Alpha Markov Measure Field model for Probabilistic Image Segmentation

Oscar Dalmau[a], Mariano Rivera[a]

[a]*Centro de Investigación en Matemáticas, A.C. Jalisco S/N. Colonia Valencia*

Abstract

We apply the theory of *metric-divergences between probability distributions* and a variational approach in order to obtain a new model for probabilistic image segmentation. We study a specific model based on a very general measure between discrete probability distributions. We show experimentally that this model is competitive with some other models of the state of the art. In this work we use a particular case of the the *measure of kind* $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ between two discrete probability distributions.

Key words: Probabilistic segmentation, Markov random vector field, Markov random measure field, Markov random field

1. Introduction

Image segmentation has been one of the most studied tasks in image processing and it is considered a bridge between low and high level image processing tasks. Image segmentation consists in obtaining a partition of the image according to some homogeneous predicate (or property). Many strategies have been proposed according to the point of view of the image modeling. If the image is considered from the deterministic modeling viewpoint (i.e., a function defined in some space) we can find different variational approaches for image segmentation [1, 2], when is modeled as a graph [3] we find Graph Cut [4] and Normalized Cut [5]; in the context of data clustering the fuzzy c-means (FCM) methods are widely used [6, 7] and if the image is modeled as a Markov Random Field (MRF) several approaches have been reported [8, 9, 10, 11]. Among them, MRF-based models have shown to be a powerful framework to design efficient and robust models for image segmentation.

The segmentation problem can be divided in two groups: hard and soft segmentation. The objective of the hard segmentation is to find a label map while the soft segmentation methods try to find piecewise smooth maps (or functions) that can be interpreted as probabilistic maps, see Ref. [12]. That is why, the soft segmentation is also known as *probabilistic segmentation*. The study of the *probabilistic segmentation* is very important because the probabilistic segmentation allows us to obtain more robust hard segmentations. On the other hand, the *probabilistic segmentation* by itself has many other applications, see for instance Refs. [13, 14] and references therein. Therefore, the *probabilistic segmentation* is an active area and the study of new probabilistic models is crucial. The main aim of this work is to present a new model for probabilistic image segmentation. The proposed model combines a variational approach with the theory of metrics between discrete distributions [14, 15, 16]. The new model is based on the *measure of kind* $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ between two discrete probability distributions [16]. We also present a theoretical study of this model and we prove some properties and discuss particular cases.

This paper is organized as follows. In Section 2 we present a new model for probabilistic segmentation. Section 3 presents some properties of the new model and some particular cases. Section 4 shows some experimental results and finally in Section 5, we present our conclusions.

2. α Markov Measure Field model

Here, we propose a new model for image segmentation that we call α-Markov Measure Field model (α-MMF). Our model is obtained from the general formulation in Ref. [12]. The new model here presented relies on a particular case of the *measure of kind* $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ between two discrete probability distributions $\boldsymbol{f}, \boldsymbol{h}$, *i.e.*, $\boldsymbol{f}, \boldsymbol{h} \in \mathcal{S}$ where $\mathcal{S}$ (the *simplex*) is the set of all points $\boldsymbol{z} \in \mathbb{R}^K$ that satisfy the following conditions: $\mathbf{1}^T \boldsymbol{z} = 1$ and $\boldsymbol{z} \succcurlyeq 0$, see Ref. [16]. This measure is defined as

$$I^{(\alpha,\beta)}_{(\gamma,\delta)}(\boldsymbol{f}, \boldsymbol{h}) \;=\; (2^{-\beta} - 2^{-\delta})^{-1} \sum_k (f_k^\alpha h_k^\beta - f_k^\gamma h_k^\delta) \tag{1}$$

where $\alpha, \beta, \gamma, \delta > 0$. Defining the particular case

$$I^{(\alpha,\beta)}(\boldsymbol{f}, \boldsymbol{h}) \;\stackrel{def}{=}\; I^{(\alpha,\beta)}_{(1,0)}(\boldsymbol{f}, \boldsymbol{h}), \tag{2}$$

then, the α-MMF model can be formulated through the following optimization problem

$$\min_{\boldsymbol{x}\in\mathcal{S}^{|\mathcal{L}|}} U_\alpha(\boldsymbol{x};\boldsymbol{y}), \tag{3}$$

where $\mathcal{L}$ is the lattice of the image, $\mathcal{S}^{|\mathcal{L}|} \stackrel{def}{=} \{\boldsymbol{x}(r)\in\mathcal{S} : r\in\mathcal{L}\}$ and $U_\alpha(\boldsymbol{x};\boldsymbol{y})$ is defined as follows

$$U_\alpha(\boldsymbol{x};\boldsymbol{y}) \stackrel{def}{=} \sum_{r\in\mathcal{L}} I^{(\alpha,\beta)}(\boldsymbol{x}(r),\boldsymbol{y}(r)) + \lambda\sum_{s\in\mathcal{N}_r}\omega_{rs} I^{(\alpha,\alpha)}(\boldsymbol{x}(r),\boldsymbol{x}(s)), \tag{4}$$

$\boldsymbol{y}\in\mathcal{S}^{|\mathcal{L}|}$ represents the likelihood of pixels to belong to some models, see Experiment section, and ω_{rs} is a weight function, *e.g.*, in the experiments we use $\omega_{rs}=\psi_\gamma(g(r),g(s))$, where γ is a small positive value, for other weight functions see Refs [17, 18]. By eliminating the constants that appear in the functional (3) we obtain

$$\min_{\boldsymbol{x}\in\mathcal{S}^{|\mathcal{L}|}} U(\boldsymbol{x};\boldsymbol{y},\alpha), \tag{5}$$

$$U(\boldsymbol{x};\boldsymbol{y},\alpha) = -\frac{1}{\alpha}\sum_{r\in\mathcal{L}}\sum_{k\in\mathcal{K}}\left[x_k^\alpha(r)y_k^\beta(r) + \lambda\sum_{s\in\mathcal{N}_r}\omega_{rs}x_k^\alpha(r)x_k^\alpha(s)\right] \tag{6}$$

where $0<\alpha\leq 1$ so that the similarity measures $I^{(\alpha,\beta)}_{(1,0)}(\cdot,\cdot)$, $I^{(\alpha,\alpha)}_{(1,0)}(\cdot,\cdot)$ are convex. It is important to note that the measure of the second term of the functional in Eq. (3) was selected in such a way that the prior energy is symmetric. The solution of the constrained optimization problem (5)-(6) can be obtained by using the Lagrange multipliers and the *Coordinate Descent* method [19]—in the MRFs context, this optimization strategy can be seen as an ICM method [20]. The Lagrangian, without the non-negativity constraints is:

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{\pi};\boldsymbol{y},\alpha) = U(\boldsymbol{x};\boldsymbol{y},\alpha) + \sum_{r\in\mathcal{L}}\pi(r)\big[\sum_{k\in\mathcal{K}}x_k(r)-1\big],$$

where $\pi(r)$ are the Lagrange multipliers associated with the constraint given by Eq. (6). Computing the first derivative w.r.t. each component $x_k(r)$ of the vector measure field $\boldsymbol{x}$ and equating to zero yields:

$$\frac{\partial\mathcal{L}(\boldsymbol{x},\boldsymbol{\pi};\boldsymbol{y},\alpha)}{\partial x_k(r)} = -x_k^{\alpha-1}(r)y_k^\beta(r) - \lambda\sum_{s\in\mathcal{N}_r}\omega_{rs}x_k^{\alpha-1}(r)x_k^\alpha(s) + \pi(r) = 0.$$

Then, we obtain, implicitly, the component $x_k(r)$ in terms of its Lagrange multiplier

$$x_k(r) = \left[\frac{n_k(r)}{\pi(r)}\right]^{\frac{1}{1-\alpha}}, \tag{7}$$

where $n_k(r)$ is defined as

$$n_k(r) \stackrel{def}{=} y_k^{\beta}(r) + \lambda \sum_{s \in \mathcal{N}_r} \omega_{rs} x_k^{\alpha}(s). \tag{8}$$

Taking into account that $\boldsymbol{x}(r) \in \mathcal{S}$ then

$$\sum_{k \in \mathcal{K}} x_k(r) = 1, \tag{9}$$

where $\mathcal{K} \stackrel{def}{=} \{1, 2, \cdots, K\}$. Substituting $x_k(r)$ in the previous equality we can compute the Lagrange multiplier $\pi(r)$. Then, using (7) we obtain the following expression for $x_k(r)$,

$$x_k(r) = \frac{[n_k(r)]^{\frac{1}{1-\alpha}}}{\sum_{j \in \mathcal{K}} [n_j(r)]^{\frac{1}{1-\alpha}}}. \tag{10}$$

Finally, the algorithm for the α-MMF model is obtained by iterating the Eqs. (8) and (10) until a convergence criterion. We note that the optimization problem (4) is a nonlinear programming problem, so this strategy produces, in general, a local minimum.

3. Analysis of the α-MMF model

To gain a better understanding of the proposed model, in this section we discuss particular cases and some properties.

3.1. Some properties

Proposition 3.1. *Let us consider the following Linear Programming problem*

$$\min_{\boldsymbol{z}} \boldsymbol{a}^T \boldsymbol{z}, \tag{11a}$$

$$s.t : \boldsymbol{z} \in \mathcal{S}, \tag{11b}$$

where $\boldsymbol{a} \in \mathbb{R}^K$. If $j = \arg\min_k a_k$ then $\boldsymbol{z}^* = \boldsymbol{e}_j$ is an optimum solution.

Proof. Clearly the problem (11a)-(11b) is a linear programming problem (LP), which is feasible and bounded. Applying *Fundamental Theorem of Linear Programming* [19], then it has an optimal solution and at least one such solutions (the basic optimal points) is a *basic feasible point*. Note the vectors of the canonical basis $\boldsymbol{e} = [e_1, e_2, \cdots, e_K]^T$ are *basic feasible points*, therefore the solutions lie on the polytope defined by the vectors of the canonical basis . Using the condition $j = \arg\min_k a_k$ then $\boldsymbol{z}^* = \boldsymbol{e}_j$ is an optimum solution. □

Proposition 3.2. *Let* $h(x) = \frac{\sum_{k\in\mathcal{K}} a_k^x \log a_k}{\sum_{k\in\mathcal{K}} a_k^x}$ *be a real function where* $0 \leq a_k \leq 1,\ \forall k \in \mathcal{K}$ *(*$a_k \geq 1,\ \forall k \in \mathcal{K}$*) and* $\sum_{k\in\mathcal{K}} a_k \neq 0$*, then* $h(x)$ *is an increasing function.*

Proof. It is enough to prove that $h'(x) \geq 0$. We will assume without loss of generality that $0 \leq a_k \leq 1,\ \forall k \in \mathcal{K}$ (the proof is similar if $a_k \geq 1,\ \forall k \in \mathcal{K}$). By applying $\log$ function to both sides of $h(x)$ and computing the derivative w.r.t x we get

$$h'(x) \;=\; h(x)\left(\frac{\sum_{k\in\mathcal{K}} a_k^x \log^2 a_k}{\sum_{k\in\mathcal{K}} a_k^x \log a_k} - \frac{\sum_{k\in\mathcal{K}} a_k^x \log a_k}{\sum_{k\in\mathcal{K}} a_k^x}\right).$$

As $h(x) < 0$ we need to prove that

$$\frac{\sum_{k\in\mathcal{K}} a_k^x \log^2 a_k}{\sum_{k\in\mathcal{K}} a_k^x \log a_k} \;\leq\; \frac{\sum_{k\in\mathcal{K}} a_k^x \log a_k}{\sum_{k\in\mathcal{K}} a_k^x}, \tag{12}$$

or equivalently

$$\frac{\sum_{k\in\mathcal{K}} a_k^x \log^2 a_k}{\sum_{k\in\mathcal{K}} a_k^x} \;\geq\; \left(\frac{\sum_{k\in\mathcal{K}} a_k^x \log a_k}{\sum_{k\in\mathcal{K}} a_k^x}\right)^2. \tag{13}$$

The inequality (13) follows from Jensen's inequality

$$f(\sum_{k\in\mathcal{K}} \omega_k x_k) \;\leq\; \sum_{k\in\mathcal{K}} \omega_k f(x_k), \tag{14}$$

using $f(x) = x^2$, $\omega_k = \frac{a_k^x}{\sum_k a_k^x}$ and $x_k = -\log a_k$. □

Proposition 3.3. *Let* $\boldsymbol{a}(x)$ *be a vector of* $\mathbb{R}^K$ *whose components are defined as follows* $\boldsymbol{a}_k(x) = \frac{a_k^x}{\sum_{k\in\mathcal{K}} a_k^x}$ *where* $0 \leq a_k \leq 1,\ \forall k \in \mathcal{K}$ *(*$a_k \geq 1,\ \forall k \in \mathcal{K}$*) and* $\sum_{k\in\mathcal{K}} a_k \neq 0$*. Then, the entropy of* $\boldsymbol{a}(x)$ *is a decreasing function of* x*.*

Proof. We are to prove that if $x < y$ then $H(\boldsymbol{a}(x)) \geq H(\boldsymbol{a}(y))$ where $H(\cdot)$ is certain entropy measure. First, we need to select an entropy measure. There have been many attempts to generalize Shannon's entropy $H(\boldsymbol{f}) = -\sum_{k\in\mathcal{K}} f_k \log f_k$, where $\boldsymbol{f} \in \mathcal{S}$ is a discrete distribution. Therefore, we will prove the proposition based on some generalizations. Rényi's entropy (or the entropy of order α) is defined as

$$H_\alpha(\boldsymbol{f}) \;=\; \frac{\log\left(\sum_{k\in\mathcal{K}} f_k^\alpha\right)}{1-\alpha},\ \alpha \neq 1,\ \alpha > 0, \tag{15}$$

Another generalization of Shannon's entropy is the entropy of type β (or the entropy of degree β) due to Havrda and Charvat, see [15], and is defined as

$$H^\beta(\boldsymbol{f}) \;=\; \frac{\sum_{k\in\mathcal{K}} f_k^\beta - 1}{2^{1-\beta}-1},\ \beta \neq 1,\ \beta > 0, \tag{16}$$

and, $H(\boldsymbol{f}) = \lim_{\alpha\to 1} H_\alpha(\boldsymbol{f}) = \lim_{\beta\to 1} H^\beta(\boldsymbol{f})$. The entropies (15) and (16) share the same term $T(\boldsymbol{f}) = \sum_{k\in\mathcal{K}} f_k^\alpha$, hence it is better to work with this term. We will assume without loss of generality that $\alpha > 1$, then the inequality $H(\boldsymbol{a}(x)) \geq H(\boldsymbol{a}(y))$ is equivalent to prove that

$$T(\boldsymbol{a}(x)) \;\leq\; T(\boldsymbol{a}(y)).$$

Defining the function

$$g(x) \;\stackrel{def}{=}\; T(\boldsymbol{a}(x)) = \frac{\sum_{k\in\mathcal{K}} a_k^{x\alpha}}{(\sum_{k\in\mathcal{K}} a_k^x)^\alpha},$$

it suffices to prove that $g(x)$ is an increasing function, or that $g'(x) \geq 0$. By applying $\log$ function to both sides of $g(x)$ and computing the derivative w.r.t x we get

$$g'(x) \;=\; \alpha g(x)\left(\frac{\sum_{k\in\mathcal{K}} a_k^{x\alpha}\log a_k}{\sum_{k\in\mathcal{K}} a_k^{x\alpha}} - \frac{\sum_{k\in\mathcal{K}} a_k^{x}\log a_k}{\sum_{k\in\mathcal{K}} a_k^{x}}\right). \tag{17}$$

Finally, the result $g'(x) \geq 0$ follows from the fact that $g(x) > 0$ and applying Proposition 3.2 to the last factor of (17). □

The proposed model, also satisfies the following properties:

Proposition 3.4. *The optimum solution $\boldsymbol{x}^*$ of the minimization of the data term in Eq. (4) satisfies*

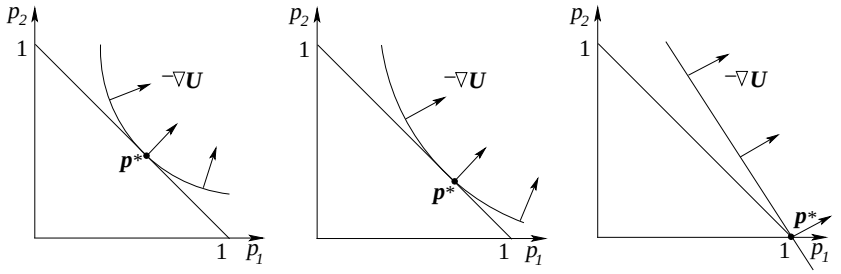

Figure 1: 2D example. The entropy of the solution decreases while the parameter α increases from 0 to 1.

i) $\arg\max_{k\in\mathcal{K}} \boldsymbol{x}^* = \arg\max_{k\in\mathcal{K}} \boldsymbol{y}$ *for all* $\alpha \in (0, 1]$,

ii) $\boldsymbol{x}^*(r) = \boldsymbol{e_{k_r}}$ *if* $\alpha = 1$ *and* $\forall r \in \mathcal{L}\ \exists k_r \in \mathcal{K}$ *such that* $y_{k_r}(r) > y_i(r)\ \forall i \neq k_r$,

iii) $\boldsymbol{x}^* = \boldsymbol{y}$ *if* $\alpha = \beta = \frac{1}{2}$.

Proof.
The proof of i) is obtained directly by using the Eq. (10).
The proof of ii) follows from Proposition 3.1.
The proof of iii) follows from Cauchy-Schwarz inequality. □

Proposition 3.5. *If there is no prior information about* $\boldsymbol{x}$ *(*i.e. $\lambda = 0$*) then the entropy of the* α*-MMF model is a decreasing function of* α.

The proof follows from proposition 3.3. In addition, the Fig. 1 depicts how the entropy of the solution of the model (5)-(6) decreases while the parameter α increases from 0 to 1. Finally, we conclude that changing the parameter α in our model (3) we can control the entropy of the solution.

Table 1: The best parameters obtained after training each method on the whole Lasso's benchmark.

Parameters	ECQMMF	αMMF
λ	3.18×10^5	2.91×10^5
γ	2.84×10^{-2}	1.94×10^{-3}
α	-7.08×10^4	4.67×10^{-1}

Figure 2: 'Ceramic' image taken from the Lasso database. a) Original image b) trimap: foreground (in white), background (in dark gray), the unknown region (in light gray) and the non-process region (in black), c) groun8dtruth: foreground (in white), background (in black), the region in gray is not taken into account in the comparisons.

3.2. Particular cases

- **Case** $\alpha = 1$: If α is set to 1 in the functional Eq. (3) we obtain the following model,

$$\min_{\boldsymbol{x}\in\mathcal{S}^{|\mathcal{L}|}} U_1(\boldsymbol{x};\boldsymbol{y}), \tag{18}$$

where

$$U_1(\boldsymbol{x};\boldsymbol{y}) \stackrel{def}{=} I^{(1,\beta)}(\boldsymbol{x}(r),\boldsymbol{y}(r)) + \lambda \sum_{s\in\mathcal{N}_r} \omega_{rs} I^{(1,1)}(\boldsymbol{x}(r),\boldsymbol{x}(s)). \tag{19}$$

Eliminating the constants values in Eq. (19) produces the functional

$$U(\boldsymbol{x};\boldsymbol{y},1) = -\sum_{r\in\mathcal{L}}\sum_{k\in\mathcal{K}} \left[x_k(r) y_k^{\beta}(r) + \lambda \sum_{s\in\mathcal{N}_r} \omega_{rs} x_k(r) x_k(s) \right]. \tag{20}$$

obtaining the following optimization problem

$$\min_{\boldsymbol{x}\in\mathcal{S}^{|\mathcal{L}|}} U(\boldsymbol{x};\boldsymbol{y},1). \tag{21}$$

The solution of the previous optimization problem could be obtained using the formula (10) when α approaches to 1, that is, $x_k(r)$ can be computed with

Table 2: Comparative performance of probabilistic segmentation methods using the Lasso's bench database. The values in the table represent the classification errors (MSE) and are expressed in percentage. These results were obtained by using the best parameters after training each method on the whole database.

Filename	Group	ECQMMF	αMMF	Filename	Group	ECQMMF	αMMF
21077	1	3.88	3.74	cross	3	1.62	1.46
86016	1	3.30	3.63	grave	3	1.64	1.19
181079	1	6.31	6.65	person2	3	0.61	0.49
271008	1	3.73	2.71	person7	3	0.88	0.66
banana1	1	2.57	1.17	stone2	3	0.41	0.17
bush	1	5.84	6.92	65019	4	0.59	0.65
flower	1	0.54	0.46	153077	4	1.66	1.49
music	1	1.66	1.49	209070	4	2.27	2.12
person5	1	3.47	3.22	376043	4	6.59	6.85
sheep	1	5.91	5.34	book	4	4.13	2.45
37073	2	1.70	1.70	doll	4	0.42	0.4
106024	2	8.68	7.52	llama	4	4.98	5.77
189080	2	4.20	4.01	person3	4	1.03	1.01
304074	2	10.75	9.67	person8	4	0.55	0.65
banana2	2	0.72	0.45	teddy	4	3.50	3.50
ceramic	2	1.37	0.98	69020	5	4.71	4.10
fullmoon	2	0.00	0.00	153093	5	4.61	5.27
person1	2	0.53	0.37	227092	5	3.4	4.14
person6	2	4.84	5.13	388016	5	1.28	1.24
stone1	2	1.04	0.78	bool	5	2.01	1.63
24077	3	4.27	4.10	elefant	5	1.14	0.79
124080	3	5.70	5.06	memorial	5	1.52	1.22
208001	3	1.76	1.86	person4	5	2.96	3.29
326038	3	7.19	5.89	scissors	5	1.97	1.73
banana3	3	1.87	1.88	tennis	5	7.73	6.74

$$x_k(r) \;=\; \lim_{\alpha\to 1} \frac{[n_k(r)]^{\frac{1}{1-\alpha}}}{\sum_{j\in\mathcal{K}} [n_j(r)]^{\frac{1}{1-\alpha}}} = \delta(k_r - k), \tag{22}$$

where $\delta(\cdot)$ is the Kronecker delta and $k_r = \arg\max_{j\in\mathcal{K}} x_j(r) = \arg\max_{j\in\mathcal{K}} n_j(r)$. That is, at each iteration we compute $n_j(r), j \in \mathcal{K}$ and then we set $x_k(r) = 1$ if the corresponding value of $n_k(r)$ is the greatest or $x_k(r) = 0$ in any other case.

- **Case** $\alpha = 1, \beta \to 0$: In this case, $I^{(1,\beta)}(\cdot,\cdot)$ becomes the Kerridge's Inaccuracy measure [21], that is

$$\lim_{\beta\to 0} I^{(1,\beta)}(\boldsymbol{f},\boldsymbol{h}) \;=\; -\sum_k f_k \log h_k,$$

and the new functional is

$$\min_{\boldsymbol{x}\in\mathcal{S}^{|\mathcal{L}|}} -\sum_{r\in\mathcal{L}}\sum_{k\in\mathcal{K}} \left[x_k(r)\log y_k(r) + \lambda \sum_{s\in\mathcal{N}_r} \omega_{rs} x_k(r) x_k(s) \right], \tag{23}$$

where $n_k(r) = \log y_k(r) + \lambda \sum_{s\in\mathcal{N}_r} \omega_{rs} x_k(s)$ and now can take different signs. The solution of the previous problem can also be obtained using the iterative method based on the expression (22). This functional, see Eq. (23), is very similar to the MPM-MAP model proposed by Marroquin et al. in [8]. However, there are two main differences. First, in the MPM-MAP model $x_k(r)$ takes discrete values, *i.e.* $x_k(r) \in \{0, 1\}$. Second, the regularization term is different too.

4. Experiments

In the interactive segmentation the image is partitioned in different regions (a trimap), see Ref. [22]: foreground ($\mathcal{F}$), background ($\mathcal{B}$), unknown ($\mathcal{U}$), and a non-process region ($\mathcal{O}$), see Fig. 2. The problem consists of estimating the class ($\mathcal{F}$ or $\mathcal{B}$) to which the pixels located in the unknown region belong.

We make a comparison using the best algorithm reported in [23], the Entropy Control Quadratic Markov Measure Field (ECQMMF) and the one proposed here,

Table 3: Summary of some statistics computed for each method.

	EQQMMF	αMMF
mean	3.08	**2.87**
median	2.14	**1.87**
std	2.45	**2.37**
min	**0.00**	**0.00**
max	10.75	**9.67**

Table 4: Cross-validation results. Training errors for each group and method, and the mean training error for each model.

Training Group	EQMMF	αMMF	Testing Group	EQMMF	αMMF
1	3.01	**2.83**	1	3.21	3.08
2	3.06	**2.97**	2	2.93	**2.50**
3	3.17	**3.01**	3	2.49	**2.43**
4	2.97	**2.82**	4	3.27	**3.08**
5	2.86	**2.71**	5	3.68	**3.53**
mean	3.01	**2.87**	mean	3.11	**2.92**

the α-MMF. For making such a comparison we use the Lasso's database available online at Ref. [24]. This database consists of 50 color images. Each image has a *trimap* and its segmented by hand image (*groundtruth*, image t). To compare the methods we use as comparison measure the *mean square error* (MSE) between the groundtruth and the segmented image (s) obtained by each method. The MSE is computed just in the region of interest (the unknown region $\mathcal{U}$), that is:

$$MSE(s,t) \;=\; \frac{1}{|\mathcal{U}|}\sum_{r\in\mathcal{U}}(s(r)-t(r))^2. \tag{24}$$

We note that the α-MMF and ECQMMF models have 3 parameters. Then, each method has the same number of parameters $\theta = [\lambda,\gamma,\alpha]^T$, where λ controls the spatial smoothness, γ controls the sharpness of the edges and α controls the entropy of the solution. To obtain the solution of the ECQMMF model we use the formula proposed in Ref. [11].

As preference measure to the models (foreground and background) we use the empirical normalized likelihoods, see Refs. [22, 23] for details:

$$y_k(r) \;=\; \frac{h_k(g(r)) + \epsilon}{\sum_{j\in\{1,2\}}[h_j(g(r)) + \epsilon]},$$

where $\epsilon = 10^{-3}$ is a small positive value that prevents the above expression is undefined. h_1 and h_2 are the empirical color histograms that corresponds to the foreground and the background respectively. We did experiments computing the histograms in Lab and RGB color spaces and the results were very similar. Therefore, in the experiments presented here the empirical histograms are computed in the RGB color space.

First, the parameter set θ was trained for each method on the whole Lasso's database by minimizing the MSE, Eq. (24), between the groundtruth and the segmented images. We follow the evaluation procedure proposed in [23]: a cross-validation with a parameter optimization stage. The minimization of the MSE was done by using the Nelder Mead method [25, 26]. In particular, we use the implementation in the Numerical Recipes [26]. Second, we applied a k-fold cross-validation, see Refs. [27, 28], by dividing the database in 5 groups. In particular, we use the same groups reported in [23].

The results of the first experiment appear in the Tables 1-3. The Table 1 shows the best parameter set obtained for each method during the training step. The Table 2 presents the results obtained for each method on the whole Lasso's dataset by using the best parameter set, Table 1, obtained in the training step. Table 3 shows a summary of some statistics which are based on the results presented in the Table 2. The results of the second experiment (cross-validation) are shown in Table 4. Note that the α-MMF has the best performance in both evaluations. The difference between the results reported in Table 2 and in Ref. [23] with respect to the ECQMMF model, is due to the color empirical histograms are computed in the RGB color space. In Ref. [23] the color empirical histograms were computed in the Lab color space.

According to the Table 1, the algorithms tried to increase the entropy during the training step. That is, for this problem, the models 'prefer' to have high entropy. Observe that, in the case of the ECQMMF model, the value of α is less than zero. Therefore, the result would have been the same if instead of penalizing the Gini's entropy we would have penalized the L_2-norm in the local prior energy. The α-MMF model also 'prefers' a solution with high entropy because $\alpha < 0.5$, see Table 1. However, we note that there problems for which solutions with low

entropy is better, for instance: the applications presented in Ref. [18] and the simultaneous estimation of segmentation and parameters problem [11, 23].

5. Conclusions

We present a new model, α-MMF, for probabilistic image segmentation. We also analyze some properties and limit cases of the presented model. For evaluating the performance of the new model we use the interactive segmentation task. We compared our model with the ECQMMF model which has been recently reported as the best method is this image processing task. We show experimentally that the α-MMF model has competitive results compared with some algorithms of the state of the art.

Acknowledgment

This research was supported in part by CONACYT (grant 61367) and PROMEP (grant 103.5/08/2919). O. Dalmau was also supported in part by a PhD scholarship from CONACYT, Mexico.

References

[1] D. Mumford, J. Shah, Optimal approximation by piecewise smooth functions and associated variational problem, Commun. Pure Appl. Math. (1989) 577–685.

[2] G. A. Hewer, C. Kenney, B. S. Manjunath, Variational image segmentation using boundary functions, IEEE Transactions on Image Processing 7 (1998) 1269–1282.

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

[4] Y. Boykov, M. P. Jolly, Interactive organ segmentation using graph cuts, in: MICCAI, LNCS 1935, 2000, pp. 276–286.

[5] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 888–905.

[6] J.-F. Yang, S.-S. Hao, P.-C. Chung, Color image segmentation using fuzzy c-means and eigenspace projections, Signal Processing 82 (3) (2002) 461–472.

[7] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, T.-J. Chen, Fuzzy c-means clustering with spatial information for image segmentation, Computerized Medical Imaging and Graphics 30 (2006) 9–15.

[8] J. L. Marroquin, S. Botello, F. Calderon, B. C. Vemuri, MPM-MAP algorithm for image segmentation, in: International Conference on Pattern Recognition (ICPR'00), Vol. 1, 2000, pp. 300–310.

[9] J. L. Marroquin, F. Velazco, M. Rivera, M. Nakamura, Gauss-markov measure field models for low-level vision, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2001) 337–348.

[10] J. L. Marroquin, E. Arce, S. Botello, Hidden Markov measure field models for image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2003) 1380–1387.

[11] M. Rivera, O. Ocegueda, J. L. Marroquín, Entropy-controlled quadratic markov measure field models for efficient image segmentation, IEEE Transactions on Image Processing 16 (2007) 3047–3057.

[12] O. D. Cedeño, M. Rivera, A general bayesian markov random field model for probabilistic image segmentation, in: IWCIA, LNCS 5852, 2009, pp. 149–161.

[13] M. Rivera, O. Ocegueda, J. L. Marroquin, Entropy controlled quadratic markov measure field models for efficient image segmentation, IEEE Trans. Image Process.

[14] O. Dalmau, M. Rivera, T. Alarcon, Bayesian Scheme for Interactive Colourization, Recolourization and Image/Video Editing, To appear in Computer Graphics Forum.

[15] B. D. Sharma, I. J. Taneja, Functional Measures in Information Theory, Funckcialaj Ekvacioj 17 (1974) 181–191.

[16] I. J. Taneja, H. Gupta, On generalized measures of relative information and inaccuracy, Applications of Mathematics 23 (1978) 317–333.

[17] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1990) 629–639.

[18] O. Dalmau, M. Rivera, P. P. Mayorga, Computing the alpha-channel with probabilistic segmentation for image colorization, in: IEEE Proc. Workshop in Interactive Computer Vision (ICV'07), 2007, pp. 1–7.

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

[20] J. Besag, On the statistical analisys of dirty pictures, J. R. Stat. Soc., Ser. B, Methodol. 48 (1986) 259–302.

[21] D. Kerrigde, Inaccuaracy and Inference, J. Royal Statist. Society, Ser. B 23 (1961) 184–194.

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

[23] M. Rivera, P. P. Mayorga, Comparative study on quadratic markovian probability fields for image binary segmentation, Tech. Rep. I-07-15, Centro de Investigacion en Matematicas, A.C. (Dec 2007).

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

[25] J. A. Nelder, R. Mead, A simplex method for function minimization, Comput. J. 7 (1965) 308–313.

[26] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, Numerical recipes in C (2nd ed.): the art of scientific computing, Cambridge University Press, New York, NY, USA, 1992.

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

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