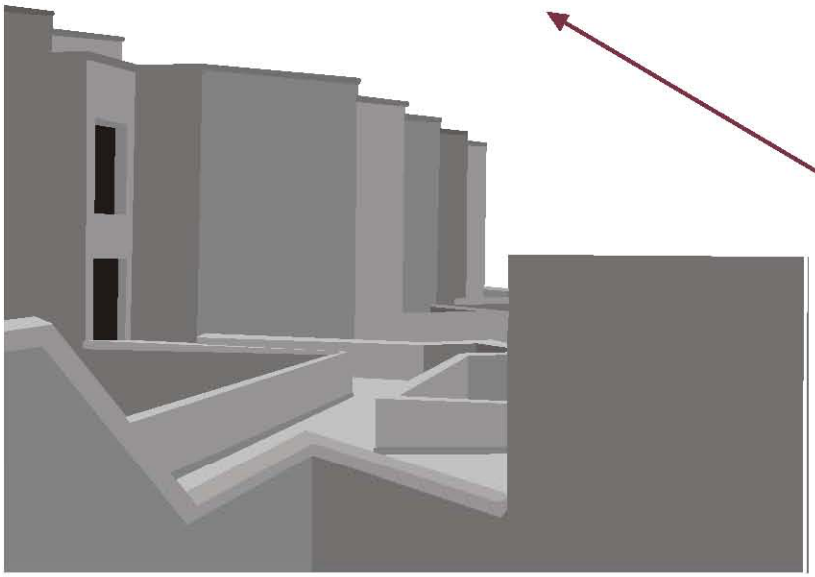


Comunicaciones del CIMAT

QUADRATIC PROGRAMMING FOR PROBABILISTIC IMAGE SEGMENTATION

Mariano Rivera & Oscar Dalmau

Comunicación del CIMAT No I-10-06/25-06-2010
(CC/CIMAT)

CIMAT

Quadratic Programming for Probabilistic Image Segmentation

Mariano Rivera and Oscar Dalmau

Comunicaciones del CIMAT

Abstract

We present a general framework for image segmentation based on quadratic programing, *i.e.* by the minimization of a quadratic regularized energy linearly constrained. In particular, we present a new and general derivation of the Quadratic Makov Measure Field models (QMMFs) that can be understood as a procedure for regularizing the model preferences (memberships or likelihood) as well as efficient optimization algorithms. In the QMMFs the uncertainty in the computed regularized probability measure field is controlled by penalizing the Gini's coefficient and hence it affects the convexity of the QP problem. The convex case is reduced to the solution of a positive definite linear system and, for that case, an efficient Gauss–Seidel scheme is presented. On the other hand, we present a efficient projected Gauss-Seidel with a subspace minimization for optimizing the non–convex case. We demonstrate the proposal capabilities by experiments and numerical comparisons with interactive two-class segmentation as well as in the simultaneous estimation of segmentation and (parametric and non-parametric) generative models.

This paper has been submitted to IEEE TRANSACTIONS ON IMAGE PROCESSING, December 2009.

Index Terms

Image segmentation, Quadratic programming, Interactive segmentation, Computer vision, Markov random fields.

I. Introduction

IMAGE SEGMENTATION is an active research topic in computer vision and is the core process in many practical applications, see for instance the listed in [1]. Given that image segmentation is a ill–posed problem that is task and user dependent, among many approaches, methods based on Markov Random Field (MRF) models have become popular for designing segmentation algorithms because their flexibility for being adapted to very different circumstances as: color, connected components, motion, stereo disparity, etc. [2], [3], [4], [5], [6], [7], [1]. See for instance the three posibble segmentation of a single scene in Fig. 1.

The MRF approach allows one to express the label assignment problem into an energy function that includes spatial context information for each pixel and thus promotes smooth segmentations. The energy function codifies the compromise of assigning a label to a pixel by depending on the value of the particular pixel and the value of the surrounding pixels. Since the label space is discrete, frequently, the segmentation problem requires of the solution of a combinatorial (integer) optimization problem. In that order, max–flow graph–cut based techniques [8], [9], [10], [11], [12], [13], [14], [2], [15] and spectral methods for graph–cut [16], [17], [18] are among the most successful solution algorithms. In particular, graph-cut based methods can solve the binary (two labels) segmentation problem in polynomial time [6]. Recently some authors have reported advances in the solution of the multi-label problem, their strategy consists on constructing an approximated problem by relaxing the integer constraint [18], [19]. Additionally, two important issues in discrete MRF are: the reuse of solutions in the case of dynamic MRF [10], [20] and the measurement of labeling uncertainty [20].

However, the combinatorial approach (hard segmentation) is neither the most computationally efficient, and, in some cases, the most precise strategy for solving the segmentation problem. A different approach is to directly estimate the uncertainties on the label assignment or memberships [5], [21], [7], [1], [22]. In the Bayesian framework,

The authors are with the Department of Computer Science, Centro de Investigacion en Matematicas AC , Guanajuato, GTO, Mexico 36000.

E-mail: mrivera@cimat.mx, see http://www.cimat.mx/˜mrivera.html

This work is supported in part by CONACYT (Grant 61367), Mexico and O. Dalmau was supported by a CIMAT AND CONACYT scholarships.

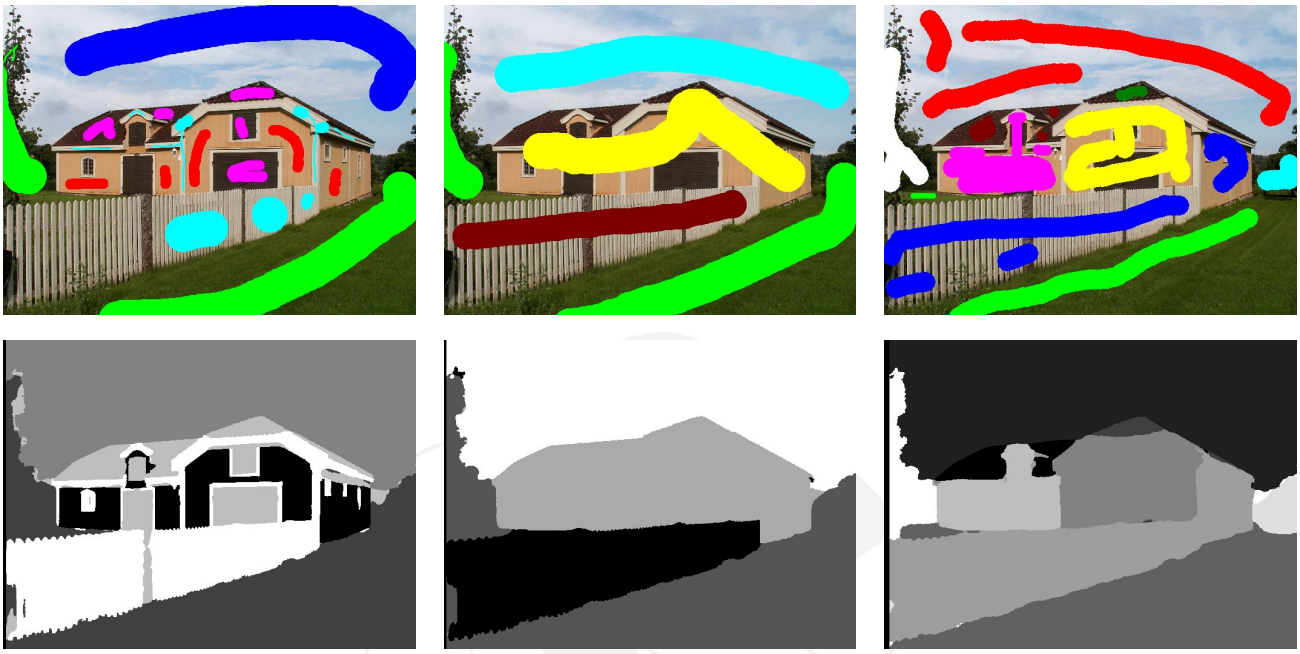


Fig. 1. Multi–class segmentation of a same scene according to different criteria (codified in the user scribbles). The columns correspond to segmentations by color, semantic objects and planar regions, respectively.

Comunicaciones del CIMAT

such a memberships can be expressed in a natural way in terms of probabilities—leading to the named probabilistic segmentation (PS) methods.

In this work we present new insights and extensions to the recent reported PS method named Quadratic Markov Measure Field models (QMMFs) [1]. In particular we present a general framework for PS. We demonstrate that the QMMF data term (potential) is a dissimilarity measure between discrete density distributions and satisfies the requirements of the proposed PS framework. We presents efficient optimization algorithms, proper for the two-classes (binary) and multi-classes segmentation. We demonstrate that the solution to a convex QMMFs is computed by solving a linear system. Since the entropy control proposed in Ref. [1] affects the convexity of the quadratic programing problem then we propose a projection strategy combined with a subspace minimization method for the nonconvex QMMF case [23].

Preliminary results of this work were in [24], [25], [26]. We organize this paper as follows. Section III reports a new derivation of the QMMF models based, the new derivation shows that the data term is an information measure between density distributions that preserve model preferences. In section IV we presents optimization algorithms for solving efficiently the QMMFs energies. Experiments that demonstrate the method performance are presented in section V. Finally, our conclusions are given in section VI.

II. Brief review of Entropy–Controlled Quadratic Markov Measure Field Models

Let r be the pixel position in the image or the region of interest, $\mathcal{R} = \{r\}$ (in a regular lattice $\mathcal{L}$). Then, $\mathcal{K} = \{1, \ldots, K\}$ denotes the set of index classes and $\mathbb{S}^K \subset \mathbb{R}^K$ denotes the simplex such that

$$z \in \mathbb{S}^K \tag{1}$$

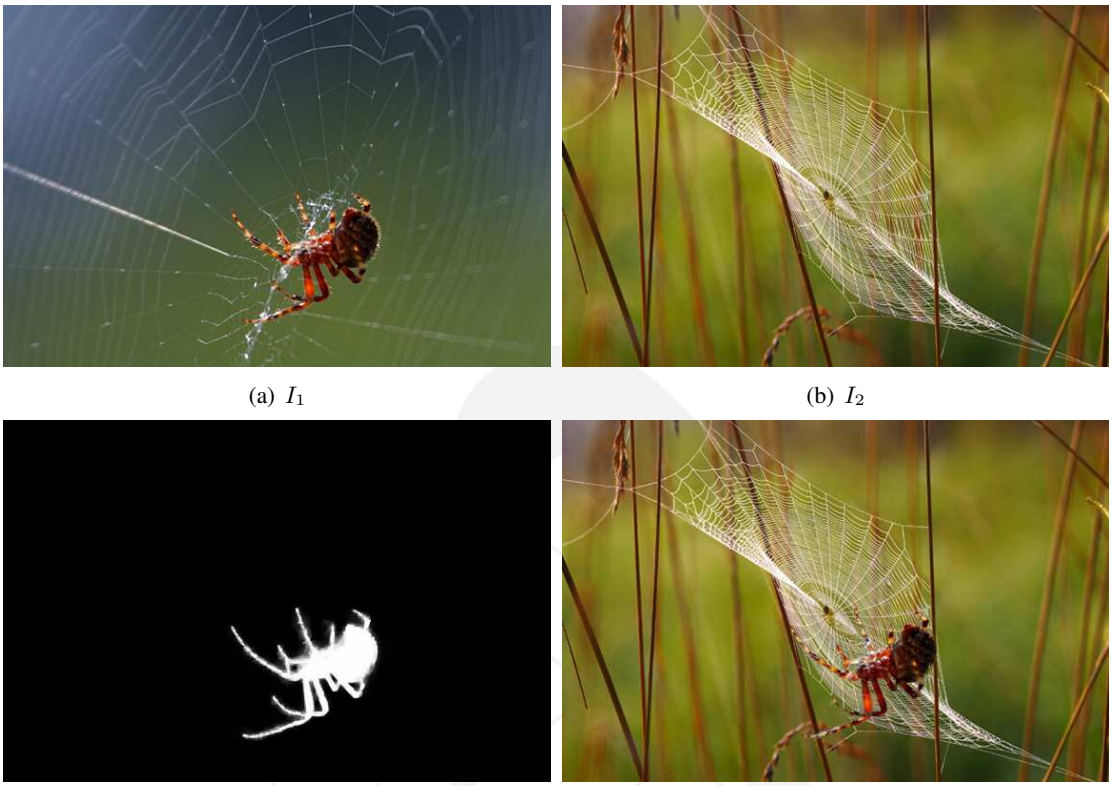
if and only if

$$\mathbb{1}^T z = 1, \tag{2}$$

$$z \succeq 0; \tag{3}$$

where the vector $\mathbb{1}$ has all its entries equal one and its size is defined by the context; in our notation $z \succeq 0$ with $z \in \mathbb{R}^K \iff z_k \geq 0$, for $k = 1, 2, \ldots, K$.

Recently, in Ref. [1] the Entropy Controlled Quadratic Markov Measure Fields (EC-QMMF) models for image multiclass segmentation were proposed. Such models are computationally efficient and produce a PS of excellent



(a) I_1 (b) I_2

(c) p_1 (d) $g = p_1 I_1 + p_2 I_2$

Fig. 2. Image model generation. I_1 and I_2 are the original data, p_1 is a matting factor vector (with $p_1 + p_2 = 1$ and $p_1, p_2 \geq 0$) and g is the observed image.

quality. Whereas hard segmentation procedures compute a hard label for each pixel, PS approaches, as QMMFs, compute the confidence of assigning a particular label to each pixel. In the Bayesian framework, the amount of confidence (or uncertainty) is represented in terms of probabilities. In this framework $p_k(r)$ denotes the unknown probability of the pixel r to belong to the class $k \in \mathcal{K}$. Hence the vector field p is a probability measure field, i.e. $p(r) \in \mathbb{S}^K$.

The QMMF formulation is based on the generative model:

$$g(r) \quad = \quad p(r)^T I(r) + \eta(r), \tag{4}$$

where g is the observed image, the images vector $I = [I_1, I_2, \ldots, I_K]^T$ is generated with a parametric model, $I_k(r) = \Phi_{\theta_k}(r)$, with known (or estimated) parameters $\theta = \{\theta_k\}, \forall k$; η is a possible noise and $p(r) \in \mathbb{S}^K$ is a matting vector that can be understood as a probability measure [1], [22]. Fig. 2 illustrates the generation image process assuming model (4). Based on this generational model, one can define the model preference as follows:

Definition The likelihood (model preference) $v_k(r)$ is the conditional probability of observing a particular pixel value $g(r)$ by assuming that such a pixel is taken from the image I_k:

$$v_k(r) \quad \stackrel{def}{=} \quad P(g(r)|p(r) = e_k, I), \tag{5}$$

where e_k is the kth canonical basis vector.

In the particular case of *i.i.d.* Gaussian Noise and gray scale images:

$$v_k(r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\|g(r) - I_k(r)\|^2}{2\sigma^2}\right]. \tag{6}$$

Then, according to [1], an effective PS of g can be computed solving the quadratic programing problem:

$$\min_p U(p) \quad \text{s.t.} \;\; p(r) \in \mathbb{S}^K, \quad \text{for} \;\; r \in \mathcal{R} \tag{7}$$

Comunicaciones del CIMAT

with

$$U(p) = \frac{1}{2} \sum_{r \in \mathcal{R}} \left\{ p(r)^T D_r p(r) + \frac{\lambda}{2} \sum_{y \in \mathcal{N}_r} \|p(r) - p(s)\|^2 w_{rs} \right\} \tag{8}$$

where $D_r = \mathrm{diag}(d(r))$ is a diagonal matrix, associated with the pixel r, with its k-th diagonal element computed with

$$d_k(r) = -\log \hat{v}_k(r) - \mu, \tag{9}$$

$\hat{v}(r) \in \mathbb{S}^K$ is the pixelwise normalized likelihood, the parameter μ controls the entropy of the solution (*i.e.* the amount of uncertainty in the PS), $\mathcal{N}_r = \{s \in \mathcal{R} : \|r - s\|_2 = 1\}$ is the set of first order pixel neighbors, the positive parameter λ controls the regularization (smoothness) process and the positive weights w lead the class border to coincide with large image gradients. In [24] the weight function w is computed with

$$w_{rs} = \frac{\gamma}{\gamma + \|T\{g(r)\} - T\{g(s)\}\|^2} \tag{10}$$

where γ is a positive parameter that controls the edge sensibility and T is in general a nonlinear transformation that depends on the task. For instance T can transform a color pixel value from the RGB space to the Lab space.

Then the memberships (p) and the parameters (θ) can be estimated by alternating partial minimizations until convergence:

1) $\min_p U(p, \theta)$ s.t. $p \in \mathbb{S}^K$, keeping fixed θ,
2) $\min_\theta U(p, \theta)$ keeping fixed p.

These minimization can be partially achieved as in a generalized EM scheme [27].

In the original proposal, QMMF models were derived from the observed model (4), assuming Gaussian noise, η, and measure vectors $p(r)$ with neglected entropy, *i.e.* the product $p_i(r)p_j(r) \approx 0$ for $i \neq j$ at any pixel r [1]. Then the generalization to other distributions than the Gaussian is justified, again in the low entropy limit, by the approach: $-\sum_k p_k^2(r) \log \hat{v}_k(r) \approx -\log \left[p(r)^T \hat{v}(r) \right]$, a dissimilarity distance between vectors. However, as we will prove, the low entropy constraint in p is not required to derive the QMMFs models and, according with our experiments, there exist problems where the best segmentation is computed with high entropic p's, experiments in subsection V-B.

III. On QMMF models

A. Probabilistic segmentation

Likelihoods v are of particular interest in image PS approaches. We frame PS methods in the next definition.

Definition *Consistence Condition Qualification* (CCQ). The information measure $M(p, q)$ preserves the CCQ if given the measure vector q, then probability measure $p^* = \mathrm{argmin}_p \, M(p, q)$ with $p \in \mathbb{S}^K$ satisfies: $\mathrm{argmax}_k \, p_k^* = \mathrm{argmax}_k \, q_k$.

Definition Let $p \in \mathbb{S}^K$ a discrete density function, then an information measure $D(p, q)$ is CCQ consistent iff $p^* = \arg\min_p D(p, q)$ s.t. $p \in \mathbb{S}^K$ is CCQ.

If a vector p holds the CCQ we say that p is CCQ. The CCQ definition provides a framework for developing probabilistic segmentation methods. In the case we are interested, Bayessian regularization, CCQ means that the data term does not alter the model's preference.

B. QMMF Data term

In this subsection we prove that the QMMF data term

$$\sum_{r \in \mathcal{R}} Q\left(p(r), \hat{v}(r)\right) \tag{11}$$

with

$$Q(p, q) = -\sum_k p_k^2 \log q_k \tag{12}$$

Communications in CIMAT

is an information measure that promotes fidelity of the measure vector $p(r)$ to the likelihood vector $\hat{v}(r)$ and holds CCQ. First we note the follows.

Proposition 3.1: The information function (12) is a dissimilarity (or inaccuracy) measure between the discrete distributions p and q.

To prove the last proposition we use the generalized $(\alpha, \beta, \gamma, \delta)$-information measure between two probability density functions [28]:

$$I_{(\gamma,\delta)}^{(\alpha,\beta)}(p;q) = \frac{\sum_k p_k^\alpha q_k^{\beta-\alpha} - p_k^\gamma q_k^{\gamma-\delta}}{\exp(\alpha-\beta) - \exp(\gamma-\delta)} \quad \text{with} \quad p, q \in \mathbb{S}^K. \tag{13}$$

The we note that (13) reduces to the Q–dissimilarity (12) when $\alpha = \gamma = \delta = 2$ and in the limit as $\beta \to 2$, a direct result of of the L'Hospital's rule.

The Q–dissimilarity can be written using the positive definite diagonal matrix $A = \operatorname{diag}(-\log q)$ as $p^T A p$. In particular all positive definite diagonal matrix is a Stieltjes matrix, see Appendix A. Now we introduce a general result for induced norms based on Stieltjes matrices.

Proposition 3.2: Let A is a Stieltjes matrix then the solution to

$$\operatorname{argmin}_p \frac{1}{2} p^T A p \quad \text{s.t.} \quad \mathbb{1}^T p = 1 \tag{14}$$

is given by $p = \pi A^{-1}\mathbb{1}$; where the positive Largrange's multiplier $\pi = (\mathbb{1}^T A^{-1}\mathbb{1})^{-1}$ acts as a normalization constant. Moreover $p \succ 0$ (is a probability measure vector) and holds CCQ.

The proof of proposition 3.2 is presented in the Appendix A. Then, from this proposition, we can conclude that the QMMF data term preserve the order on the minimizer distributions ($p_i \geq p_j \iff \hat{v}_i \geq \hat{v}_j$) and hence is CCQ. Note that last result preserves for unnormalized likelihood ($p_i \geq p_j \iff v_i \geq v_j$); *i.e.* the QMMF models can directly use unnormalized v. Hence, in the case of Gaussian likelihood the model parameters (mean and covariance) can be computed with standard formulas just weighting with $p_k^2(r)$ the contribution of the r-th pixel value to the k-th class [1], [26].

C. Relationship with other information measures

For comparison purposes, we review three CCQ consistent information measures (the Kerridge's inaccuracy, the Q-disimilarity and the Euclidean distance) that guarantee CCQ, there are however important differences in the properties of the computed solution and algorithmic implications. In [1] is remarked that the second power ($\alpha = 2$) in (11) is justified by its numerical advantage: it leads to a quadratic programming problem. However, here we show that such a selection has beneficial implications on the solution p itself.

First we review the Kerridge's inaccuracy

$$K(p,q) = -\sum_k p_k \log q_k \tag{15}$$

is derived from (13) with $\alpha = \gamma = \delta = 1$ and $\beta \to 1$[29] [30].

Proposition 3.3: The solution to

$$\operatorname{argmin}_p K(p,q) \quad \text{s.t.} \quad \mathbb{1}^T p = 1, p \succeq 0 \tag{16}$$

is given by

$$p_i = \begin{cases} 1 & q_i \geq q_j,\, i \neq j, \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

Hence p is a indicator vector and holds CCQ.

The proof is presented in the Appendix A. This result can be contrasted with the corresponding for the Q–disimilarity: the Kerridge's inaccuracy results in a hard labeling (entropy zero), this is a disadvantage of the measure given the lack of information on the solution's confidence. In addition, the Euclidean distance

$$d(p,q) = \frac{1}{2}\|p - q\|^2, \tag{18}$$

Comunicaciones del CIMAT

TABLE I
DISSIMILARITIES BETWEEN THE DISTRIBUTIONS p AND q.

Name	Information Measure	Minimizer p given q	Preserve order	Gaussian model parameters	Optimization Problem
Kerridge	$-\sum_i p_i \log q_i$	$p_i = \begin{cases} 1 & q_i \geq q_j, i \neq j, \\ 0 & \text{otherwise.} \end{cases}$	No	Easily computable	Combinatorial optimization
Inner product	$1 - \sum_i p_i q_i$	$p_i = \begin{cases} 1 & q_i \geq q_j, i \neq j, \\ 0 & \text{otherwise.} \end{cases}$	No	Easily computable	Combinatorial optimization
Q-dissimilarity	$-\frac{1}{2}\sum_i p_i^2 \log q_i$	$p_i = \frac{(\log q_i)^{-1}}{\sum_j (\log q_j)^{-1}}$	Yes	Easily computable	Quadratic Programming
Euclidean	$\frac{1}{2}\sum_i (p_i - q_i)^2$	$p_i = q_i$	Yes	No appropriate	Quadratic minimization

base of the Gaussian Markov Measure Models (GMMFs), has the straightforward solution:

$$p = q. \tag{19}$$

Table I presents a resume of the discussed information measures. —we have included the Inner Product (second row) which solution, easily to verify, is also an indicator vector. We can see that both the the Q-dissimililarity and the Euclidean distance lead to Quadratic optimization problems. However, the Q-disimilarity is preferred over the Euclidean distance because experimentally has been demonstrated that produces results with lower entropy [1]. Moreover, in the case of the joint estimation of segmentation and distribution parameters, the parameters of Gaussian distributions can be computed with simple formulas: the mean corresponds to the mean of the p^2-weighted data. Differently, the use of the Euclidean distance results in a collapse to a single model [1], [26]..

IV. MINIMIZATION ALGORITHMS

We develop two efficient minimization algorithms for solving the QMMF's optimization problem. First, we presents a memory efficient algorithm that update the measure field p position–wise and vector–component–wise. Second we presents a faster algorithm with a vector-wise update scheme. Although both schemes are initially developed for the QMMFs convex case, we show that they can be adapted to the non-convex case using a subspace minimization strategy.

A. Memory Limited Gauss Seidel Scheme

If μ is chosen such that the energy (8) is kept convex then the computation of p consists in solving a linear system. This is stated in next proposition.

Proposition 4.1: (Convex QMMF) Let $U(p)$ be the energy function defined in (8) and assume $d_r \succeq 0$, then the solution to

$$\min_p U(p) \quad \text{s.t.} \;\; \mathbb{1}^T p(r) = 1, \quad \text{for} \;\; r \in \Omega$$

is a probability measure field: it holds $p_r \succeq 0$.

Proof: We present an algorithmic proof to this Proposition. The optimal solution satisfies the Karush-Kuhn-Tucker (KKT) conditions:

$$p_k(r) d_k(r) + \lambda \sum_{s \in \mathcal{N}_r} \left(p_k(r) - p_k(s)\right) w_{rs} = \pi(r) \tag{20}$$

$$\mathbb{1}^T p(r) = 1 \tag{21}$$

where π is the vector of Lagrange's multipliers. Note that the KKT conditions are a symmetric and positive definite linear system that can be solved with very efficient algorithms as Conjugate Gradient or Multigrid Gauss-Seidel (GS). In particular, a simple GS scheme results from integrating (20) w.r.t. k (i.e. by summing over k) and using (21):

$$\pi(r) = \frac{1}{K} p(r)^T d(r). \tag{22}$$

Conmunicaciones del CIMAT

Thus, from (20):

$$p_k(r) = b_k(r)\left[\bar{p}_k(r) + \pi(r)\right] \tag{23}$$

where we define:

$$\bar{p}_k(r) \stackrel{def}{=} \lambda \sum_{s \in \mathcal{N}_r} w_{rs} p_k(s) \tag{24}$$

and

$$b_k(r) \stackrel{def}{=} \frac{1}{d_k(r) + \lambda \sum_{y \in \mathcal{N}_r} w_{rs}}. \tag{25}$$

Eqs. (22) and (23) define a two steps iterative algorithm. Moreover, if (22) is substituted into (23), we can note that if an initial positive p is chosen, then the GS scheme (23) will produce a convergent nonnegative sequence. ■

One can see that the GS scheme, here proposed [Eqs. (22) and (23)], is simpler than the originally reported in [1]. In the non-convex QMMF case we can use the projection strategy for enforcing the non-negativity constraint. Then at each iteration, the projected p can be computed with

$$p_k(r) = \max\left\{0, b_k(r)[\bar{p}(r) + \pi(r)]\right\}. \tag{26}$$

In Addition, the GS scheme for the binary (two classes) segmentation can be simplified with the elimination of the variable p_2 (using $p_2 = 1 - p_1$). In such a case, the GS update formula is given by

$$p_1(r) = \frac{d_2(r) + \lambda \sum_{s \in \mathcal{N}_r} w_{rs} p_1(s)}{d_1(r) + d_2(r) + \lambda \sum_{s \in \mathcal{N}_r} w_{rs}}. \tag{27}$$

Once more a projection can be applied in the the non-convex case.

B. Vector-wise Gauss Seidel Scheme

Since the iterative update formula (23) [and its projected version(26)] requires of a reduced amount of memory, its is proper for processing large data, as video or tomographic images (MRI or TC volumes). On the other hand, we can improve the computational performance (convergence rate) with a extra memory requirement if, instead of update $p(r)$ component by component, we update the entire vector in a single step. First we write the KKT conditions (20) for the full vector $p(r)$:

$$D_r p(r) + \lambda \sum_{s \in \mathcal{N}_r} \left(p(r) - p(s)\right) w_{rs} \quad = \quad \pi(r)\mathbb{1} \tag{28}$$

Note that the KKT conditions (28) and (21) are yet a symmetric and positive definite linear system that can be solved with very efficient algorithms as Conjugate Gradient or Multigrid Gauss-Seidel (GS). Following a similar algebraic procedure than that the used in section IV-A we have the positive definite and diagonal dominant system:

$$H_r p(r) = \bar{p}_r \tag{29}$$

where $H_r = D_r + \Lambda_r - \hat{\mathbb{1}} d_r^T$, $\hat{\mathbb{1}} = \mathbb{1}/K$, we define the diagonal matrix $\Lambda_r \stackrel{def}{=} \left(\lambda \sum_{y \in \mathcal{N}_r} w_{rs}\right) I$ and the kth component of the vector $\bar{p}(r)$ is computed with (24). Since H_r is a Stieltjes matrix, then $H^{-1} \succ 0$ (has only positive elements). Thus, the iteration of

$$p(r) = H_r^{-1} \bar{p}(r), \forall r \tag{30}$$

keeps $p \succ 0$ if an initial p is positive. Moreover the inverse matrix H_r^{-1} can efficiently be computed with the Sherman-Morrison formula:

$$H_r^{-1} \quad = \quad B_r + \frac{B_r \hat{\mathbb{1}} d(r)^T B_r}{1 - \hat{\mathbb{1}}^T B_r d(r)} \tag{31}$$

where we define the diagonal matrix $B_r = \mathrm{diag}[\ b(r)]$ with the elements of $b(r)$ computed with (25). Then by using $B_r \hat{\mathbb{1}} = b(r)/K$ and the vector:

$$\tilde{d}(r)^T = \frac{1}{K - b(r)^T d(r)} d(r)^T B_r. \tag{32}$$

Comunicación del CIMAT

Algorithm 1 Convex QMMF

1: {Initialization}
   i. Let K be the number of classes, $\lambda \geq 0$ the regularization parameter, v the normalized likelihoods and w the intra-pixel affinity;
   ii. Set $d_k(r) = -\log \hat{v}_k(r)$;
   iii. Compute $b_k(r)$ with (25) ;
   iv. Compute $\tilde{d}(r)^T$ with (32);
   v. Initialize $p \succ 0$; {*e.g.* $p = \hat{v}$}
2: **repeat**
3: **for** all r **do**
4: Compute the elements of $\bar{p}(r)$ with (24);
5: Compute $c = \tilde{d}(r)^T \bar{p}(r)$;
6: Update $p_k(r) = b_k(r)\left[\bar{p}_k(r) + c\right]$ for $k = 1, 2, \ldots, K$.
7: **end for**
8: **until** convergence

Thus, the kth component of $p(r)$ in (30) is computed the simple formula:

$$p_k(r) = b_k(r)\left[\bar{p}_k(r) + \tilde{d}(r)^T \bar{p}(r)\right] \tag{33}$$

where the vectors b and $\tilde{d}$ can be pre–computed. This procedure is resumed in Algorithm 1.

The entropy of the solution p can be controlled be means of the μ parameter that penalize the Gini's (entropy) coefficient. A positive μ reduce the entropy but can result in a negative value of $d_k(r)$, see (9), and hence leads us to a nonconvex quadratic programing problem. In such a case, the solution to (7) can be computed by using the minimization strategy in Algorithm 1 combined with a subspace minimization strategy. First we solve for $p(r)$ the problem (7) neglecting the nonnegative constraints. The active set $\mathcal{A}_r$ at each pixel is estimated from the non-positive coefficients in $p(r)$. Then we refine the previous solution by fixing $p_i(r) = 0$ for $i \in \mathcal{A}_r$ and solve (7) for the remanning $p_i(r)$ with $i \notin \mathcal{A}_r$. If a new negative coefficient is computed, the active set $\mathcal{A}_r$ is updated and a new solution is computed. The partial solution after few subspace minimization (we used 2 recursions in our experiments) is used as starting point for a new Gauss-Seidel iteration. We present the procedure details in the Algorithm 2. Note that the subspace minimization, line 10, can be computed with the same algorithm, in a recursive procedure.

Algorithm 2 Non–Convex QMMF with subspace minimization

1: {Initialize as Algorithm 1.}
2: **repeat**
3: **for** all r **do**
4: Compute the elements of $\bar{p}(r)$ with (24);
5: Compute $c = \tilde{d}(r)^T \bar{p}(r)$;
6: Update $p_k(r) = b_k(r)\left[\bar{p}_k(r) + c\right]$ for $k = 1, 2, \ldots, K$.
7: Compute an estimate of the active set for $p(r)$: $\mathcal{A} = \{i : \tilde{p}_i \leq 0\}$;
8: Solve approximatelly (7) for $\tilde{p}_r$ fixing $\tilde{p}_{ir} = 0$ for $i \in \mathcal{A}$ and $p(s)$ for $s \neq r$.
9: Set $p(r) = \tilde{p}_r$.
10: **end for**
11: **until** convergence

V. Experiments

In this paper, we have mainly presented theoretical aspects of the QMMFs. However, we have also presented practical implication of the QMMFs. In following experiments we focus on demonstrate the method capabilities for:

Comunicaciones del CIMAT

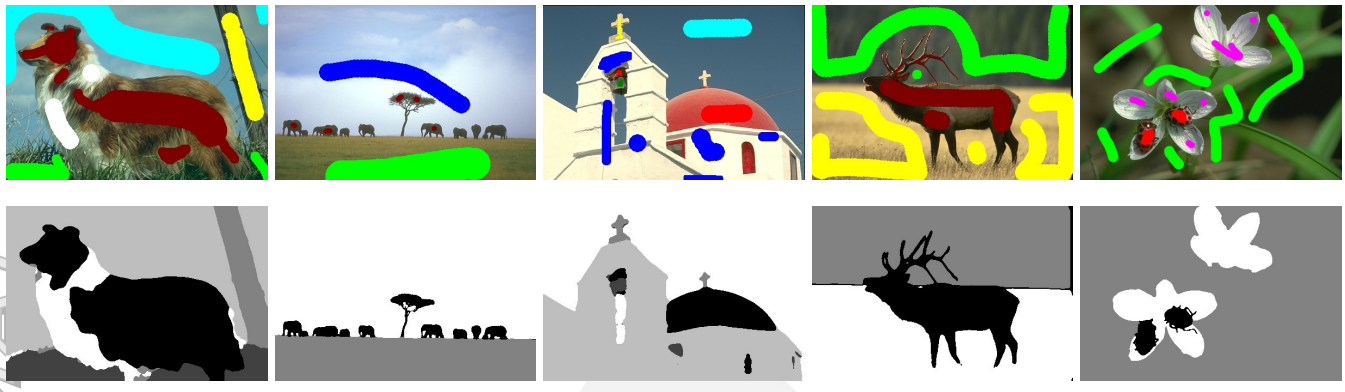


Fig. 3. Interactive multi–class segmentations.

a) Multi–class interactive image segmentation;
b) Efficient binary segmentation;
c) Simultaneous segmentation and parameter estimation.

A. Multiclass interactive segmentation

The image segmentation task is obviously a very ill-posed problem, i.e. there exists multiple"valid" segmentations for a particular image. User interaction is a popular form of introducing prior (high level) knowledge for segmenting images with complex scenes. In this paradigm the user labels by hand a subset of pixels and then the unknown labels are estimated with a segmentation algorithm that takes into account the distribution of the labelled pixels and the smoothness of the spatial segmentation. Interactive segmentation is a powerful technique that allows to develop general purpose tools. To illustrate this we can see three possible segmentations of the image in Fig. 1. The first column shows scribbles given by the user and the second column the corresponding segmentations computed with the method here presented. The rows correspond to segmentation by color, semantical objects (house, vegetation, fence, etc.) and planar regions, respectively.

In our multi-class interactive segmentation implementation, we assume that some pixels in the region of interest, $\mathcal{R}$, are correctly labelled by hand, thus we have a partially labelled field (*multimap*):

$$\mathcal{A}(r) \in \{0\} \cup \mathcal{K}, \ \forall r \in \Omega \tag{34}$$

where $\mathcal{A}(r) = k > 0$ indicates that the pixel r was assigned to the class k and $\mathcal{A}(r) = 0$ indicates that such a pixel class is unknown and needs to be estimated. Hence, by assuming that the user's labels are correct, then, in the data term in (8), the sum over all the pixels in the region of interest, $r \in \mathcal{R}$, is replaced by sum only over the unlabeled pixels: *i.e.* for $\{r : \mathcal{A}(r) = 0\}$. Alternatively, by leaving the sum for all pixels $r \in \mathcal{R}$ we assume uncertainty in the hand labeled data.

Let g an image such that $g(r) \in t$, with $t = \{t_1, t_2, \ldots, t_T\}$ the pixel values (maybe vectorial values as in the case of color images), then the density distribution for the classes are empirically estimated by using a histogram technique. That is, if H_{ki} is the number of hand labelled pixels with value t_i for the class k [24] then h is the smoothed histogram version. We implement the smoothing operator by a homogeneous diffusion process. Thus the normalized histograms are computed with $\hat{h}_{ki} = h_{ki}/\sum_l h_{kl}$ and the likelihood of the pixel r belonging to a given class k (likelihood function, LF) is computed with:

$$LF_{ki} = \frac{\hat{h}_{ki} + \epsilon}{\sum_j (\hat{h}_{ji} + \epsilon)}, \ \forall k; \tag{35}$$

with $\epsilon = 1 \times 10^{-8}$, a small constant. The ϵ constant introduces introduces an uniform distribution that avoids a possible division by zero and guarantee positive likelihoods. Thus the likelihood of an observed pixel value is computed with $v_k(r) = LF_{ki}$ such that $i = \arg\min_j \|g(r) - t_j\|^2$. Figure 3 shows multi–class interactive segmentations computed with the proposed algorithm implemented in Matlab (in *.m* and *.mex* files). Given the inter–pixel affinity w_{rs} can be understood as the likelihood that the pixels r and s belong to the same class, then we used $T\{g(r)\} = \hat{v}(r)$ in (10).

Comunicaciones del CIMAT

(a) Original (b) Trimap (c) GraphCut (d) QMMF+EC

Fig. 4. Segmentation example from the Lasso's data set.

TABLE II
CROSS-VALIDATION RESULTS: PARAMETERS, AKAIKE INFORMATION CRITERION, TRAINING AND TESTING ERROR.

Algorithm	Params.	AIC	Training	Testing
Graph Cut	λ, γ	8.58	6.82%	6.93%
Rand. Walk.	λ, γ	6.50	5.46%	5.50%
GMMF	λ, γ	6.49	5.46%	5.49%
QMMF	λ, γ	6.04	5.02%	5.15%
QMMF+EC	λ, γ, μ	3.58	3.13%	3.13%

B. Quantitative Comparisson: Image Binary Interactive Segmentation

Next we resume our results of a quantitative study on the performance of the segmentation algorithms: the proposed Binary variant of QMMF, the maximum flow (minimum graph cut, GC), GMMF and Random Walker (RW). The reader can find more details about this study in our technical report [25]. The task is to segment color images into background and foreground allowing interactive data labeling. The generalization capabilities of the methods are compared with a *cross-validation* procedure [31]. The comparison was conducted on the Lasso benchmark database [8]; a set of 50 images available online [32]. Such a database contains a natural image set with their corresponding trimaps and the ground truth segmentations. Actually, a Lasso trimap is an image of class labels: no–process mask ($\mathcal{M} = \mathcal{L}\backslash\mathcal{R}$), background, foreground and unknown; where no error is assumed in the initial labelled pixels. First column in Fig. 4 shows an image form the Lasso database and second column the corresponding trimap; the gray scale corresponds with the above class enumeration. In this case, the region to process is labeled as "unknown" and the boundary conditions are imposed by the foreground and background labeled regions.

We opted to compute the weights using the standard formula (10) (*i.e.* $T\{g(r)\} = g(r)$ in (10)), in order to focus our comparison on the data term of the different algorithms: QMMF, GC, GMMF and RW. In this task, empirical likelihoods are computed from the histogram of the hand labeled [10], the potential minimized was the inner product (see Table I) for both the data and regularization parameter.

The hyper parameters (λ, μ, γ) were trained by minimizing the mean of the segmentation error in the image set by using the Nelder and Mead simplex descent [33]. We implement a cross–validation procedure following the recommendation in Ref. [31] and split the data set into 5 groups, 10 images per set. Figure 4 shows an example of the segmented images. Table II shows the resume of the training and testing error and the Akaike information criterion (AIC) [31]. The AIC was computed for the optimized (trained) parameters with the 50 image in the database. Note that the AIC is consistent with the cross-validation results: the order of the method performance is preserved. Moreover the QMMF algorithm has the best performance in the group.

This automatic learning parameter process confirm that GMMF and RW as close variants have similar performance [25]. However, it produces two unexpected results:

i. Our GC based segmentation improves significantly the reported results in [8]. Indeed, our basic GC implementation of the method in [10] overcomes, significantly, the reported results with Likelihood Functions based on Gaussian Mixtures [8].
ii. Contra–intuitively the learned parameter μ (QMMF+EC) promotes large entropy. We believe that such results are because of: the lasso data set have a narrow band of unknown pixels, the trimaps are correct (they have not miss-labeled pixels) and the hand–segmentations (ground truth) favor smooth boundaries.

Comunicaciones del CIMAT

(a) QMMF $\mu = 10$. (b) $\mu = 0$. (c) $\mu = -123$.

(d) GMMF. (e) Rand. Walk. (f) GraphCut.

Fig. 5. First row, results computed with the proposed method with a) low-entropy, b) without entropy control and c) high entropy. Second row, results computed with state of the art methods .

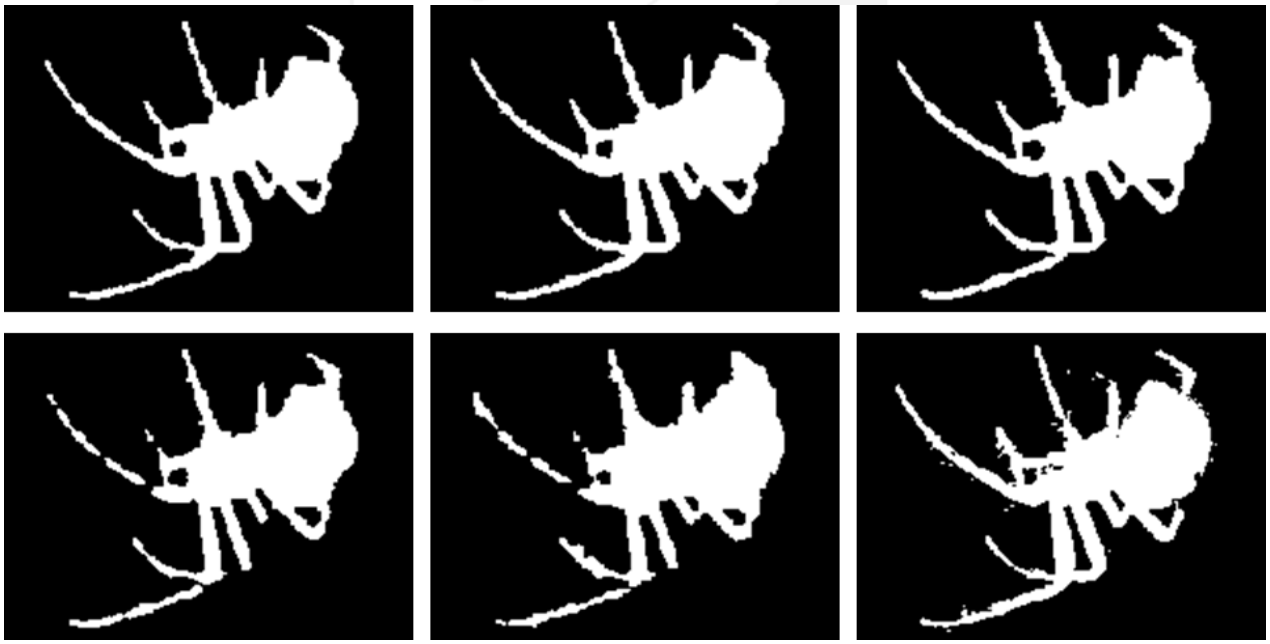


Fig. 6. Label maps corresponding to Fig. 5, same order.

An advantage of the entropy control is that allows one to adapt the algorithm for different tasks, for instance for the case of simultaneously estimation of segmentation and model parameters low entropy produces better results. The effect of the entropy control is illustrated in Figs. 5 and 6. The QMMF method algorithm produces, in all the cases, better segmentation with smooth boundaries than GMMF, RW and GC. In particular the matting factor shown Fig. 2 is computed with QMMF using $\mu = 0$.

Comunicaciones del CIMAT

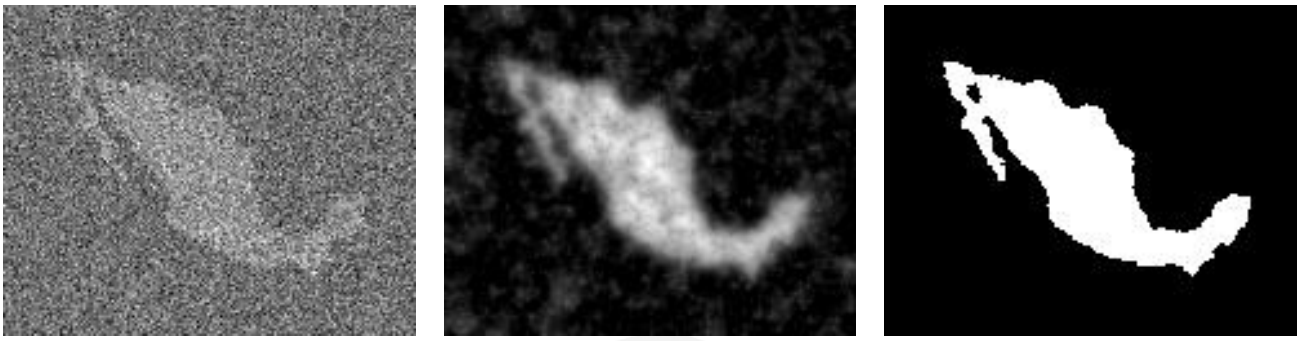


Fig. 7. Simultaneous segmentation and parameter estimation. From left to right: Noisy image (160×128 pixels) , computed p field and segmentation.

C. Robust model parameter estimation

Fig. 7 shows the segmentation of a corrupted binary image with white gaussian noise (means zero and std. dev equals 0.9 of the image dynamic range). These results were computed by estimating simultaneously the distribution parameters and the p field. Eight iterations of the two–steps scheme were required and the computation time was 0.2 sec., compare with results in [18]. On the other hand, our implementations based on Gauss Markov Measure Fields (GMMF, an early variant of Random Walker [7]) collapsed to a single model [5]. This limitation of the GMMF model is discussed in [34], see also [25]. Finally, Fig. 8 demonstrates generalization of the QMMFs for computing LF based on histogram techniques. The histograms are computed by p^2–weighting the pixel values, we initially set $p_k(r) = v_k(r), \forall k, r$. The erroneous segmentation at the first iteration is product of inaccurate scribbles and thus inaccurate initial LF (class histograms). The segmentation after two iteration demonstrates the ability of the QMMFs for estimating nonparametric class distributions.

VI. Conclusions and Discussion

We presented a derivation of the QMMF model independent of the minimal entropy constraint. Therefore, based on prior knowledge, we can control the amount of entropy increment, or decrement, in the computed probability measures. We demonstrated that the QMMF models are general and accept any marginal probability functions. As demonstration of such a generalization we presented experiments with iterative estimation of likelihood functions based on histogram techniques. We proposed robust likelihoods that improve the method performance for segmenting textured regions.

Our contributions in this work are mainly a more general derivation of the QMMF models and more efficient optimization algorithms. Along the paper we present a series of experiments for demonstrating our proposals. Additionally, we present an experimental comparison with respect algorithms of the state of the art. We selected the task of binary interactive segmentation for conducting our comparison, first because it demonstrates the use of the entropy control in the case of generic likelihood functions. Second, a benchmark database is online available, and finally our hyper–parameter training scheme demonstrates to be objective by, significantly, improving the previously reported results with a graph cut based method.

Appendix A

Definition (Stieltjes matrices [35]). A $K \times K$ Stieltjes matrix $A = (a_{ij})$ with $i, j = 1, 2, \ldots, K$ satisfies:

- is symmetric and positive definite;
- has positive diagonal elements, $a_{ii} > 0$;
- has nonpositive off-diagonal elements, $a_{ij} \leq 0,\ i \neq j$;
- its inverse $B = (b_{ij})$ is nonnegative, $b_{ij} > 0, \forall i, j = 1, 2, \ldots, K$.

Proof of Proposition 3.2: The KKT conditions of (14) are

$$Ap - \pi \mathbb{1} = 0 \tag{36}$$

$$\mathbb{1}^T p - 1 = 0, \tag{37}$$

Comunicaciones del CIMAT

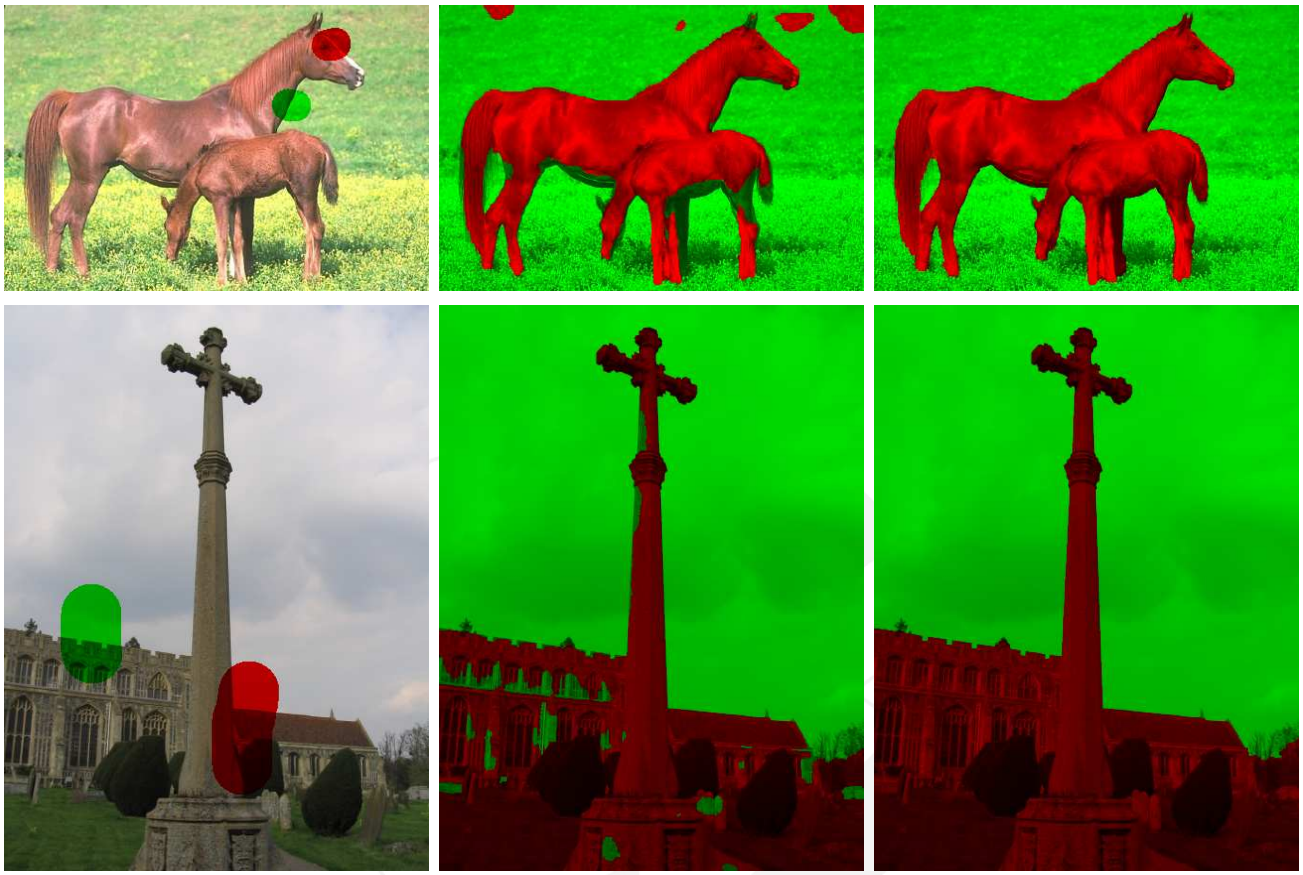


Fig. 8. Iterative estimation of empirical likelihood functions by histograms of p^2–weighted data. Binary segmentation: initial scribbles, first iteration and second iteration; respective columns.

where π is the Lagrange's multiplier. Then from (36):

$$p = \pi A^{-1}\mathbb{1}. \tag{38}$$

Substituting this result in (37), we have $\pi \mathbb{1}^T Q^{-1}\mathbb{1} = 1$, thus

$$\pi = \frac{1}{\mathbb{1}^T A^{-1}\mathbb{1}}. \tag{39}$$

We can conclude that $p \succeq 0$, since, A is a Stieltjes matrix, A^{-1} is positive and thus $A^{-1}\mathbb{1} > 0$ and $\mathbb{1}^T A^{-1}\mathbb{1} > 0$. ■

Proof of Proposition 3.3: The KKT conditions of (16) are

$$\begin{aligned} d - \pi\mathbb{1} &= s, && (40)\\ \mathbb{1}^T p - 1 &= 0, && (41)\\ p^T s &= 0, && (42)\\ p, s &\succeq 0. && (43) \end{aligned}$$

where $d_k \overset{def}{=} -\log q_k$, π is the Lagrange's multiplier of the equality constraint and s is the vector of Lagrange's multiplier of the nonnegativity constraints. Then from (40) and (42): $p^T(d - \pi\mathbb{1}) = 0$. Then it can be seen that the KKTs are fullfil with: $p = e_{k^*}$ with $k^* = \underset{k}{\operatorname{argmin}}\, d_k$, $\pi = d_{k^*}$ and s given by (40). ■

REFERENCES

[1] M. Rivera, O. Ocegueda, and J. L. Marroquin, "Entropy-controlled quadratic Markov measure field models for efficient image segmentation," *IEEE Trans. Image Processing*, vol. 8, no. 12, pp. 3047–3057, Dec. 2007.

Comunicaciones del CIMAT

[2] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. Springer-Verlag, Tokyo, 2001.

[3] J. Besag, "On the statistical analisys of dirty pictures," *J. R. Stat. Soc., Ser. B, Methodol.*, vol. 48, pp. 259–302, 1986.

[4] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images," *IEEE PAMI*, vol. 6, no. 6, pp. 721–741, 1984.

[5] J. L. Marroquin, F. Velazco, M. Rivera, and M. Nakamura, "Probabilistic solution of ill–posed problems in computational vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 337–348, 2001.

[6] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts," in *European Conference on Computer Vision (ECCV02).*, 2002.

[7] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, 2006.

[8] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr, "Interactive image segmentation using an adaptive GMMRF model," in *ECCV*, vol. 1, 2004, pp. 414–427.

[9] C. A. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," *IEEE Trans. Image Processing*, vol. 3, no. 2, pp. 162–177, 1994.

[10] Y. Boykov and M.-P. Jolly, "Interactive graph cut for optimal boundary & region segmentation of objects in N–D images," in *ICIP (1)*, 2001, pp. 105–112.

[11] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 11, pp. 1222–1239, 2001.

[12] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 721–741, 1984.

[13] O. Juan and R. Keriven, "Trimap segmentation for fast and user-friendly alpha matting," in *VLSM, LNCS 3752*, 2005, pp. 186–197.

[14] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross, and C. Rother, "Probabilistic fusion of stereo with color and contrast for bi-layer segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1480–1492, 2006.

[15] C. Rother, V. Kolmogorov, and A. Blake, "Interactive foreground extraction using iterated graph cuts," in *ACM Transactions on Graphics*, no. 23 (3), 2004, pp. 309–314.

[16] Y. Weiss, "Segmentation using eigenvectors: A unifying view," in *ICCV (2)*, 1999, pp. 975–982.

[17] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 8, pp. 888–905, 2000.

[18] C. Olsson, A. P. Eriksson, and F. Kahl, "Improved spectral relaxation methods for binary quadratic optimization problems," *Computer Vision and Image Understanding*, vol. 112, pp. 30–38, 2008.

[19] N. Komodakis, G. Tziritas, and N. Paragios, "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*, vol. 112, pp. 14–29, 2008.

[20] P. Kohli and P. H. S. Torr, "Measuring uncertainty in graph cut solutions," *Computer Vision and Image Understanding*, vol. 112, pp. 30–38, 2008.

[21] M. Rivera, O. Ocegueda, and J. L. Marroquin, "Entropy controlled Gauss-Markov random measure fields for early vision," in *VLSM*, vol. LNCS 3752, 2005, pp. 137–148.

[22] A. Levin, A. Rav-Acha, and D. Lischinski, "Spectral matting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1–14, 2008.

[23] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer Series in Operation Research, 2000.

[24] M. Rivera and P. P. Mayorga, "Quadratic markovian probability fields for image binary segmentation," in *in Proc. ICCV, Workshop ICV 07*, 2007, pp. 1–8.

[25] ——, "Comparative study on quadratic Markovian probability fields for image binary segmentation," CIMAT A.C., Mexico, Tech. Rep. 10.12.2007, I-07-15 (CC), December 2007.

[26] M. Rivera, O. Dalmau, and J. Tago, "Image segmentation by convex quadratic programming," in *Int. Conf. on Pattern Recognition (ICPR08)*, 2008.

[27] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models*, M. I. Jordan, Ed. Kluwer Academic Publishers, Boston MA., 1998, pp. 355–368.

[28] I. Taneja and H. C. Gupta, "On generalized measures of relative information and inaccuracy," *Aplikace Matematiky*, vol. 23, pp. 317–333, 1978.

[29] B. D. Sharma and R. Autar, "Relative information functions and their type (alpha, beta) generalizations," *Metrika*, vol. 21, no. 1, pp. 41–50, 1974.

[30] D. F. Kerridge, "Inaccuracy and inference," *J. Royal Stat. Soc, Ser. B*, vol. 23, no. 1, pp. 184–194, 1961.

[31] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*. Springer, 2001.

[32] http://research.microsoft.com/vision/cambridge/ i3l/segmentation/GrabCut.htm.

[33] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, pp. 308–313, 1965.

[34] J. L. Marroquin, B. C. Vemuri, S. Botello, F. Calderon, and A. Fernandez-Bouzas, "An accurate and efficient Bayesian method for automatic segmentation of brain MRI," *IEEE Trans. Medical Imaging*, vol. 21, pp. 934–945, 2002.

[35] R. S. Varga, *Matrix Iterative Analysis*, 2nd ed. Springer Series in Computational Mathematics, 2000, vol. 27.

Contiguaciones del CIMAT