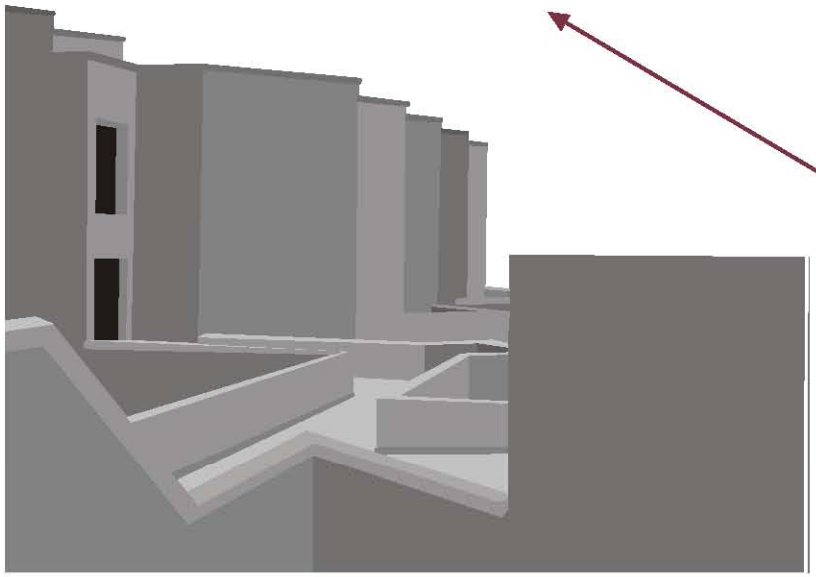


Comunicaciones del CIMAT

CONVEX QUADRATIC PROGRAMMING FOR IMAGE SEGMENTATION

Mariano Rivera and Oscar Dalmau

Comunicación del CIMAT No I-09-01/09-02-2009
(CC /CIMAT)

CIMAT

CONVEX QUADRATIC PROGRAMMING FOR IMAGE SEGMENTATION

MARIANO RIVERA AND OSCAR DALMAU

Abstract. A quadratic programming formulation for multiclass image segmentation is investigated. It is proved that, in the convex case, the global minima of Quadratic Markov Measure Field (QMMF) models holds the non-negativity constraint. This allows one to design efficient optimization algorithms. We also proposed a (free parameter) inter–pixel affinity measure more related with the classes memberships than with color or gray gradient based standard methods. Moreover, it is introduced a formulation for computing the pixel likelihoods by taking into account local context and texture properties. We demonstrate the QMMFs 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.
KEYWORDS: Image segmentation, Interactive segmentation, Quadratic energy function, Matting, Binary image segmentation, Markov random fields

1. 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]. Among many approaches, Markov random field (MRF) models based methods 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].

1.1. Markov random fields for Image segmentation.

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 shows

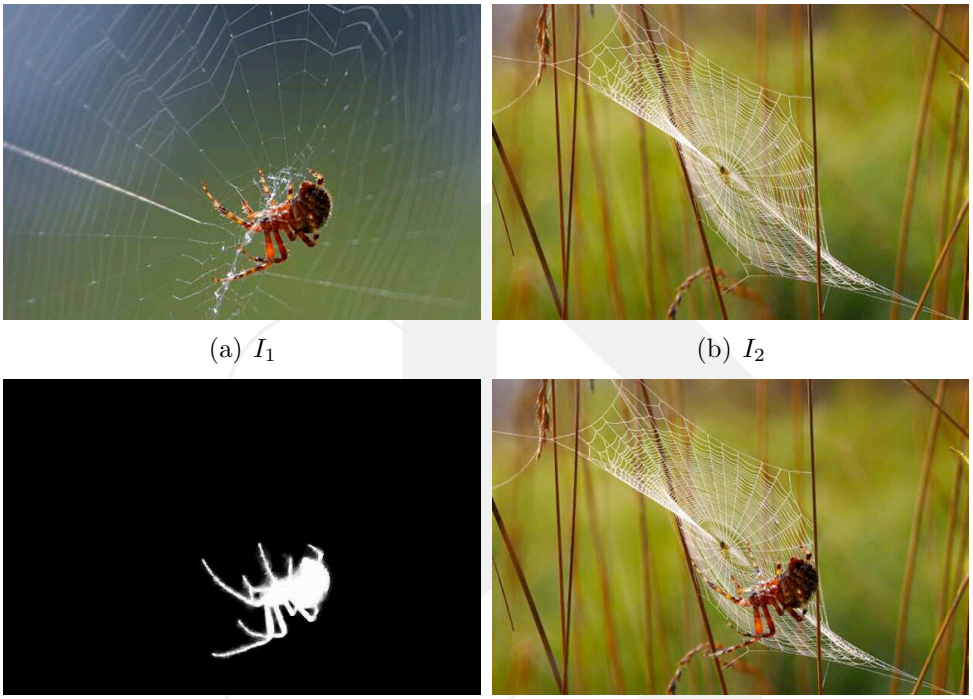
The authors are with the Centro de Investigacion en Matematicas AC, Guanajuato, GTO, Mexico 36000.
M. Rivera is also with the Department of Mathematics, Florida State University.
This work is supported in part by CONACYT (Grant 61367), Mexico.

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, graph-cut based techniques [8, 9, 10, 11, 12, 13, 14, 2, 15] and spectral methods [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, such a memberships can be expressed in a natural way in terms of probabilities—leading to the so named probabilistic segmentation (PS) methods.

In this work we present new insights and extensions to the recent reported PS method called Quadratic Markov Measure Fields models (QMMFs) [1]. In particular we investigate the convex (positive defined) and the binary (two classes) cases of the QMMFs. QMMFs are computationally efficient because they lead to the minimization of a quadratic energy function. Such a quadratic minimization is achieved by solving a linear system with a standard iterative algorithm as Gauss-Seidel (GS) or Conjugate Gradient (CG) [23]. As it is well known, the convergence ratio of such algorithms can be improved by providing a good initial guess (starting point)—a useful property in the case of dynamic models. Moreover gradient descent based algorithms (as GS or CG) produce a sequence of partial solutions that reduce successively the energy function. Thus, for applications with limited computational time, a good partial solution can be obtained by stopping the iterations even if the global optimum has not reached yet. These characteristics allow one to, naturally, implement computationally efficient multigrid algorithms [24].

1.2. **QMMF models notation.** Recently, in Ref. [1] was proposed the Entropy Controlled Quadratic Markov Measure Fields (EC-QMMF)



(a) I_1 (b) I_2

(c) α_1 (d) $g = \alpha_1 I_1 + \alpha_2 I_2$

FIGURE 1. Image model generation. I_1 and I_2 are the original data, α is a matting factor vector (with $\alpha_1+\alpha_2 = 1$) and g is the observed image.

models for image multiclass segmentation. Such models are computationally efficient and produce probabilistic segmentations of excellent quality.

Mathematically, let r be the pixel position in the image or the region of interest, $\mathcal{R} = \{r\}$ (in a regular lattice $\mathcal{L}$), $\mathcal{K} = \{1, \ldots, K\}$ the index set of known images, I_k, and $\mathbb{S}^K \subset \mathbb{R}^K$ the simplex such that

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

if and only if

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

$$z_k \geq 0, \quad \forall k \in \mathcal{K}; \tag{3}$$

where the vector $\mathbb{1} \in \mathbb{R}^K$ has all its entries equal one. Then the QMMF formulation is constructed on the assumption that the observed image g is generated with the model:

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

where η is a possible noise and $\alpha(r) \in \mathbb{S}^K$ is a matting factor that can be understood as a probability measure [1, 22]. Fig 1 illustrates the generation image process assuming model (4).

According to [1], an effective segmentation of the observed image, g, can be computed if the probabilities measures α are constrained to be as informative as possible, i.e. they have neglected entropy. Then the probabilistic segmentation by means of QMMF models consist on the solution of a quadratic programing problem of the form:

$$\arg\min_{\alpha \in \mathbb{S}^K} \quad \sum_r \psi(\alpha(r)) + \lambda \sum_r \sum_{s \in \mathcal{N}_r} w_{rs} \phi(\alpha(r), \alpha(s)) \tag{5}$$

where the potentials ψ and ϕ are quadratic and (for practical purposes) a first order neighborhood is used: $\mathcal{N}_r = \{s \in \mathcal{R} : \|r - s\| = 1\}$. The positive parameter λ controls the regularization (smoothness) process and the positive weights w lead the class border to coincide with large image gradients.

1.3. **Summary of contributions.** Our contributions in this paper are summarized as follows:

- We present a derivation of the QMMF model that relax significantly 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 demonstrate that the QMMF models are general and accept any marginal probability functions. independently of the entropy control.
- We note that the inter-pixel interactions, w_{rs} in (5), needs be understood as the probability that the neighbor pixels (r, s) belong to the same class. Thus, if the pixel values are assumed independent samples then the inter-pixel affinity is computed more accurately in the likelihood space than the image value space.
- Based on the independence assumption, we propose robust likelihoods that improve the method performance for segmenting textured regions.
- We objectively evaluate the methods performance by using a hyper-parameter training method based on cross-validation.
- We present a simpler and memory efficient algorithm for minimizing the quadratic energy functional.

Preliminary results of this work were in [25, 26, 27]. We organize this paper as follows. Section 2 presents the new derivation of the QMMF models; the convex and binary cases are studied in depth. That section also presents a new optimization procedure—simpler than the early reported. Section 3 presents extensions to the QMMF models. Experiments that demonstrate the method performance are presented in section 4. Finally, our conclusions are given in section 5.

2. Entropy–Controlled Quadratic Markov Measure Field Models

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.

2.1. **Posterior Probability.** For the purpose of this section, we assume that the images set $I = \{I_k\}, \forall k$, is either given or generated by a parametric model, $I_k(x) = \Phi_{\theta_k}(r)$, with known parameters $\theta = \{\theta_k\}, \forall k$. Such an assumption is equivalent to own a procedure for computing the marginal likelihood $v_k(r)$—the likeness between the observed pixel $g(r)$ and the model pixel $I_k(r)$ (for all pixel r and class k). In interactive approaches the marginal likelihoods are computed from empirical distribution [10, 8, 15, 13, 28, 25, 29, 30]. The computation of the α factor is the solution to the, ill–posed, inverse problem stated by model (4), subject to the constraints (2) and (3). In the Bayesian Regularization framework one computes the solution α^* as an estimator of the posterior distribution $P(\alpha|g, I)$. Given the dependency of I on θ we write indistinctly $P(\alpha|g, \Theta)$ or $P(\alpha|g, I)$; this will be explained in detail in next subsection. In terms of the Bayes' rule, such a posterior distribution is expressed as:

$$P(\alpha|g, I) = \frac{1}{Z} P(g|\alpha, I) P(\alpha, I); \tag{6}$$

where $Z = P(g)$ is a normalization constant (independent on α), $P(g|\alpha, I)$ is the Likelihood (conditional probability) of the data by assuming given (α, I), $P(\alpha, I)$ is the joint prior distribution of the unknowns α and the image set I (or parameters θ). This prior is the Bayesian mechanism for leading the solution to have known properties. Next subsections are devoted to derive the terms on the left side of (6).

2.2. **Likelihood.** The here presented development is constructed on the Gaussian distribution assumption. Posteriorly, in subsection 3.1, we show that the presented results can be generalized to other distributions than the Gaussian.

From the generation model (4), and by assuming i.i.d. Gaussian independent noise $\eta \sim \mathcal{N}(0, \sigma^2)$, the likelihood of α is given by

$$P(g|\alpha, I) \propto \prod_r \exp\left[-\frac{1}{\sigma^2}\left[d^T(r)\alpha(r)\right]^2\right], \tag{7}$$

where we define the residual

$$d_k(r) \stackrel{def}{=} g(r) - I_k(r). \tag{8}$$

We note however that:

Proposition 2.1. *In general, the maximum likelihood (ML) estimator of the posterior probability* (7) *,*

$$\tilde{\alpha} = \arg\max_{\alpha \in \mathbb{S}^K} P(g|\alpha, I),$$

is not unique.

Proof. Without loss of generality we assume $d_k(x) \in \mathbb{R}^1$, then the non–uniqueness is easy demonstrated from the fact that

$$\text{rank}(d(r)d^T(r)) = 1.$$

Thus $\tilde{\alpha}(r)$ lays in the intersection of $\mathbb{S}^K$ and the Null space of $d(r)d^T(r)$. □

Now we introduce the follows definition:

Definition The marginal likelihood $v_k(r)$ 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:

$$v_k(r) \stackrel{def}{=} P(g(r)|\alpha(r) = e_k, I), \tag{9}$$

where e_k is the kth canonical basis vector.

In the particular case of *i.i.d* Gaussian Noise:

$$v_k(r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{d_k^2(r)}{2\sigma^2}\right]. \tag{10}$$

Marginal likelihoods are of particular interest in image PS approaches. We frame PS methods in the next definition.

Definition *Consistence Condition Qualification* (CCQ). Let $v(r)$ be the marginal likelihood vector at pixel r, then, in absence of prior knowledge (uniform prior distributions), PS procedures compute a probability measure field p, with $p(r) \in \mathbb{S}^K, \forall r$, that satisfies:

$$\max_k p_k(r) = \max_k v_k(r), \forall r. \tag{11}$$

If a vector p holds (11) we say that p is CCQ. Note that if p is CCQ w.r.t. $v(r)$, then it is also CCQ w.r.t. the pixel-wise normalized marginal likelihood $\hat{v}_k(r) = v_k(r)/(\mathbb{1}^T v(r))$.

Proposition 2.2. *In general, the ML estimators of* (7) *are not CCQ.*

Rather to give a rigorous proof to Proposition (2.2), we present an informal counterexample to its contradiction: if p is a ML estimator of (7), then p is CCQ. Suppose that the observed pixel has a particular value, says $g = 2$, generated with $g = (\alpha^*)^T m$; where $m = (1, 2, 3)^T$ and $\alpha^* = (0, 1, 0)^T$ is CCQ. Then, for instance, the vector $\tilde{\alpha} = (1/2, 0, 1/2)^T$ is also an ML estimator. However $\tilde{\alpha}$ is not CCQ. Thus the contradiction is false.

In last example, we can note that α^* is more compact (has the fewer coefficient different from zero) than any other ML estimator. Our intuition says that, for explaining the data, we should prefer simple models over complex ones: the parsimony principle. Since α can be seen as a discrete distribution, compact representations have smallest entropy. This discussion will be useful for constraining the ML estimators of (7) for being CCQ. Following we present and discuss three candidate constraints:

i.: *Zero entropy* (maxima information):

$$\alpha_k(r)\alpha_l(r) = 0, \ \ \forall r, \forall k \neq j. \tag{12}$$

This constraint (entropy equal zero) implies $\alpha_k(r) \in \{0, 1\}$ and results in a hard segmentation approach. The original QMMF formulation is constructed on this constraint despite computes a soft–segmentation [1].

ii.: *Zero entropy at imperfectly explained data*:

$$\alpha_k(r)\alpha_l(r)\left[1 - v_k(r)\right]\left[1 - v_l(r)\right] = 0, \ \ \forall r, \forall k \neq l. \tag{13}$$

Imperfect explained pixels are those without a model with marginal likelihood equal one. Thus, if we have a pixel r such that $[1 - v_k(r)]\,[1 - v_l(r)] > 0$ (for $k \neq l$) then the entropy at such a pixel is enforced to be equal zero, i.e. $\alpha_k(r)\alpha_l(r) = 0$, for $k \neq l$. Note that $1 - v_k(r)$ can be replaced by $|d_k(r)|$, where $d_k^2(r) \stackrel{def}{=} -\log v_k(r)$, and (because of the equality constraint)

the absolute value can also be neglected. Thus constraint (13) can be rewrite as:

$$\alpha_k(r)\alpha_l(r)d_k(r)d_l(r) = 0, \quad \forall r, \forall j \neq k. \tag{14}$$

iii.: Last constraints are too restrictive. Thus we proposed to relax constraint (14) by introducing the expected value at each pixel:

$$\sum_k \sum_{l \neq k} [\alpha_k(r)\alpha_l(r)d_k(r)d_l(r)] = 0, \quad \forall r. \tag{15}$$

We establish that if any constraint among (12)–(15) is enforced then the corresponding ML estimator is unique and is CCQ. Note that constraints (12)–(14) necessarily imply constraint (15). Thus, without loss of generality, the next theorem is written in terms of the less restrictive constraint (15).

Theorem 2.3. *If $d_k(r) > 0, \forall k, r$ then the constrained ML estimator computed as the solution to*

$$\begin{aligned} \arg\max_{\alpha \in \mathbb{R}^K} \quad & P(g|\alpha, I) \\ s.t. \quad & \\ & \mathbb{1}^T \alpha(r) = 1, \forall r \\ & \sum_k \sum_{l \neq k} [\alpha_k(r)\alpha_l(r)d_k(r)d_l(r)] = 0, \forall r \end{aligned} \tag{16}$$

is a probability measure field and holds CCQ

The proof, presented later, relies on Theorem 2.5. First, we present the definition of Stieltjes matrices and an important property [31].

Definition A $K \times K$ Stieltjes matrix $Q = (q_{ij})$ with $i, j = 1, 2, \ldots, K$ satisfies:

- is symmetric and positive definite;
- has positive diagonal elements, $q_{ii} > 0$;
- has nonpositive off-diagonal elements, $q_{ij} \leq 0, \ i \neq j$.

A well known property of Stieltjes matrices is the following [31]:

Proposition 2.4. *Let Q be a $K \times K$ Stieltjes matrix, then its inverse matrix $Q^{-1} = (\tilde{q}_{ij})$ is nonnegative: $\tilde{q}_{ij} > 0, \forall i, j = 1, 2, \ldots, K$.*

Next Theorem present an important property of some quadratic programing problems based on Stieltjes matrices.

Theorem 2.5. *Let Q be a Stieltjes matrix, then the minimizer of*

$$\min_{x\in\mathbb{R}^K} \quad \frac{1}{2}x^TQx \quad s.t. \quad \mathbb{1}^Tx = 1 \tag{17}$$

holds $x > 0$.

Proof. The Karush-Kuhn-Tucker (KKT) conditions of (17) are

$$Qx - \pi\mathbb{1} = 0 \tag{18}$$

$$\mathbb{1}^Tx - 1 = 0, \tag{19}$$

where π is the Lagrange's multiplier. Then from (18):

$$x = \pi Q^{-1}\mathbb{1}. \tag{20}$$

Substituting this result in (19), we have $\pi\mathbb{1}^TQ^{-1}\mathbb{1} = 1$, thus

$$\pi = \frac{1}{\mathbb{1}^TQ^{-1}\mathbb{1}} \tag{21}$$

and using this formula into (20):

$$x = \frac{Q^{-1}\mathbb{1}}{\mathbb{1}^TQ^{-1}\mathbb{1}}. \tag{22}$$

We can conclude that $x > 0$, since, from Proposition 2.4, Q^{-1} is positive and thus its sums by row and over all its entries are positive; $Q^{-1}\mathbb{1} > 0$ and $\mathbb{1}^TQ^{-1}\mathbb{1} > 0$, respectively. □

In last proof we can note that the Lagrange multiplier is nonnegative [see (21)], hence we generalize last results in the next Corollary.

Corollary 2.6. *Let Q be a Stieltjes matrix, then the minimizer of $\min_{x\in\mathbb{R}^K} \frac{1}{2}x^TQx$ subject to $\mathbb{1}^Tx \geq 1$ holds: $x > 0$ and $\mathbb{1}^Tx = 1$.*

Now we are in position to present the proof to Theorem 2.3.

Proof. (Theorem 2.3) If (15) is enforced, we have

$$\left[\alpha^T(r)d(r)\right]^2 = \sum_k \left[\alpha_k(r)d_k(r)\right]^2.$$

Thus, because of Theorem 2.5, the ML estimator is a probability measure field. Such an ML estimator is the optimizer (for all r) of the Lagrangian:

$$\min_{\alpha}\max_{\pi} \frac{1}{2}\sum_k \alpha_k^2(r)d_k^2(r) - \pi(r)\left[\mathbb{1}^T\alpha(r) - 1\right] \tag{23}$$

where π are the Lagrange's multipliers. Since the ML estimator satisfies the KKT conditions:

$$\begin{aligned} \alpha_k(r) d_k^2(r) &= \pi(r), \forall k \\ \mathbb{1}^T \alpha(r) &= 1, \end{aligned}$$

we have $\alpha_k(r) = \pi(r)/d_k^2(r) = \pi(r)/[-\log v_k(r)]$. Given that $\pi(r)$ is a positive scalar [from (21)], we have that $\alpha_k(r)$ is maxima for $v_k(r)$ maxima; thus the ML estimator is CCQ. □

Now that we have constrained the ML estimators to be CCQ, we present prior distributions for introducing special characteristics in the segmentation. Our general prior assumes independency between mattings α and images I, and a uniform distribution on the images; then the prior probabilities are of the form:

$$P(\alpha, I) \propto P_s(\alpha) P_h(\alpha)$$

where the priors P_s and P_h introduce, respectively, an explicit control on the smoothness and on the entropy of the probability measures $\alpha(r)\ \forall r$ [1]. Next we discuss these particular priors.

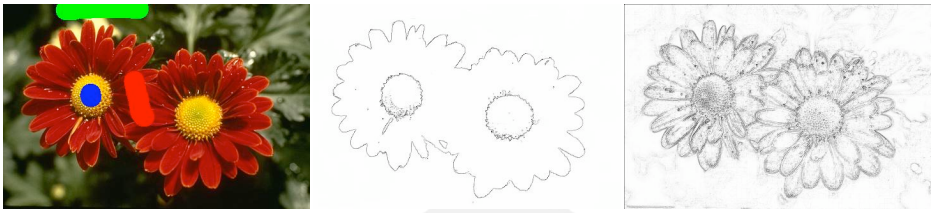
2.3. **Smoothness control prior P_s.** The intra–region smoothness is promoted by imposing a Gibbsian distribution, based on MRF models, that reduces the granularity of the regions. A popular prior form is the first order potential [11, 32, 33, 8, 28, 34, 1]:

$$P_s(\alpha) = \frac{1}{Z_c} \exp\left[-\frac{\lambda}{2} \sum_r \sum_{s \in \mathcal{N}_r} \|\alpha(r) - \alpha(s)\|^2 w_{rs} \right]; \tag{24}$$

where Z_c is a normalization constant and λ is a positive parameter that controls the smoothness amount. The positive weights w are of special interest, they should be chosen such that $w_{rs} \approx 1$ if the neighboring pixels r and s are likely to belong to the same class and $w_{rs} \approx 0$ in the opposite case. In the literature are commonly reported wights that depend on the magnitude of the gradient. In the task of color image segmentation, an instance of such weight–functions is [10, 35]:

$$w_{rs} = \frac{\gamma}{\gamma + \|\mathrm{Lab}(g(r)) - \mathrm{Lab}(g(s)\|^2}, \tag{25}$$

where $\mathrm{Lab}(\cdot)$ is an operator that transforms a vector in the RGB space into the Lab space. The use of the Lab-space based distance is, in that context, motivated by its close relationship with the distance of color human perception. However the Lab-space distance (as the color human perceptual distance) hardly represents the inter–class (objects) distances. Inter-class distances are context and task dependent. For



(a) Scribbles (b) Likelihood based (c) Color based

FIGURE 2. Interpixel affinity, w_{rs}.

instance, if the task is to segment the image in Fig. 2 into flowers and foliage, the weights should be close to zero (black) at the petals borders. Other possibility is to segment the image into three classes: petals, foliage and headflowers; in such a case Fig. 2b shows a proper weight map. As can be noted on Fig. 2c, the weights based on gradient magnitude may not represents the intra-classes edges. Here, we propose a new inter–pixel affinity measure based on the marginal likelihoods and thus incorporates, implicitly, the non-euclidean distances of the feature space. Our weight proposal is

$$w_{rs} = \frac{v^T(r)v(x)}{\|v(r)\|\|v(s)\|}. \tag{26}$$

Although other variants need be investigated, here we remark that:

Remark The intra–pixel affinity w_{rs} can be understood as the probability that the neighbor pixels r, s belongs to the same class. Assuming i.i.d. samples, such a probability can be approached more precisely from likelihood vectors than directly from the observed data.

Fig. 2 compares our likelihood based weights versus gradient based ones. Panel 2a shows scribbles for three classes: petals, foliage and headflowers. Panel 2b shows the weights based on the likelihood based (26) and 2c shows the standard weights based on the image gradient (25).

2.4. **Entropy control prior P_h.** Entropy control steers the sharpness of the probability measure vector at each pixel. In general, the entropy control prior is of the form:

$$P_h(\alpha) \propto \prod_r \exp\left[-\mu H\left(\alpha(r)\right)\right]$$

where μ is a parameters that promotes entropy increment (if $\mu < 0$) or decrement (if $\mu > 0$) and $H(z)$ is an entropy measure of the discrete distribution z [36]. In order to keep quadratic the potential, in [1] is

used the Gini's potential as entropy measure. The Gini's potential can be seen the negative variance of the z values: $H(z) = \mathbb{E}^2(z) - \mathbb{E}(z^2)$. Since $\mathbb{E}(\alpha(r)) = 1/K$, one can neglect such an α-independent term:

$$H(\alpha(r)) = -\sum_k \alpha_k^2(r).$$

2.5. **QMMF posterior energy and Minimization Algorithm.** The posterior distribution of the matting factor α is of the form:

$$P(\alpha|g, I) \propto \exp\left[-U(\alpha)\right].$$

Then

Proposition 2.7. *The constrained MAP estimator is computed by solving the quadratic programing problem:*

$$\min_{\alpha} U(\alpha) \quad s.t. \;\; \alpha_k(r) \in \mathbb{S}^K, \quad for \;\; r \in \Omega.$$

where the posterior energy is defined as

$$\begin{aligned} U(\alpha) \;\; = \;\; & \sum_{r\in\mathcal{R}} \Big\{ \sum_k \alpha_k^2(r) \left[d_k^2(r) - \mu \right] \\ & + \frac{\lambda}{2} \sum_{y\in\mathcal{N}_r} \|\alpha(r) - \alpha(s)\|^2 w_{rs} \Big\}. \end{aligned} \tag{27}$$

In addition, if μ is chosen such that the energy $U(\alpha)$ is kept convex (i.e. $d_k^2(r) - \mu > 0, \; \forall k, r$), then the non-negativity constraints are inactive at the global optimal solution. In such a case, the non-negativity constraints are neglected and thus the optimization procedure can be achieved with simple and efficient minimization procedures for convex quadratic minimization. This is stated in next theorem.

Theorem 2.8. (Convex QMMF) *Let $U(\alpha)$ be the energy function defined in* (27) *and assuming*

$$l_k(r;\mu) \stackrel{def}{=} -\log v_k(r) - \mu > 0, \;\; \forall k, r; \tag{28}$$

then the solution to

$$\min_{\alpha} \frac{1}{2} U(\alpha) \quad s.t. \;\; \mathbb{1}^T \alpha(r) = 1, \quad for \;\; r \in \Omega$$

is a probability measure field.

Proof. We present an algorithmic proof to this Theorem. The optimum solution holds the KKT conditions:

$$\alpha_k(r) l_k(r;\mu) + \lambda \sum_{s\in\mathcal{N}_r} \left(\alpha_k(r) - \alpha_k(s)\right) w_{rs} = \pi(r) \tag{29}$$

$$\mathbb{1}^T \alpha(r) = 1 \tag{30}$$

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 of integrating (29) w.r.t. k (i.e. by summing over k) and using (30):

$$\pi(r) = \frac{1}{K}\alpha^T(r) l(r;\mu). \tag{31}$$

Thus, from (29):

$$\alpha_k(r) = \frac{a_k(\alpha, r) + \pi(r)}{b_k(r)} \tag{32}$$

where we define: $a_k(\alpha, r) \stackrel{def}{=} \lambda \sum_{s\in\mathcal{N}_r} w_{rs}\alpha_k(s)$ and $b_k(r) \stackrel{def}{=} l_k(r;\mu) + \lambda \sum_{y\in\mathcal{N}_r} w_{rs}$. Eqs. (31) and (32) define a two steps iterative algorithm. Moreover, if (31) is substituted into (32), we can note that if an initial positive guess for α is chosen, then the GS scheme (32) will produce a convergent nonnegative sequence. □

One can see that the GS scheme, here proposed [Eqs. (31) and (32)], is simpler than the originally reported in [1]. In the non-convex QMMF case we can use the projection strategy. Then at each iteration, the projected α can be computed with

$$\alpha_k(r) = \max\left\{0, \frac{a_k(\alpha, r) + \pi(r)}{b_k(r)}\right\}. \tag{33}$$

2.6. **Binary Segmentation.** The binary case (segmentation in two classes) is of particular interest given that many problems in computer vision, image processing and image analysis require of segmenting the image into two clases

Theorem 2.9. Convex Quadratic Markov Probability Field. *Let μ be chosen such that $l_k(r;\mu) > 0$ for $k = 1, 2$ and $\lambda > 0$, then the unconstrained minimizer α^* of the energy functional*

$$\begin{aligned} B(\alpha) &= \sum_r \Big\{ \alpha^2(r) l_1(r;\mu) + [1-\alpha(r)]^2 l_2(r;\mu) \\ &\quad + \frac{\lambda}{2} \sum_{y\in\mathcal{N}_r} [\alpha(r) - \alpha(s)]^2 w_{rs} \Big\} \end{aligned} \tag{34}$$

is a probability field, i.e. $\alpha^*(r) \in [0, 1]$.

By defining $\alpha = \alpha_1$ and substituting $\alpha_2 = 1 - \alpha$ in the energy $U(\alpha)$ then the proof is straightforward from Theorem 2.8 . For this convex binary case the GS scheme is given by:

$$\alpha(r) = \frac{l_2(r) + \lambda \sum_{s \in \mathcal{N}_r} w_{rs} \alpha(s)}{l_1(r) + l_2(r) + \lambda \sum_{s \in \mathcal{N}_r} w_{rs}}. \tag{35}$$

and in the non-convex binary case the projection strategy can also be used.

3. Generalizations

In this section we extend the presented QMMF formulation: other likelihood functions than the Gaussian, local structure information and model parameter estimation.

3.1. Non–Gaussian Likelihood Functions.

In last section we derive the QMMF formulation by assuming that the image noise η is i.i.d. Gaussian. Although the Gaussianity assumption is supported by the central limit theorem, the use of the exact data distribution improves significantly the accuracy of the estimation. This is the case of multimode distributions as the ones empirically estimated from scribbles in an interactive segmentation approach. For removing the Gaussian noise assumption, we first note that

Proposition 3.1. *Let v_k be the smooth density distribution of the pixel values, $g(r), \forall r$; then it can be expressed with a Gaussian mixture model* [36]*:*

$$v_k(r) = \sum_{i=1}^{M} \pi_{ki} G_{ki}(r), \tag{36}$$

with $G_{ki}(r) \stackrel{def}{=} 1/\sqrt{2\pi}\sigma_k \exp\left[-d_{ki}^2(r)/2\sigma_k^2\right]$, where we defined $d_{ki}(r) \stackrel{def}{=} g(r) - m_{ki}$. The known parameters are denoted by $\theta_k = (\sigma_k, \pi_k, m_k)$; where $\pi_k \in \mathbb{S}^K$ is the mixture coefficients vector, $m_k = (m_{k1}, m_{k2}, \ldots, m_{kM})$ are the Gaussians' means, σ_k are the variances and M is the number (maybe very large) of Gaussians.

Thus, if we assume that such a mixture is composed by a large summation of narrow Gaussians, we can state the next theorem that generalize of the QMMF approach for other distribution than the Gaussian:

Theorem 3.2. *Let $v = \{v_k\}$, for $k \in \mathcal{K}$, be the smooth density distributions of the corresponding $I = \{I_k\}$; then the likelihood of the observed*

image g is given by

$$P(g|\alpha, I) = \prod_k \prod_x \Big[v_k(r) \Big]^{\alpha_k^2(r)}, \tag{37}$$

if constraint (15) *is provided.*

The proof is presented in Appendix A.

In Section 2 we derive the QMMF formulation by assuming Gaussian marginal likelihoods. Now, by using theorem 3.2, we generalize the QMMF formulation as follows.

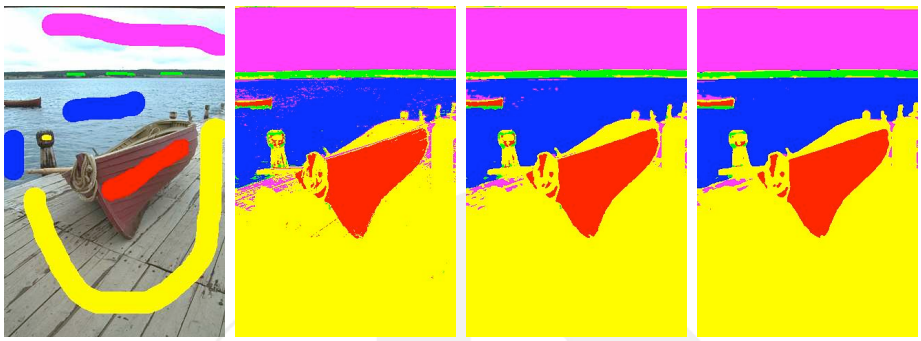
Proposition 3.3. *Theorem 2.3 holds for other marginal likelihoods, v, than the Gaussian. Consequently the QMMF model can be used independently of the particular form of v.*

As immediate consequence we have the next corollary.

Corollary 3.4. *The QMMF model is independent of the particular form of the negative log marginal likelihood (distance), $d_k^2(r) \stackrel{def}{=} [-\log v_k(r)]$.*

For demonstrating last corollary, we show in the experiments a wide use of empirical marginal likelihood functions computed with histogram methods.

3.2. **Local structured data.** Frequently, the pixel color coordinates (given in a color space) are used as features for color based segmentation. In contrast, in texture based segmentation the features result of applying a set of operators that take into account the neighborhood of the pixel. Instances of such operators are Gabors filters [37] and statistical moments [38]. The underlaying idea in statistical moments is to consider the pixel neighborhood as samples of a (unknown) distribution determined by some statistics: variance, skewness, kurtosis, etc. Such statistics are then arrayed as a feature vector [39]. The feature vector length depends on the number of operators used and of the dimension reduction technique used [40, 41]. If the generative distributions are known (or estimated in a two step algorithm, see next subsection 3.3), then we can use an efficient procedure for introducing texture information without the explicitly computation of texture features. Similar to statistical moments, we suppose that textured regions are generated with i.i.d. random samples of particular distributions. Thus, the contextualized marginal likelihood, V, at a pixel r should consider its neighborhood pixels that are, almost in all the cases, generated with



(a) Scribbles (b) $\rho = 0$ (c) $\rho = 1$ (d) $\rho = 2$

FIGURE 3. ML estimator for different Neighborhood sizes.

the same distribution:

$$V_k(r) \propto \prod_{y \in \mathcal{M}_r} v_k(s) = \exp\left(\sum_{s \in \mathcal{M}_r} \log v_k(s) \right), \tag{38}$$

where $\mathcal{M}_r = \{s : |r - s| \leq \rho\}$ is the neighborhood of r and the parameter ρ define the neighborhood size. Figure 3 shows Maximum Likelihood maps of (38) using different neighborhood sizes, ρ. Note that for large ρ–values the Maximum Likelihood (ML) estimator has a reduced granularity but at the same time small details are lost. Moreover, according with Corollary 3.4, the marginal likelihoods $\{V_k\}$ are compatible with our QMMF model.

3.3. **Parameter estimation of generative models.** In Ref. [1] was studied the particular case estimating the mean of Gaussian Likelihood functions. Now, we presents the generalization of the Gaussian parameter estimation presented in [1] to both parametric and nonparametric models.

In the derivation of the QMMF model we have assumed that the image set $I = \{I_k\}$ is given and thus the marginal distributions $\{v_k\}$. In a generative approach such an assumption is equivalent to suppose known the noise, η, distribution and the parameters, $\theta = \{\theta_k\}$, of the generative modelsΦ , where $I_k(r) = \Phi(\theta_k, r)$, see model (4). From the posterior distribution, $P(\alpha,\theta\,|g) = \frac{1}{Z} P(g|\alpha,\theta\,) P(\alpha,\theta\,)$, we estimate both: memberships (α) and parameters (θ) by alternating partial MAP estimations:

(1) $\max_\alpha P(\alpha,\theta\,|g)$ keeping fixed θ,
(2) $\max_\theta P(\theta,\alpha\,|g)$ keeping fixed α;

until convergence. These minimization can be partially achieved as in a generalize EM scheme [42]. The maximization we concern in this subsection is the one in the second step. This maximization is achieved by minimizing the negative posterior energy $D(\alpha,\theta\,) \stackrel{def}{=} -\log P(\alpha,\theta\,|g)$:

$$D(\alpha,\theta\,) = \sum_r \sum_k \alpha_k^2(r)(-\log v_k(r)), \tag{39}$$

where we neglect constant terms on θ. Thus keeping α fixed, the parameters are computed by solving the system that results of $\nabla_\theta D(\alpha,\theta\,) = 0$, where:

$$\nabla_{\theta_k} D(\alpha,\theta\,) \;=\; \sum_r \left[\frac{-\alpha_k^2(r)}{v_k(r)}\right] \nabla_{\theta_k} v_k(r); \tag{40}$$

where ∇_θ denotes the partial gradient w.r.t. θ.

In the case of Gaussian Likelihood functions the update formulas have a simple form.

Proposition 3.5. *If the marginal likelihoods are Gaussians of the form*

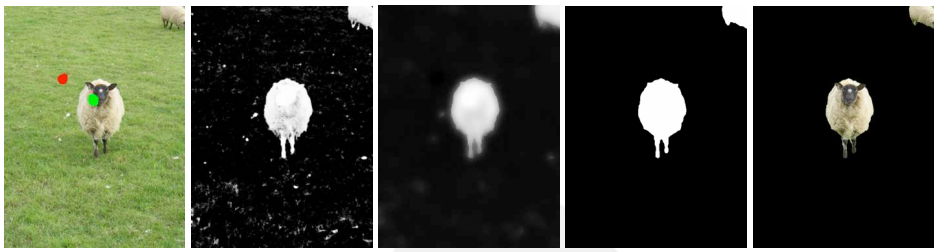
$$v_k(r) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left[-\frac{1}{2\sigma_k^2} d_k^2(r)\right] \tag{41}$$

with $d_k(r) = g(r) - m_k$, then parameter estimation step is computed with the formulas:

$$m_k \;=\; \frac{\sum_r \alpha_k^2(r) g(r)}{\sum_x \alpha_k^2(r)} \tag{42}$$

$$\sigma_k^2 \;=\; \frac{\sum_r \alpha_k^2(r)(g(r) - m_k)^2}{\sum_r \alpha_k^2(r)}. \tag{43}$$

The proof is presented in Appendix A. Excepting the precise definition of the weight, $\alpha_k^2(r)$, formulas (42) and (43) are similar to those used in the Expectation-Maximization (EM) procedure. The class mean, m_k, computed with (42) can be understood as the mean of the data contributions to each class, k. Such contributions correspond to the normalized $\alpha_k^2(r)$. Then non-parametric representations of likelihood functions require of considering such data contributions. For instance, we generalize QMMF to histogram based likelihood functions by weighting each pixel value with α^2 at the time of computing the histograms. Experiments that demonstrate this procedure are presented in subsection 4.3.



(a) (b) (c) (d) (e)

FIGURE 4. Interactive binary segmentation process illustration: (a) Pixels labelled by hand (red and green scribbles), (b) Likelihoods computed from the empirical class distributions, (c) computed α map, (d) label map (maximum α) and (e) segmented image.

4. Experiments

In this paper, we have mainly presented theoretical aspects of the QMMFs. However, we have also presented practical implication of the QMMFs: computation of models parameters, the binary segmentation case and minimization algorithms. In following experiments we focus in demonstrate:

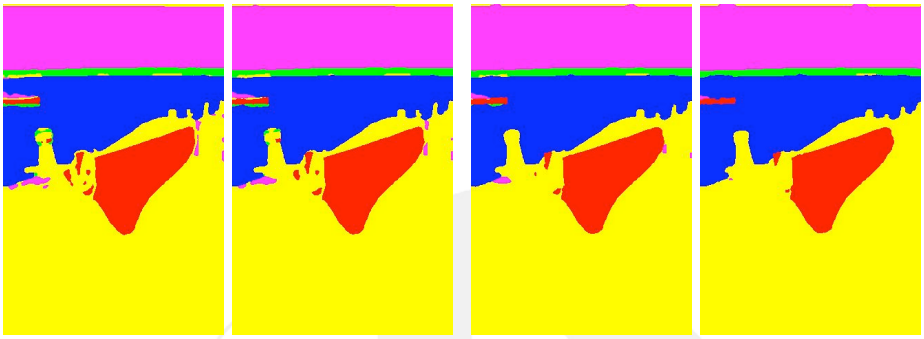
a): Practical aspects of the QMMFs;
b): Performance comparison of the binary QMMFs;
c): Generative models estimation.

4.1. **Multiclass interactive segmentation.** User interaction is a popular form for introducing prior (high level) knowledge for segmenting images with complex scenes. In that 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. Fig. 4 illustrates the interactive segmentation process. These results were computed with the proposed algorithm, see subsection 4.2.

We illustrate our multiclass segmentation method by implementing an interactive procedure, *i.e.* we assume that some pixels in the region of interest, $\mathcal{R}$, are labelled by hand, thus we have a partially labelled field (*multimap*):

$$\mathcal{A}(r) \in \{0\} \cup \mathcal{K}\,,\ \forall r \in \Omega \tag{44}$$

where $\mathcal{A}(r) = k > 0$ indicates that the pixel r was assigned to the class k and $\mathcal{A}(r) = 0$ indicates that pixel class is unknown and needs to be



(a) 10 (b) 20 (c) 50 (d) 100

FIGURE 5. Partial solutions (segmentations) for different iteration numbers.

estimated. If we assume correct user's labels , then the sum on the data term in (34) is replaced by:

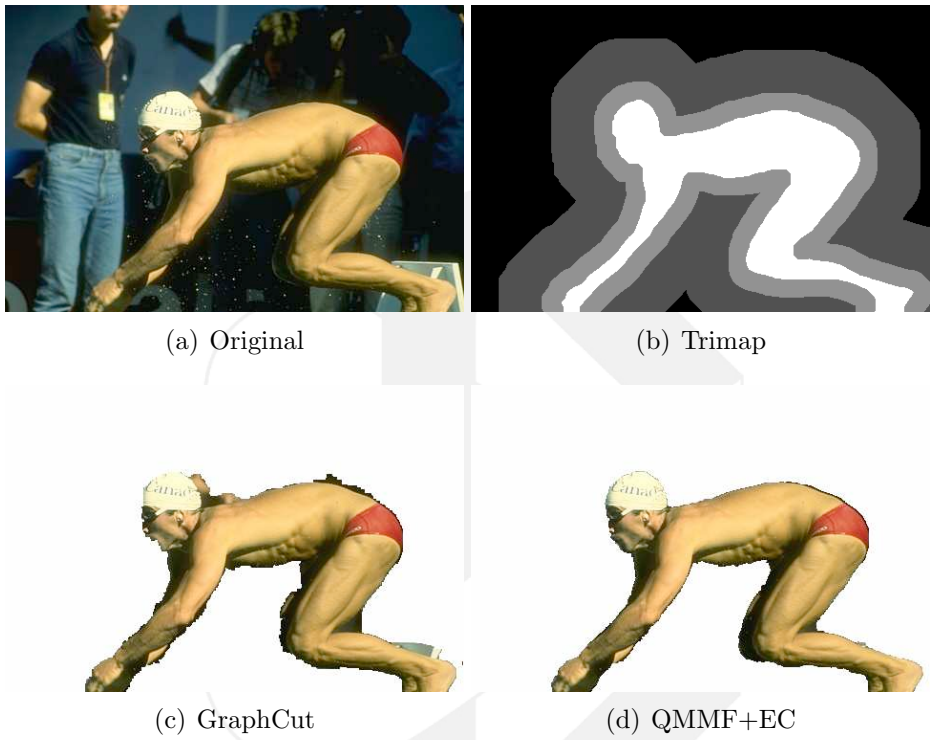
$$\sum_{r:\mathcal{A}(r)=0} \sum_{k} \alpha_k^2(r)(-\log v_k(r)). \tag{45}$$

On the other hand, 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 [25] 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 to a given class k (likelihood function, LF) is computed with:

$$LF_{ki} = \frac{\hat{h}_{ki} + \epsilon}{\sum_j (\hat{h}_{ji} + \epsilon)}, \ \forall k; \tag{46}$$

with $\epsilon = 1 \times 10^{-8}$, a small constant. Thus the likelihood of an observed pixel value is computed with $v_k(r) = LF_{ki}$ such that $i = \min_j \|g(r) - t_j\|^2$. In the experiment of Fig. 5 we used the proposed likelihood computation (with $\rho = 1$), the inter–pixel affinity measure (26) and the two step GS scheme, Eqs. (31) and (32). The shown sequence corresponds to partial solutions computed with different iteration number.



(a) Original (b) Trimap

(c) GraphCut (d) QMMF+EC

FIGURE 6. Segmentation example from the Lasso's data set.

4.2. **Quantitative Comparisson: Image Binary Interactive Segmentation.** Following 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 found more details about this study in our technical report [26]. The task is to segment into background and foreground (binary segmentation) color images allowing interactive data labeling. The generalization capabilities of the methods are compared with a *cross-validation* procedure [36]. The comparison was conducted on the Lasso benchmark database [8]; a set of 50 images online available [43]. Such a database contains a natural image set with their corresponding trimaps and the ground truth segmentations.

We opted to compute the weights using the standard formula (25), 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 labeled by hand pixels [10].

TABLE 1. Adjusted parameters for the results in table 2.

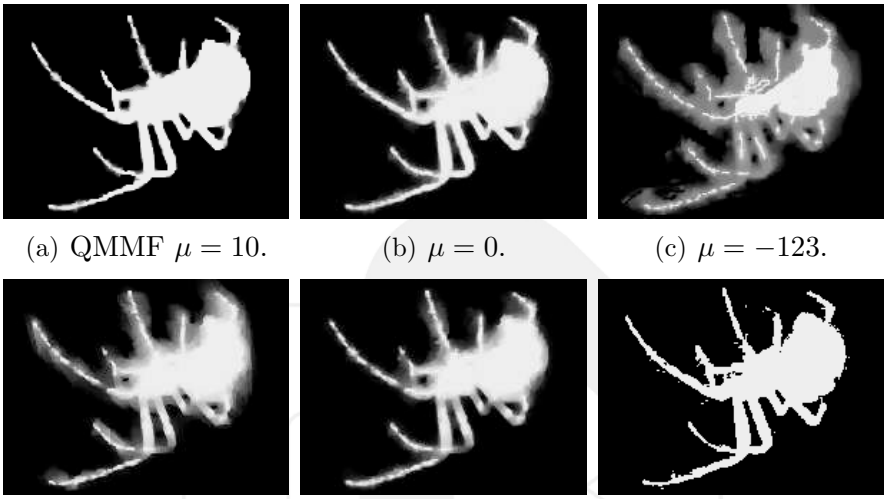
Parameter	QMMF	QMMF+EC
λ	4.7×10^3	2.28×10^5
γ	9.14×10^{-6}	5.75×10^{-3}
μ	0.0	-5.75×10^5

TABLE 2. 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%

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 [44]. We implement a cross–validation procedure following the recommendation in Ref. [36] and split the data set into 5 groups, 10 images per set. The learned parameters are reported in Table 1. Figure 6 shows an example of the segmented images. Table 2 shows the resume of the training and testing error and the Akaike information criterion (AIC) [36]. 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. Its important to note that our GC based segmentation improves significantly the reported results in [8].

We remark that the learned parameter μ for controlling the entropy (version QMMF+EC) promotes large entropy, such a parameter was appropriated for the trimap segmentation task and should not produce the expected results in other tasks. However the entropy control allows one to adapt the algorithm for different tasks, for instance for the case of simultaneously estimation of segmentation and model parameters, see subsection 3.3. The effect of the entropy control is illustrated in Figs. 7 and 8. 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. 1 was computed with QMMF using $\mu = 0$.



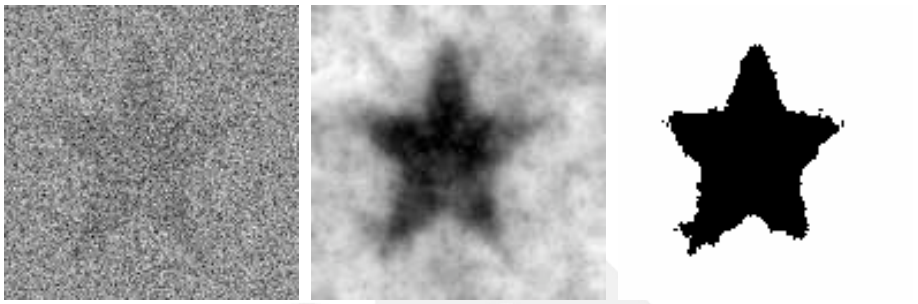
(a) QMMF $\mu = 10$. (b) $\mu = 0$. (c) $\mu = -123$.

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

FIGURE 7. 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 methods of the state of the art.

FIGURE 8. Label maps corresponding to Fig. 7, same order.

4.3. **Robust model parameter estimation.** In parametric segmentation, the computation of the exact Gaussian parameters is as important as the robustness to noise. Figure 9 shows the results when the regions are generated by i.i.d. samples of Gaussian distribution: $\mathcal{N}(0, 0.9^2)$ and $\mathcal{N}(0.5, 0.9^2)$ (see also experiments in our technical report [26]) At this SNR, the solution computed with GC, GMMF and



(a) (b) (c)

FIGURE 9. Robustness to noise: (a) Synthetic image, (b) computed α_1 and (c) segmentation.

FIGURE 10. Binarization of gray-scale random images: original images and their corresponding computed α field.

RW algorithms (do not illustrated) collapse to a single model because their large entropy of the α fields.

Fig. 10 shows the α fields computed on set of random images. Those sharper α fields result of the entropy control.

Fig. 11 shows the restoration of a corrupted binary image image. Such results were computed by estimating simultaneously the distribution parameters and the α field, assuming Gaussian distributions. 5 iteration of the two–steps scheme were required and the computation time was less than a second, compare with results in [18].

(a) (b) (c)

FIGURE 11. Binary restoration: a) Iceland corrupted map, b) computed α and c) segmentation.

In all the cases shown in Figs. 9, 10 and 11 implementations based on Gauss Markov Measure Fields (GMMF, an early variant of Random Walker [7]) collapsed to a single model [5]. That limitation of the GMMF model is discussed in [45], see also [26].

Finally, Fig. 12 demonstrates generalization of the QMMFs for computing LF based on histogram techniques. The histograms are computed by α^2–weighting the pixel values, we initially set $\alpha_k(r) = v_k(r), \forall k, r$. 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.

5. Conclusions and Discussion

We presented a derivation of the QMMF model that relax significantly 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 theoretical extensions and generalization to the QMMF model. 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

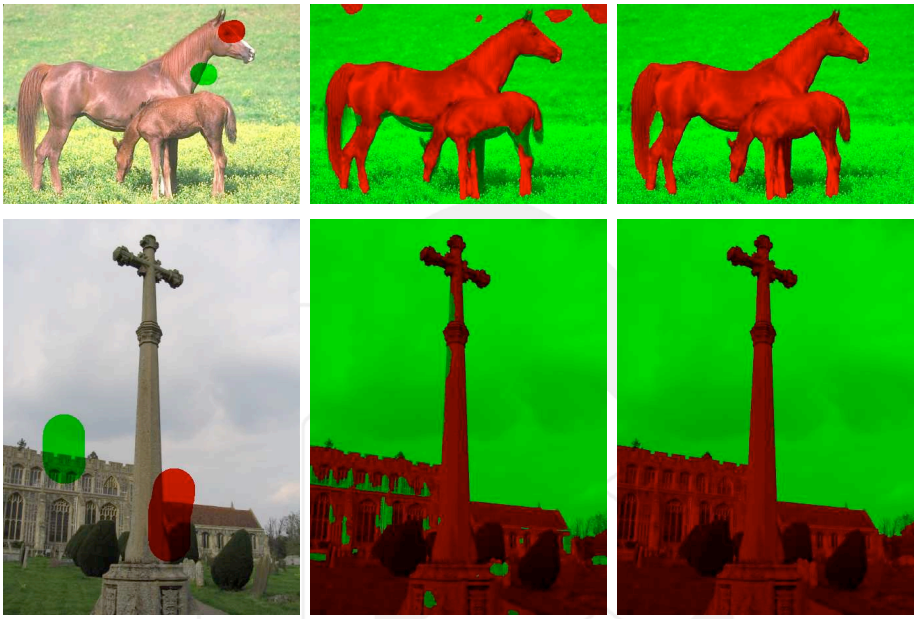


FIGURE 12. Iterative estimation of empirical likelihood functions by histograms of α^2–weighted data. Binary segmentation: initial scribbles, first iteration and second iteration; respective columns.

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.

6

Proof for Theorem 3.2. According to the generation model (7) and using (15), we have:

$$P(g|\alpha, I) \propto \prod_k \prod_r \left\{ \sum_{i=1}^{M} \frac{\pi_{ki}}{\sqrt{2\pi}\sigma_k} \exp\left[-\frac{d_{ki}^2(r)\alpha_k^2(r)}{2\sigma_k^2} \right] \right\}$$

$$= \prod_k \prod_r \left\{ \sum_{i=1}^{M} \pi_{ki} \left[G_{ki}(r) \right]^{\alpha_k^2(r)} \right\}. \tag{47}$$

Then, assuming narrow Gaussians and a large number, M, of them:

$$\lim_{M\to\infty, \sigma_k \to 0} G_{ki}(r) G_{kj}(r) = 0; \quad i \neq j, \forall k \tag{48}$$

if $m_{ki} \neq m_{kj}$ is provided. Thus, in the limit

$$\begin{aligned} \sum_{i=1}^{M} \pi_{ki} \left[G_{ki}(r)\right]^{\alpha_k^2(r)} &= \left[\sum_{i=1}^{M} \pi_{ki} G_{ki}(r)\right]^{\alpha_k^2(r)} \\ &= \left[v_k(r)\right]^{\alpha_k^2(r)}. \end{aligned} \tag{49}$$

Then (37) results of substituting (49) into (47). □

Proof for Proposition 3.5. The result is proved if we substitute in (40) the partial derivatives:

$$\begin{aligned} \frac{\partial D}{\partial m_k} &= \frac{1}{\sigma_k^2} d_k(r) v_k(r), \\ \frac{\partial D}{\partial \sigma_k} &= \left[\frac{1}{\sigma_k^2} d_k^2(r) - \frac{1}{\sigma_i}\right] v_i(r). \end{aligned}$$

Then, we solve for m_k and σ_k^2 the system $\nabla_\theta D(\alpha,\theta) = 0$. □

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.

[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] W. L. Briggs, S. McCormick, and V. Henson, *A Multigrid Tutorial*, 2nd ed. SIAM Publications, 2000.

[25] 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.

[26] ——, "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.

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

[28] L. Grady, "Computing exact discrete minimal surfaces: Extending and solving the shortest path problem in 3D with application to segmentation," in *CVPR (1)*, June 2006, pp. 69–78.

[29] O. Dalmau, M. Rivera, and P. P. Mayorga, "Computing the alpha–channel with probabilistic segmentation for image colorization," in *in Proc. ICCV, Workshop ICV 07*, 2007, pp. 1–7.

[30] P. Kohli and P. H. S. Torr, "Dynamic graph cuts for efficient inference in Markov random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2079–2088, 2007.

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

[32] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *Int'l J. Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006.

[33] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 689–694, 2004.

[34] S. Kumar and M. Hebert, "Discriminative random fields," *Int'l J. Computer Vision*, vol. 68, no. 2, pp. 179–201, 2006.

[35] L. Grady, Y. Sun, and J. Williams, "Interactive graph-based segmentation methods in cardiovascular imaging," in *Handbook of Mathematical Models in Computer Vision*, N. P. et al., Ed. Springer, 2006, pp. 453–469.

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

[37] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using gabor filters," *Pattern Recogn.*, vol. 24, no. 12, pp. 1167–1186, 1991.

[38] M. Tuceryan, "Moment based texture segmentation," *Pattern Recognition Letters*, vol. 15, pp. 659–668, 1994.

[39] S. Belongie, C. Carson, H. Greenspan, and J. Malik, "Color- and texture-based image segmentation using em and its application to content-based image retrieval," in *ICCV'98*, 1998, pp. 675–682.

[40] D. A. Clausi and H. Deng, "Design-based texture feature fusion using gabor filters and co-occurrence probabilities," *IEEE Trans. Image Processing*, vol. 14, no. 7, pp. 925–936, 2005.

[41] J. Ramirez-Ortegon and M. Rivera, "Probabilistic rules for automatic texture segmentation," in *Proc MICAI 2006*, vol. LNAI 4293. Springer, 2006, pp. 778–788.

[42] 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.

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

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

[45] 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.