indices $\mathcal{I} = \{1, 2, \ldots, |\mathcal{L}|\}$ where $r_i, i \in \mathcal{I}$ represents a site in the image. The image is denoted by I, $I(r)$ is a scalar (the intensity in case of a gray level image) or a vector (color channels in the case of a color image). Let $\mathcal{K} = \{1, 2, \ldots, K\}$ be a subset of labels. A segmentation of an image is a mapping $\mathcal{R} : \mathcal{L} \to \mathcal{K}$ from the set of sites $\mathcal{L}$ to the set of labels $\mathcal{K}$. That is to say, it is a partition of $\mathcal{L}$ in regions $\mathcal{R}_k \subseteq \mathcal{L}$ in such a way that: $\cup_{k \in \mathcal{K}} \mathcal{R}_k = \mathcal{L}$ and $(i \neq j; i, j \in \mathcal{K}) \Rightarrow \mathcal{R}_i \cap \mathcal{R}_j = \emptyset$, where $\mathcal{R}_k = \{r \in \mathcal{L} : \mathcal{R}(r) = k\}$.

According to [11] the observation model at each pixel r of the image can be written as follows: $I(r) = \sum_{k \in \mathcal{K}} p_k(r) I_k(r) + \eta(r)$, where $p_k(r) \in [0, 1]$ is the component of a vector measure field $\boldsymbol{p}$ at a pixel r, $\boldsymbol{p}(r) = (p_1(r), p_2(r), \ldots, p_K(r))$, and represents the probability that the pixel r belongs to the class k (or to the region $\mathcal{R}_k$). Then $p_k(r)$ must satisfy $\sum_{k \in \mathcal{K}} p_k(r) = 1$. Moreover, $\eta(r)$ is white noise and $I_k(r)$ is the intensity value of the layer I_k at the pixel r.

Definition 1. *The set of points* $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ *that satisfy* $x_i \geq 0, \forall i \in \{1, 2, \ldots, n\}$ *and*

$$\sum_{i=1}^{n} (x_i)^\gamma = 1, \tag{1}$$

will be called γ positive unit n-hypersurface *and will be denoted as:* S_γ^n.

For instance, if $\gamma = 1$ the positive unit n-hypersurface will be named *positive unit n-hyperplane* or a *simplex* (S^n) and if $\gamma = 2$ will be called *positive unit n-hypersphere*.

2.2 Review

Rivera et al. [11] defined a Consistence Condition Qualification (CCQ) that an Energy Functional for Gauss-Markov Measure Fields (GMMF) models should satisfy in an image segmentation problem.

Definition 2. (Consistence Condition Qualification) *If no a priori information about the vector measure field* $\boldsymbol{p}$ *is available, and also, its Maximum Likelihood (ML) estimator satisfies the condition:* $\arg\max_k p_k(r) = \arg\max_k \hat{v}_k(r), \quad \forall r \in \mathcal{L}$*, where* $\hat{\boldsymbol{v}}$ *represents the normalized likelihood*[1]*, then we say that the vector measure field* $\boldsymbol{p}$ *holds the Consistence Condition Qualification (CCQ).*

---

[1] The likelihood is defined as $v_k(r) = p(I(r)|p_k(r) = e_k, I_k)$.

In this work, we are interested in two particular CCQ cases. The first CCQ, allows us to obtain a soft segmentation of the image and from the second CCQ we can formulate a functional to obtain a hard segmentation. Now we introduce two definitions that will guide our study.

Definition 3. (Equality Consistence Condition) *If no a priori information about the vector measure field $\boldsymbol{p}$ is available, and also, its ML estimator satisfies the condition $p_k(r) = \hat{v}_k(r), \ \ \forall r \in \mathcal{L}$, where $\hat{\boldsymbol{v}}$ represents the normalized likelihood, we say that the vector measure field $\boldsymbol{p}$ holds the Equality Consistence Condition (ECC).*

Definition 4. (Hard Consistence Condition) *If no a priori information about the vector measure field $\boldsymbol{p}$ is available, and also, its ML estimator satisfies the following conditions*

$$k_r^* = \arg\max_k p_k(r) = \arg\max_k \hat{v}_k(r), \ \ \forall r \in \mathcal{L}, \tag{2}$$

$$p_{k_r^*}(r) = 1, \ \ \forall r \in \mathcal{L}, \tag{3}$$

where $\hat{\boldsymbol{v}}$ represents the normalized likelihood, then we say that the vector measure field $\boldsymbol{p}$ holds the Hard Consistence Condition (HCC).

2.3 General model

In the Bayesian regularization formulation, based on Markov Random Field, for instance: GMMF [8], HMMF [9] and QMMF [10], the modeling (observation model) begins with some assumptions about the generative model of the observation (the image). After some mathematical derivation, an energy functional that depends on the likelihood is obtained. In our formulation, different from the above models, we consider that the vector field $\boldsymbol{v}$ is the observation. Using a Bayesian formulation, the posterior distribution is

$$P(\boldsymbol{p}|\boldsymbol{v}) \propto P(\boldsymbol{v}|\boldsymbol{p})P(\boldsymbol{p}) = e^{-U(\boldsymbol{p};\boldsymbol{v})}, \tag{4}$$

the conditional probability $P(\boldsymbol{v}|\boldsymbol{p})$ is obtained from

$$P(\boldsymbol{v}|\boldsymbol{p}) = \prod_{r \in \mathcal{L}} P(\boldsymbol{v}(r)|\boldsymbol{p}(r)) = e^{-D(\boldsymbol{p};\boldsymbol{v})},$$

where $P(\boldsymbol{v}(r)|\boldsymbol{p}(r))$ is the observation model and $P(\boldsymbol{p}) \propto e^{-\lambda R(\boldsymbol{p})}$ is a Gibbsian distribution. In our case, we are not interested in the model that generates the data, but in finding energies that have particular properties, for instance those which satisfy CCQ, in particular HCC or ECC. Of course, each particular energy must have an underlaying observation model as in the cases mentioned above but we are not interested in this point. Therefore, we formulate the segmentation problem as the minimization of an energy functional (the MAP estimation of Eq. (4)). This problem has the general form:

$$\arg\min_{\boldsymbol{p} \in S_\gamma^n} U(\boldsymbol{p}; \boldsymbol{v}) = D(\boldsymbol{p}; \boldsymbol{v}) + \lambda R(\boldsymbol{p}), \tag{5}$$

where the parameter λ is positive and controls the granularity of regions to be segmented, $D(\boldsymbol{p}; \boldsymbol{v})$ is the data term and can be related to the likelihood term in a Bayesian

formulation, $R(\boldsymbol{p})$ is the regularization term and represents the prior knowledge about the vector measure field $\boldsymbol{p}$. As we have said before, $D(\boldsymbol{p}; \boldsymbol{v})$ must be chosen in such a way that enforces CCQ, and in particular we are interested in the HCC and ECC cases. In our formulation, we consider that the observations are given by the likelihood $\boldsymbol{v}$, different from [8], [9] and [10] where the image is assumed to be the observation. So, we regularize directly on the likelihood. Then, the image segmentation problem can be solved by minimizing a metric-divergence between models associated to a certain feature vector [8,9]. Under this scheme many functionals can be obtained for segmentation purposes. A particular case of the functional in Eq. (5) is the following

$$U(\boldsymbol{p}; \hat{\boldsymbol{v}}) = D(\boldsymbol{p}; \hat{\boldsymbol{v}}) + \lambda R(\boldsymbol{p}), \tag{6}$$

that depends on the normalized likelihood $\hat{\boldsymbol{v}}$. A natural choice for the data term $D(\boldsymbol{p}; \hat{\boldsymbol{v}})$ in the functional (6) is to select a distance between distributions. On the other hand, in the data term of the potential (5) we can use other measures, for instance: divergence measures, inaccuracy measures, metrics and distances. The problem now is what measure to choose. In the literature we can find a lot of metric-divergences[2] between distributions [12]. Some of these metric-divergences have a symmetric version, for example, the Kullback and Leibler divergence, the main drawback is that they are highly non-linear. This leads us to a highly non-linear optimization problem. So, the algorithmic properties are an important issue in the selection of metric-divergences.

2.4 Model for ECC

In this section we study the Matusita Distance (MD)[3] [13]. Let $\boldsymbol{f}, \boldsymbol{h}$ be discrete probability distributions, i.e. $\sum_{i=1}^{K} f_i = \sum_{i=1}^{K} g_i = 1$ and $f_i, g_i \geq 0$, then the MD is defined as:

$$d(\boldsymbol{f}||\boldsymbol{h}) = \sqrt{2 - 2\sum_{i=1}^{K} \sqrt{f_i h_i}} \; . \tag{7}$$

This measure has a geometric interpretation and satisfies the axioms to be a metric. Hence, $d(\boldsymbol{f}||\boldsymbol{h}) = 0 \iff \boldsymbol{f} = \boldsymbol{h} \iff f_i = g_i$ [4] and therefore, MD can be used to obtain ECC models. Using (7) in the data term of (6) we obtain:

$$D(\boldsymbol{p}; \hat{\boldsymbol{v}}) = \sum_{r \in \mathcal{L}} \frac{1}{2}\big[d(\boldsymbol{p}(r)||\hat{\boldsymbol{v}}(r))\big]^2 = \sum_{r \in \mathcal{L}} \Big[1 - \sum_{i=1}^{K} \sqrt{p_i(r)\hat{v}_i(r)}\Big] \; . \tag{8}$$

---

[2] From the mathematical point of view, they are not metrics because they do not satisfy the axioms to be a metric, most of them only satisfy the non-negativity and identity of indiscernible axioms, i.e.: $d(a, b) \geq 0$ and $d(a, b) = 0$ iff $a = b$, where a, b are elements of a metric space.

[3] MD is proportional to Hellinger Distance (HD), see [13].

[4] Observe that $d(\boldsymbol{f}||\boldsymbol{h}) = 0 \iff \boldsymbol{f}||\boldsymbol{h}$, i.e. the vectors $\boldsymbol{f}, \boldsymbol{h} \in \mathbb{R}^K$ are parallels, using the Cauchy-Bunyakovsky-Schwarz inequality. Then, there exists $\lambda \in \mathbb{R}$ such that $\boldsymbol{f} = \lambda \boldsymbol{h}$. Using the condition $\sum_{i=1}^{K} f_i = \sum_{i=1}^{K} g_i = 1$ we obtain that $\lambda = 1$ and finally $f_i = g_i$.

Similarly, the regularization potential is proposed in terms of the metric (7):

$$R(\boldsymbol{p}) = \sum_{r \in \mathcal{L}} \sum_{s \in \mathcal{N}_r} \frac{1}{2} \big[d(\boldsymbol{p}(r) \| \boldsymbol{p}(s)) \big]^2 = \sum_{r \in \mathcal{L}} \sum_{s \in \mathcal{N}_r} \Big[1 - \sum_{i=1}^{K} \sqrt{p_i(r) p_i(s)} \Big] \ . \quad (9)$$

As we can see, both terms (8) and (9) are non-rational and the minimization of (6) becomes a non-linear system. To deal with this problem, we propose to minimize the following functional that is a transformation of the original energy functional (6):

$$U(\tilde{\boldsymbol{p}}; \tilde{\boldsymbol{v}}) = U(\mathcal{T}_1 \boldsymbol{p}; \mathcal{T}_2 \hat{\boldsymbol{v}}) = D(\mathcal{T}_1 \boldsymbol{p}; \mathcal{T}_2 \hat{\boldsymbol{v}}) + \lambda R(\mathcal{T}_1 \boldsymbol{p}), \quad (10)$$

where $\mathcal{T}_j : [0,1] \to [0,1]$ with $j \in \{1,2\}$ are increasing real functions.

So, instead of computing directly the vector measure field $\boldsymbol{p}$, we propose to compute the transformed $\tilde{\boldsymbol{p}} = \mathcal{T}_1 \boldsymbol{p}$ and $\tilde{\boldsymbol{v}} = \mathcal{T}_2 \hat{\boldsymbol{v}}$. To simplify the notation it is understood that when we apply $\mathcal{T}_j$ to a vector, we really mean to apply $\mathcal{T}_j$ on each component of the vector. Hence, substituting in Eq. (5) we have a new functional (10) in terms of $\tilde{\boldsymbol{p}}$ and $\tilde{\boldsymbol{v}}$.

A natural way for selecting $\mathcal{T}_j$ in the case of MD is choosing power functions $\mathcal{T}_j(.) = (.)^n$. We study the case $\mathcal{T}_1 = \mathcal{T}_2$ and $n = \frac{1}{2}$, that corresponds to a mapping of the simplex S^K onto the positive unit K-hypersphere S_2^K. Then, the original optimization problem is transformed into:

$$\min_{\tilde{\boldsymbol{p}} \in S_2^K} U(\tilde{\boldsymbol{p}}; \tilde{\boldsymbol{v}}), \ \tilde{\boldsymbol{v}} \in S_2^K \ . \quad (11)$$

The optimization problem (11) satisfies the following proposition:

Proposition 1. *The constraint quadratic optimization problem defined by Eq. (11) satisfies ECC when no a priori information is available (i.e. $\lambda = 0$).*

The proof is straightforward. We can see by using the Cauchy-Bunyakovsky-Schwarz inequality, that the vector measure field $\tilde{\boldsymbol{p}}$ under no prior knowledge (setting $\lambda = 0$) satisfies the ECC. Otherwise, the above optimization problem is a quadratic optimization problem with quadratic constraints that can be solved by using the Lagrange multiplier method. The Lagrangian for the constrained optimization problem, without including the non-negativity constraints, is

$$\mathcal{L}(\tilde{\boldsymbol{p}}, \boldsymbol{\pi}; \tilde{\boldsymbol{v}}) = U(\tilde{\boldsymbol{p}}; \tilde{\boldsymbol{v}}) + \sum_{r \in \mathcal{L}} \pi(r) \Big[\sum_{i=1}^{K} \tilde{p}_i^2(r) - 1 \Big], \quad (12)$$

where $\pi(r), \ r \in \mathcal{L}$ are the Lagrange multipliers.

Now, we obtain the first-order necessary conditions for optimality, the Karush-Kuhn-Tucker (KKT) conditions, see [14]:

$$\frac{\partial \mathcal{L}(\tilde{\boldsymbol{p}}, \boldsymbol{\pi}; \tilde{\boldsymbol{v}})}{\partial \tilde{p}_k(r)} = 0, \forall r \in \mathcal{L}, k \in \mathcal{K}; \quad (13)$$

$$\sum_{k \in \mathcal{K}} p_k(r) = \sum_{k \in \mathcal{K}} \mathcal{T}_1^{-1}(\tilde{p}_k(r)) = 1, \forall r \in \mathcal{L} \ . \quad (14)$$

From the KKT condition in Eq. (13) we obtain

$$-\tilde{v}_k(r) - \lambda \sum_{s \in \mathcal{N}_r} \tilde{p}_k(s) + 2\pi(r)\tilde{p}_k(r) = 0 \ \forall r \in \mathcal{L} \ . \tag{15}$$

Defining $n_k(r) \overset{def}{=} \tilde{v}_k(r) + \lambda \sum_{s \in \mathcal{N}_r} \tilde{p}_k(s)$ and using a Gauss-Seidel scheme in the equation system (15), we can compute the vector measure field $\tilde{p}$ through an iterative process. By using the equality constraints (14) and substituting (15) we can compute the Lagrange multipliers $\pi(r)$. Substituting $\pi(r)$ in Eq. (15) we obtain the vector measure field $\tilde{p}$ whose components are computed with

$$\tilde{p}_k(r) = \frac{n_k(r)}{\sqrt{\sum_{i=1}^{K} n_i^2(r)}} \ . \tag{16}$$

Note that, if the vector measure field $\tilde{p}$ is initialized with non-negative values (i.e. $\tilde{p}^{(0)} \geq 0$) then the solution $\tilde{p}^*$ implicitly satisfies the non-negativity constraints and then $\tilde{p}^* \in S_2^K$. Finally, the optimum vector measure field p^* is computed by applying the inverse transformation of $\mathcal{T}_1$ to $\tilde{p}^*$: $p_k^*(r) = [\tilde{p}_k^*(r)]^2$.

2.5 Model for HCC

Based on the formulation defined by Eq. (11) we can think of changing the hypersurface on which $\tilde{p}$ is located. We study the case when $\tilde{p} \in S^K$ (i.e. we change the equality restrictions expressed in the L_2-Norm and use the L_1-Norm). Then, we have the following optimization problem:

$$\min_{\tilde{p} \in S^K} U(\tilde{p}; \tilde{v}) \ . \tag{17}$$

Under no prior knowledge, the above problem satisfies the following proposition:

Proposition 2. *If $\forall r \in \mathcal{L} \ \exists k_r \in \mathcal{K}$ such that $\tilde{v}_{k_r}(r) > \tilde{v}_k(r) \ \forall k \neq k_r$ then the constraint quadratic optimization problem defined by Eq. (17) satisfies HCC when no a priori information is available (i.e. $\lambda = 0$).*

Observe that if we do not have prior knowledge, then the problem (17) is a linear programming (LP) problem. As is well-known in LP, under the conditions of the proposition the solution lies on a vertex of the polytope defined by the constraint and hence the solution at each pixel is $p(r) = e_{k_r}$, where e_{k_r} is a vector of the canonical base, i.e. $\tilde{p}_k(r) = \delta(k - k_r)$ and $k_r = \arg\max_{k \in \mathcal{K}} \hat{v}_k(r), \forall r \in \mathcal{L}$.

On the other hand, if prior energy (i.e. $R(\tilde{p}) = \sum_r \sum_{s \in N_r} 1 - \tilde{p}(r)^T \tilde{p}(s)$) is given, then the optimization problem in (17) becomes a quadratic programming (QP) problem. In such a case, we still can use the idea explained above, but adapted to the QP problem (17).

We propose to solve the QP problem using a sequential linear programming (SLP) strategy. First, let us define the linear functional, at each step t

$$U^l(\tilde{p}; \hat{V}^{(t)}) = \sum_{i=1}^{|\mathcal{L}|} -\tilde{p}^T(r_i)\hat{V}^{(t)}(r_i), \tag{18}$$

where the components of $\tilde{\boldsymbol{V}}^{(t)}(r_i)$ are defined as $\tilde{\boldsymbol{V}}_k^{(t)}(r_i) \stackrel{def}{=} \tilde{v}_k(r_i) + \lambda \sum_{r_j \in \mathcal{N}_{r_i}} \tilde{p}_k^{(t)}(r_j)$. Then, we need to solve the following LP subproblem at each step:

$$\min_{\tilde{\boldsymbol{p}} \in S^{\mathcal{K}}} U^l(\tilde{\boldsymbol{p}}; \tilde{\boldsymbol{V}}^{(t)}) \ . \tag{19}$$

Observe that the functional defined in (18) is the linearization around $\tilde{\boldsymbol{p}}^{(t)}$ of the QP problem (17). In practice, we are not going to minimize the problem (19) as a whole, but instead we are going to reduce the energy minimizing pixel by pixel. So, we need to redefine $\tilde{V}_k^{(t)}(r_i)$ as follows:

$$\tilde{V}_k^{(t)}(r_i) \stackrel{def}{=} \tilde{v}_k(r_i) + \lambda \left(\sum_{r_j \in \mathcal{N}_{r_i}, j<i} \tilde{p}_k^{(t+1)}(r_j) + \sum_{r_j \in \mathcal{N}_{r_i}, j>i} \tilde{p}_k^{(t)}(r_j) \right) \ . \tag{20}$$

Therefore, in the current site $c \in \mathcal{I}$ we use the information of the previous step t and the updated sites in the current step $t+1$ (i.e. all the sites i such that $i < c$), see Eq. (20), as a Gauss-Seidel scheme, or *coordinate descent method*, see [14]. The strategy previously explained is summarized in the SLP algorithm.

Algorithm 1 SLP Algorithm

Given a starting point $\tilde{\boldsymbol{p}}^{(0)}$
repeat
    for $i \in \mathcal{I}$ **do**
        Evaluate $\tilde{\boldsymbol{V}}^{(t)}(r_i)$ using (20)
        Solve (19) to obtain $\tilde{\boldsymbol{p}}^{(t+1)}(r_i)$
    end for
until Convergence

Proposition 3. *The SLP algorithm converges, if at each step the vector $\tilde{\boldsymbol{V}}^{(t)}(r_i)$ has a unique maxima component for all $r_i, \ i \in \mathcal{I}$.*

For proof see next Subsection.

2.6 Relation between ECC and HCC proposed models

The models studied in Subsections 2.4 and 2.5 can be enclosed in a more general model by changing the hypersurface on which $\tilde{\boldsymbol{p}}$ is defined. Now, we study the case in which $\tilde{\boldsymbol{p}}$ belongs to the γ unit positive hypersurface S_γ^K

$$\min_{\tilde{\boldsymbol{p}} \in S_\gamma^K} U(\tilde{\boldsymbol{p}}; \tilde{\boldsymbol{v}}) \ . \tag{21}$$

This model is a generalization of the models presented in the previous sections. It establishes the relation between ECC and HCC models and allows us to give a proof of

the Algorithm 1 presented in Subsection 2.5. Following the same methodology used in Subsection 2.4 we obtain the following expression for $\tilde{p}_k(r)$:

$$\tilde{p}_k(r) = \frac{[n_k(r)]^{\frac{1}{\gamma-1}}}{\left\{\sum_{i=1}^{K} [n_i(r)]^{\frac{\gamma}{\gamma-1}}\right\}^{\frac{1}{\gamma}}} \ . \tag{22}$$

Observe that HCC model is the limit case of the problem (21) as γ approaches to 1 from the right side. If we make the assumption that for all $r \in \mathcal{L}$ there exists k_r such that $n_{k_r}(r) > n_i(r)$, $\forall i \neq k_r$ and we let γ tend to 1, then for $p_k(r) = \tilde{p}_k^\gamma(r)$ where $k \in \mathcal{K}$ we have:

$$\lim_{\gamma \to 1} \tilde{p}_k^\gamma(r) = \delta(k_r - k) = p_k(r) \tag{23}$$

where $k_r = \arg\max_{i \in \mathcal{K}} p_i(r)$. This means that for $\gamma = 1$ the update procedure for the vector $p(r)$ is simply to assign $p(r) = e_{k^*}$ where $k^* = \arg\max_k \tilde{p}_k(r)$. The relation obtained in Eq. (23) is a proof of Proposition 3.

3 Experiments

In this section, we make two quantitative comparisons between the proposed methods and others of the state of the art. Also, we present some experimental results of the ECC method proposed in Subsection 2.4 for interactive image segmentation and for color image segmentation.
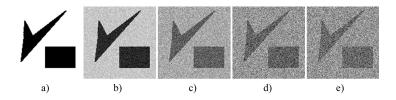
As a comparison measure we use the Jaccard similarity coefficient (or Jaccard index).

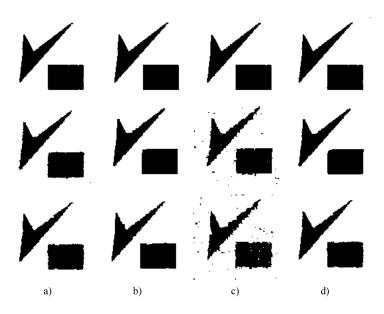$$J_{I_1,I_2}(k) = \frac{P_{I_1 \cap I_2}(k)}{P_{I_1 \cup I_2}(k)} \ . \tag{24}$$

The Jaccard coefficient (JC) measures the similarity between two sets. It is defined as the size of the intersection of the sets divided by the size of the union of the sets. In Eq. (24) $P_{I_1 \cap I_2}(k)$ denotes the number of pixels that belong to class k in both the original image (taken as the ground truth) and the segmented image. $P_{I_1 \cup I_2}(k)$ denotes the number of pixels that belong to class k in the original image or in the segmented image.

For evaluating the robustness, we compare the proposed methods with the following methods of the state of the art: Graph cut and QMMF. We use the binary image in Fig. 1 a). When we apply typical noise values to image Fig. 1 a), i.e. between 20 and 40 dB of Peak Signal-to-Noise Ratio (PSNR) where a lower level of noise (in dB) is worse, all the methods obtained a JC equal to 1 for both classes. This means that all the methods have excellent performance when using typical values of PSNR.

In the next experiments we reduce the PSNR (i.e. we increase the noise level) up to 3 dB, see Fig. 1 c)-e). For the image with 10 dB PSNR, Fig. 1 c), the proposed methods presented as good results as those of the state of the art, see the JC in Table 1 and for segmentation results see Fig. 2. In the images with 5 and 3 dB PSNR, see Fig. 1 d)-e),

Fig. 1. Binary image with different Peak Signal-to-Noise Ratio, a) Original image, b) 20 dB, c) 10 dB, d) 5 dB, e) 3 dB.

Fig. 2. Segmentation using different methods a) QMMF b) Graph cut c) Hard-segmentation proposed model, d) Soft-segmentation proposed model. First row corresponds to Fig. 1 c), second row corresponds to Fig. 1 d) and third row corresponds to Fig. 1 e).

the resulting quality of the hard-segmentation proposed method is reduced compared with the remainder methods, see Table 1 and Fig. 2 c). But experimentally, the HCC proposed model presents competitive results when the PSNR is greater than 10 dB, and is very fast. In an experiment with 1000 runs the reported average time was 0.00273 seconds, so it can be used in some real time image processing tasks. On the other hand, the soft-segmentation proposed method is very robust but it is slower, see the JC of the experiment in Table 1 and the segmentation in Fig. 2 d). For this experiment, the ECC proposed model presents competitive results according to the Jaccard index.

Table 1. Jaccard index results using different segmentation methods. The segmented images are Fig. 1 c)-e) that are obtained from Fig. 1 a) after applying the noise levels 10, 5 and 3 dB PSNR respectively.

Method	10 dB PSNR		5 dB PSNR		3 dB PSNR	
	white class	black class	white class	black class	white class	black class
QMMF	0.9907	0.9972	0.9707	0.9911	0.9618	0.9885
Graph cut	0.9936	0.9981	0.9759	0.9927	0.9720	0.9916
HCC	0.9907	0.9972	0.9274	0.9773	0.8743	0.9597
ECC	0.9923	0.9977	0.9772	0.9931	0.9631	0.9888

We also evaluate the performance of the ECC proposed method in the interactive color Image Segmentation task based on trimaps. For the experiments we use the Lasso benchmark database available online in Ref. [15]. According to the results reported in Refs. [16,17] our proposal demonstrates a better performance compared with methods of the state of the art. In Refs. [16,17] the authors compared the following methods: GraphCut, Gaussian Markov Measure Fields (GMMF), Random Walker (RW), Quadratic Markov Measure Fields (QMMF or QMPF) and Entropy Controlled Quadratic Markov Measure Fields (EC-QMMF or EC-QMPF).

Table 2 shows, for each image of Lasso's benchmark database, a comparison between the errors obtained by the best method reported in Ref. [17] (EC-QMPF) and our ECC method. In Table 3 we summarize the experimental results reported in [17] and our results. In the last column of Table 3 we show Akaike's information criterion(AIC). As we can see, our method presents the best results. We notice that the improvement of ECC is marginal compared to the results obtained by EC-QMMF. The main advantage of our ECC method is that it has 2 free parameters.

Fig. 3 illustrates our method performance in the context of color image segmentation using the quasi-automatic segmentation method. In this experiment, we use the perceptual color likelihood obtained by Alarcon and Marroquin where the likelihood field is divided into 11 basic color categories, see [18] for details.

In Fig. 4 we show a possible application of the ECC proposed model for the colorization task. In this example we used the scheme for colorization proposed by Dalmau et al. in [19].

4 Conclusions

We propose a general model for image segmentation in the context of Bayesian formulation with MRF prior. Under this formulation the modeling of image segmentation is reduced to select an appropriate metric-divergence. If the data term depends on the likelihood then a more general data term can be conveniently chosen. When the data term depends on the normalized likelihood, a metric or distance between distributions should be used. We study two particular cases of the general formulation. As metric-divergence we use the Matusita Distance. In the first case, HCC-based formulation, we

Table 2. Comparative performance between EC-QMMF and our ECC method in the context of interactive image segmentation. For the experiments we use Lasso's benchmark database available online in Ref. [15]. The results of EC-QMMF method are reported in Ref. [17]. The values in the table represent the classification percentage errors for each method.

Filename	EC-QMMF	ECC	Filename	QMPF+EC	ECC
21077	4.01	4.03	bush	7.86	5.93
20077	4.21	3.60	ceramic	1.73	1.66
37073	1.44	1.62	cross	1.75	1.60
65019	0.27	0.47	doll	1.03	0.52
69020	2.95	4.08	elefant	2.05	1.96
86016	1.99	2.28	flower	0.61	0.44
106024	7.55	7.19	fullmoon	0.27	0.00
124080	3.43	5.57	grave	1.27	1.65
153077	1.65	1.61	llama	4.32	3.82
153093	4.08	4.45	memorial	1.49	1.49
181079	7.41	7.70	music	2.26	2.03
189080	6.22	4.61	person1	1.16	0.25
208001	1.5	1.64	person2	0.71	0.30
209070	2.25	2.27	person3	0.87	0.75
227092	3.46	4.01	person4	3.27	3.59
271008	2.33	2.61	person5	2.48	2.22
304074	10.9	10.81	person6	5.19	4.92
326038	7.53	7.08	person7	0.96	0.59
376043	6.14	6.30	person8	0.93	0.92
388016	1.5	1.16	scissors	2.87	1.88
banana1	3.91	3.31	sheep	4.53	5.64
banana2	1.49	1.09	stone1	0.73	1.67
banana3	1.91	1.75	stone2	0.78	0.48
book	3.52	4.83	teddy	1.91	3.33
bool	1.74	1.65	tennis	7.31	5.60

propose a model for hard segmentation and we show experimentally that it gives competitive results compared to some methods of the state of the art. For this model, we give a fast algorithm based on a sequential linear programming and also we prove that under certain conditions the proposed method converges.
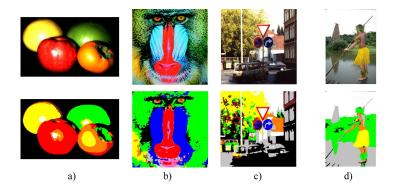
We study a second approach that satisfies the ECC condition. This model gives a soft segmentation method that is more robust to noise. We prove experimentally that this method gives excellent results and can be used for other image processing tasks.

References

1. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problem. Commun. Pure Appl. Math. (1989) 577–685 1
2. Hewer, G.A., Kenney, C., Manjunath, B.S.: Variational image segmentation using boundary functions. IEEE Transactions on Image Processing **7** (1998) 1269–1282 1

Table 3. Comparative summary of the results reported in Ref. [17] and our results. The values in the table are some statistics of the classification errors expressed in percentages.

	mean	median	stddev	parameters	AIC
GC	6.84	6.07	4.66	2	8.58
RW	5.47	4.66	3.58	2	6.50
GMMF	5.47	4.59	3.57	2	6.49
QMPF	5.08	3.87	3.46	2	6.04
QMPF+EC	3.03	2.15	2.40	3	3.58
ECC	2.98	2.12	2.34	2	3.42

a)   b)   c)   d)

Fig. 3. Color image segmentation using the soft-segmentation proposed method based on MD and the perceptual color likelihood proposed by Alarcon and Marroquin in [18].

3. Weiss, Y.: Segmentation using eigenvectors: A unifying view. In: ICCV (2). (1999) 975–982 1

4. Boykov, Y., Jolly, M.P.: Interactive organ segmentation using graph cuts. In: MICCAI, LNCS 1935. (2000) 276–286 1

5. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE PAMI **22**(8) (2000) 888–905 1

6. 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

7. 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

8. Marroquin, J.L., Velazco, F., Rivera, M., Nakamura, M.: Gauss-markov measure field models for low-level vision. IEEE PAMI **23** (2001) 1, 3, 4

9. Marroquin, J.L., Arce, E., Botello, S.: Hidden Markov measure field models for image segmentation. IEEE PAMI **25** (2003) 1380–1387 1, 3, 4

10. 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, 3, 4

Fig. 4. Colorization experiment using the soft-segmentation proposed method based on MD. a) Original image b) Multimap c)-d) Colorizations

11. Rivera, M., Ocegueda, O., Marroquin, J.L.: Entropy controlled Gauss-Markov random measure fields for early vision. In: VLSM. Volume LNCS 3752. (2005) 137–148 2
12. Taneja, I.: Generalized Information Measures and Their Applications, on-line book. (2001) 4
13. Cha, S.H.: Comprehensive survey on distance/similarity measures between probability density functions. INTERNATIONAL JOURNAL OF MATHEMATICAL MODELS AND METHODS IN APPLIED SCIENCES (2007) 300–307 4
14. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer Series in Operation Research (2000) 5, 7
15. : http://research.microsoft.com/vision/cambridge/ i3l/segmentation/GrabCut.htm 10, 11
16. Rivera, M., Mayorga, P.: Quadratic Markovian probability fields for image binary segmentation. In: Interactive Computer Vision (ICV'07). (2007) 10
17. Rivera, M., mayorga, P.P.: Comparative study on quadratic markovian probability fields for image binary segmentation. Technical Report I-07-15, Centro de Investigacion en Matematicas, A.C. (dec 2007) 10, 11, 12
18. Alarcón, T.E., Marroquín, J.L.: Linguistic color image segmentation using a hierarchical bayesian approach. *to appear in* Color Res Appl (2009) 10, 12
19. 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 10