Image Segmentation by Convex Quadratic Programming

Mariano Rivera, Oscar Dalmau and Josue Tago
Centro de Investigacion en Matematicas A.C.
{*mrivera, dalmau, tago*}@*cimat.mx*

Abstract

A quadratic programming formulation for multiclass image segmentation is investigated. It is proved that, in the convex case, the non-negativity constraint on the recent reported Quadratic Markov Measure Field model can be neglected and the solution preserves the probability measure property. This allows one to design efficient optimization algorithms. Additionally, it is proposed a (free parameter) inter–pixel affinity measure which is more related with 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.

1. Introduction

Image segmentation is an active research topic in computer vision and it is the core process in many practical applications, see for instance the listed in [15]. In spite of a large number of segmentation algorithms have been proposed, there is no general segmentation approach with superior performance in all the circumstances. For instance, in the literature there are reported algorithms based on: thresholds [11], morphological operators (watersheds) [2], grouping [7], level set approach [4], active contours (snakes) [13], distance based clustering (K-means and its variants) [10], maximum likelihood estimation (the EM algorithm) [1], graph cutting [3, 5], Markov Random Fields (MRF) [8, 12], random walking [9]. Among many approaches, MRF 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. In this paper, we present a probabilistic segmentation (PS) method based on MRF. Given a generative model set, PS techniques compute the probability of each model to generate the observed pixel value. Using this approach, an inherently combinatorial optimization problem is transformed into a real optimization problem. In this work we investigate the convex (positive defined) case of the recent reported Quadratic Markov Measure Fields (QMMF) [15] . QMMF models performance comparisons versus other segmentation approaches are reported in Refs. [14, 15]). Herein we prove that the optimization procedure for the convex case is much simpler than for the non–convex case: the optimization can be achieved by neglecting the non-negativity constraint. Additionally, we introduce a robust estimation of inter–pixel affinities (edges) and likelihoods in the QMMF formulation. The method capabilities are demonstrated by experiments.

In this work we assume known Likelihood Function. Thus we demonstrate the method capabilities in the framework of interactive segmentation where the data generative model can easy estimated from users scribbles. However, it is important remark that the, convex QMMF models allows to estimate the generative models parameters, see Ref. [15].

2. Convex QMMF Model

2.1 Review of the QMMF Model

Let r be a pixel in the image or the region of interest $\Omega = \{r_i : i = 0, 1, \ldots, N\}$, the class set is denoted by $\mathcal{K} = \{1, 2, \ldots, K\}$. and $v_k(r)$ represents the normalized likelihood of the pixel r of belonging to the class $k \in \mathcal{K}$, such that $\sum_k v_k(r) = 1, \ \forall r \in \Omega$. Then probabilistic (soft) segmentation approaches compute a probability measure field $\alpha = \{\alpha_k(r) : r \in \Omega, k \in \mathcal{K}\}$ such that satisfy the consistence constraint qualification (CCQ):

$$\arg\max_k \alpha_k(r) \quad = \quad \arg\max_k v_k(r), \ \forall r \in \Omega; \quad (1)$$

for the case in which not prior knowledge is provided. In addition, the probability measure field α should satisfy: $\sum_{k=1}^{K} \alpha_k(r) = \mathbb{1}^T \alpha(r) = 1, \qquad \forall r \in \Omega$ and

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

$\alpha_k(r) \geq 0, \quad \forall k \in \mathcal{K}, \forall r \in \Omega$; where $\mathbb{1} \in \mathbb{R}^K$ is a vector with all its entries equal to one.

From (1) we can see that without prior information about the nature of the solution, the most probable label for a given pixel r is the Maximum Likelihood (ML) estimator. On the other hand, if prior knowledge is available then it should be used for biasing the estimation to other different from the ML estimator. In our problem, we suppose that the image should be segmented in relative large regions, i.e. we expect, for almost all the pixels:

$$\alpha(r) \approx \alpha(s), \quad \forall r \in \Omega, \forall s \in \mathcal{N}_r \tag{2}$$

where $\mathcal{N}_r$ denotes the set of first neighbors of r: $\mathcal{N}_r = \{s \in \Omega : |r - s| = 1\}$.

Rivera *et al.* proposed in Ref. [15] the Entropy Controlled Quadratic Markov Measure Field (EC-QMMF) models for image multiclass segmentation. EC-QMMF based algorithms are computationally efficient and produce probabilistic segmentations of excellent quality, see experiments in Refs. [6, 14, 15]. The QMMF cost function has the form:

$$U(\alpha) = \frac{1}{2}\sum_k \sum_{r\in\Omega} \Big\{ \alpha_k^2(r) d_k(r) + \frac{\lambda}{2} \sum_{s\in\mathcal{N}_r} w_{rs} [\alpha_k(r) - \alpha_k(s)]^2 \Big\}, \tag{3}$$

subject to

$$\mathbb{1}^T \alpha(r) = 1, \quad \forall r \in \Omega; \tag{4}$$
$$\alpha_k(r) \geq 0, \quad \forall k \in \mathcal{K}, \forall r \in \Omega \tag{5}$$

where we define $d_k(r) \stackrel{def}{=} -\log v_k(r) - \mu$; with the parameter μ for controlling the entropy of the discrete distribution [15]. The second term in the energy (3) codifies a Gibbisan prior, based on MRF models, that enforce the soft constraint (2). Such a prior controls the granularity of the regions, i.e. promotes smooth regions. The spatial smoothness is controlled by the positive parameter λ and weights w are chosen such that: $w_{rs} \approx 1$ if the neighbor pixels r and s are likely to belong to the same class and $w_{rs} \approx 0$ in the opposite case. In subsection 2.4 we will address the weight computation process.

2.2 Non-negative Global Optimum of Convex QMMF Models

We prove that if μ is chosen such that the QMMF problem is kept convex [$d_k(r) > 0, \ \forall k, r$], then the non-negativity constraints are inactive at the optimal global solution. That means that, in such a case, the non-negativity constraints can be neglected and the optimization procedure can be achieved by using simple and efficient minimization procedures for quadratic optimization problems.

Let the vector $x \in \mathbb{R}^K$, then we define the sets $\mathcal{K}_x^+ = \{i : i \in \mathcal{K} : x_i \geq 0\}$, $\mathcal{K}_x^- = \mathcal{K} \backslash \mathcal{K}_x^+$ and the summation $S^+ = \sum_{i \in \mathcal{K}_x^+} x_i$. Moreover, we denote by $L_{x,y}(t) = x + ty$, with $t \in [0, 1]$, the line segment that links the points x and $x + y$, for a given $y \in \mathbb{R}^K$.

Theorem 2.1 (Diagonal norms) *Let $E(x) = x^T D x$ be a quadratic form with $x \in \mathbb{R}^K$ that satisfies $\mathbb{1}^T x = 1$ and $D = \operatorname{diag}\{d_1, d_2, \ldots, d_K\}$ a diagonal positive definite matrix. Thus, if there exists an index j for which $x_j < 0$ then also exists a vector y that satisfies $\mathbb{1}^T L_{x,y}(t) = 1$ and $E(L_{x,y}(t))$ decreases along the line that links x and $x + y$ meanwhile t increases from 0 to 1.*

Proof If we choose y as:

$$y_i = x_j \begin{cases} \frac{x_i}{S^+} & i \in \mathcal{K}_x^+ \\ -1 & i = j \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where $j = \arg\min_{i \in \mathcal{K}_x^-} x_i$. Then:

i. As $\mathbb{1}^T y = 0$ thus $\mathbb{1}^T L_{x,y}(t) = \mathbb{1}^T (x + ty) = \mathbb{1}^T x = 1$.

ii. To prove that the norm decreases along the segment line, first we need to prove:

- $L_{x_i,y_i}(t_1) > L_{x_i,y_i}(t_2) \geq 0$ for $0 \leq t_1 < t_2 \leq 1$ and $i \in \mathcal{K}_x^+$.
  We have $L_{x_i,y_i}(t) \geq 0$ since $S^+ + t x_j > 0$ (in other case $L_{x,y}(t)^T \mathbb{1} < 0$ in contradiction with **i**). Moreover, $L_{x_i,y_i}(t_1) > L_{x_i,y_i}(t_2)$ since $x_j \frac{x_i}{S^+}(t_1 - t_2) > 0$.
- $L_{x_i,y_i}(t_1) \leq L_{x_i,y_i}(t_2) < 0$ for $0 \leq t_1 < t_2 \leq 1$ and $i \in \mathcal{K}_x^-$. This proof is similar to the last case.

Therefore we have $L^2_{x_i,y_i}(t_1) > L^2_{x_i,y_i}(t_2)$ for all $i \in \mathcal{K}_x^-$ and, since $d_i > 0$, we conclude $E(L_{x,y}(t_1)) > E(L_{x,y}(t_2))$. ∎

Corollary 2.2 *Assuming $\mathcal{K}_x^- \neq \emptyset$ and y computed with* (6)*, then the full step $x + y$ satisfies $(x+y)^T D (x+y) < x^T D x$ and reduces in one the number of violations to non-negativity constraints: $\sharp\mathcal{K}_{x+y}^- < \sharp\mathcal{K}_x^-$, where the operator $\sharp$ computes the cardinality of a set.*

Corollary 2.3 *Let D be a diagonal positive matrix, then the solution x^* to the QP problem:* $\min_{x\in\mathbb{R}^n} x^T D x$ *s.t.* $\mathbb{1}^T x = 1$ *is non–negative* ($x_i^* \geq 0,\ \forall i$).

We assume that the vector x can be iteratively updated: $x \leftarrow x + y(x)$, note the y dependency on the actual vector x. Then the proof is straightforward from theorem 2.1 and corollary 2.2. Similarly we can state the following:

Corollary 2.4 *If $d_k(r) > 0$ is satisfied, then the solution to*

$\min_\alpha \sum_k \sum_{r\in\Omega} \alpha_k^2(r) d_k(r)$ *s.t.* $\mathbb{1}^T \alpha(r) = 1$

is a probability measure field.

Theorem 2.5 (Convex QMMF) *Let be the energy function $U(\alpha)$ defined in* (3) *and assuming $\lambda \geq 0$, then the solution to*

$$\min_\alpha U(\alpha) \quad \text{s.t.} \quad \mathbb{1}^T \alpha(r) = 1, \quad \forall r \in \Omega$$

is a probability measure field.

Proof From corollary 2.4, the unregularized solution, α^0, (setting $\lambda = 0$) is a measure probability field, i.e. α^0 satisfies (4) and (5) and the regularized solution α^λ for $0 \leq \lambda < \infty$ satisfies $\inf \alpha^0 \leq \inf \alpha^\lambda$ thus $\alpha^\lambda \geq 0$. In particular, in the limit when $\lambda \to \infty$ we have that $\alpha^\lambda = 1/K$. ∎

2.3 Optimization Algorithm

In the case of convex QMMF ($d_k(r) > 0, \forall r, k$) the optimum solution can be computed by solving the Karush Kuhn Tucker (KKT) conditions:

$$\alpha_k(r) d_k(r) + \lambda \sum_{s\in\mathcal{N}_r} w_{rs} \left(\alpha_k(r) - \alpha_k(s)\right) = \pi(r) \tag{7}$$

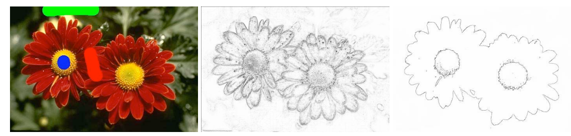
$$\mathbb{1}^T \alpha(r) = 1 \tag{8}$$

where π is the vector of Lagrange's multipliers. Note that, given π, the matrix of the linear system defined by (7) is symmetric and positive definite. We develop a two step-iterative algorithm that alternates between the computation of π and α until convergence. By integrating (7) w.r.t. k (i.e. by summing over k) and using (8):

$$\pi(r) = \frac{1}{K} \sum_j \alpha_j(r) d_j(r). \tag{9}$$

Thus, from (7):

$$\alpha_k(r) = \frac{a_k(\alpha, r) + \pi(r)}{b_k(r)} \tag{10}$$



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

Figure 1. Interpixel affinity, w_{rs}.

where we define:

$$a_k(\alpha, r) \stackrel{def}{=} \lambda \sum_{s\in\mathcal{N}_r} w_{rs} \alpha_k(s), \tag{11}$$

$$b_k(r) \stackrel{def}{=} d_k(r) + \lambda \sum_{s\in\mathcal{N}_r} w_{rs}. \tag{12}$$

Eqs. (9) and (10) define the two-step iterative algorithm. We also note that it is possible to make implicit the computation of π by substituting (9) into (10). We note that if QMMF convex, the GS scheme (10) will produce a convergent nonnegative sequence if an initial nonnegative guess for α is provided. As additional remark, one can see that the GS scheme, here proposed, is simpler than the originally reported in [15].

In the non-convex QMMF case, it is also applicable the Projected GS procedure. The projected $\tilde{\alpha}$ can be computed with $\tilde{\alpha}_k(r) \leftarrow \max\left\{0, \frac{a_k(\alpha,r)+\pi(r)}{b_k(r)}\right\}$.

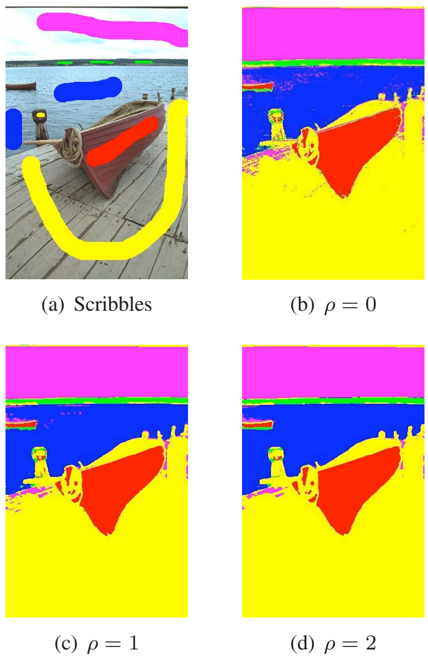
2.4 Extensions

Although the color Euclidean distance on the Lab-space is close related with the distance of human perception, it badly represents the inter–class (objects) distances. Therefore, we propose the new inter–pixel affinity measure

$$w_{rs} = \frac{v^T(r) v(s)}{|v(r)||v(s)|} \tag{13}$$

that incorporates implicitly the non-euclidean distances of the feature space by introducing prior knowledge about the feature distributions, this is illustrated in Fig. 1: panel 1a shows scribbles for three classes, panel 1b shows the gradient based inter–pixel affinity measure (computed according to [15]) and panel 1c the likelihood based edges.

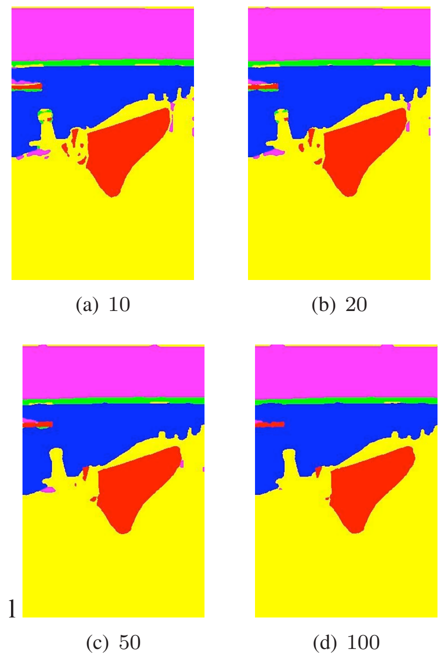
A common way for extending the likelihood computation for introducing texture information is enlarging the feature vector with new characteristics (for instances with Gabor's filter responses or local statistics [1]). Herein we propose a novel and efficient procedure for introducing texture information without explicitly including new texture features. We consider that textured regions are generated with *i.i.d.* random samples



(a) Scribbles (b) $\rho = 0$

(c) $\rho = 1$ (d) $\rho = 2$

Figure 2. ML estimator for different Neighborhood sizes.



(a) 10 (b) 20

(c) 50 (d) 100

Figure 3. Partial solutions (segmentations) for different iteration numbers.

of particular distributions, then for computing the likelihood in a given pixel r we need to examine the likelihood of surrounding pixels $\mathcal{M}_r = \{s : |r - s| \leq \rho\}$ at a distance ρ, i.e.

$$V_k(r) \propto \prod_{s \in \mathcal{M}_r} v_k(s). \tag{14}$$

Figure 2 shows the computed likelihood with (14) using different neighborhood sizes, ρ. Note that large ρ–values reduces the granularity of the Maximum Likelihood (ML) estimator, and over–smooth small details.

3. Experiments

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

$$\mathcal{R}(r) \in \{0\} \cup \mathcal{K}\,,\ \forall r \in \Omega \tag{15}$$

where $\mathcal{R}(r) = k > 0$ indicates that the pixel r was assigned to the class k and with $\mathcal{R}(r) = 0$ that class is unknown and needs to be estimated. 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 [14] then $h = S(H)$ is the smoothed histograms, where S represents the smoothing operator implemented by a homogeneous diffusion process. Thus the normalized histograms are computed with $\hat{h}_{ki} = \frac{h_{ki}}{\sum_l h_{kl}}$ and the likelihood of the pixel r to a given class k (likelihood function, LF) is computed with:

$$LF_{ki} = \frac{\hat{h}_{ki} + \epsilon}{\sum_j (\hat{h}_{ji} + \epsilon)},\ \forall k; \tag{16}$$

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. 3 we used the proposed: likelihood computation (with $\rho = 1$), inter–pixel affinity measure (13) and the two step GS scheme (subsection 2.3). The shown sequence corresponds to partial solution computed with different iteration number. Fig. 4 shows two possible segmentations from a same image considering the task of segmenting by connected regions (with a semantical meaning as: grass, fence, etc.) or by color. The task corresponds to the first and second row, respectively.

Finally, Fig. 5 shows a different application of the

(a) Scribbles.

(b) Connected regions.

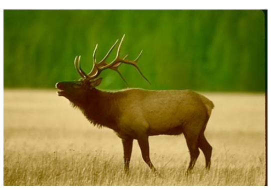
(c) Scribbles.

(d) Equal colored regions.

Figure 4. Segmentation example.

(a) Original

(b) Recolorized



(c) Recolorized

(d) Recolorized

Figure 5. Application example.

computed α probabilities. The application corresponds to the colorization method reported in [6].

4. Conclusions

In this work we have introduced extensions and presented important theoretical aspects to the probabilistic segmentation models based on QMMF. We have proved that the non-negativity constraint, in the optimization of a convex QMMF model, can be neglected and the solution preserves the measure probability field qualification. We have introduced two important modifications to the model: the inter–pixel affinity measure and the likelihood computation that considers local information and texture properties. On one hand, our inter–pixel affinity measure leads the segmentation process by likelihoods edges instead of color (or gray) gradient edges. On the other hand, the proposed likelihood takes into account the local pixel context and, implicitly, texture features. Additionally we derived a new GS scheme that is simpler and computationally more efficient (with less numerical operations and required memory) than the originally reported QMMF algorithm.

References

[1] 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*, pages 675–682.

[2] S. Beucher. Watersheds of functions and picture segmentation. In *ICASSP 82, Proc. IEEE Intern. Conf. on Acoustics, Speech and Signal Processing*, pages 3–5, 1982.

[3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE PAMI*, 23(11):1222–1239, 2001.

[4] V. Caselles, F. Catte, T. Coll, and F. Dibos. A geometric model for active contours in image processing. *Numerische Mathematik*, 66(1):1–31, 1993.

[5] B. P. D. Greig and A. Seheult. Exact maximum a posteriori estimation for binary images. *J. R. Statist. Soc. B*, 51:271–279, 1989.

[6] O. Dalmau, M. Rivera, and P. P. Mayorga. Computing the α–channel with probabilistic segmentation for image colorization. In *Proc. ICV'07*, 2007.

[7] Y. Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE PAMI*, 23(8):800–810, 2001.

[8] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE PAMI*, 6(6):721–741, 1984.

[9] L. Grady, T. Schiwietz, S. Aharon, and R. Westermann. Random Walks for interactive organ segmentation in two and three dimensions: Implementation and validation. In *MICCAI (2), LNCS 3750*, pages 773–780, 2005.

[10] A. K. Jain and R. C. Dubes. *Algorithm for Clustering Data*. Prentice Hall, Englewood Cliffs, NJ, 1998.

[11] J. Kittler and J. Illingworth. Minimum error thresholding. *Pattern Recogn.*, 19(1):41–47, 1986.

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

[13] A. W. M. Kass and Terzopoulos. Snakes: active contour models. *IJCV*.

[14] M. Rivera and P. P. Mayorga. Quadratic Markovian probability fields for image binary segmentation. In *Proc. ICV'07*, 2007.

[15] M. Rivera, O. Ocegueda, and J. L. Marroquin. Entropy-controlled quadratic Markov measure field models for efficient image segmentation. *IEEE Trans. Image Processing*, 8(12):3047–3057, Dec. 2007.