Two–level MRF Models for Image Restoration and Segmentation

Mariano Rivera
Centro de Investigacion en Matematicas
Guanajuato, GTO 36240 MEXICO

James C. Gee
University of Pennsylvania
Philadelphia, PA 19104, USA

Abstract

We present a new general Bayesian formulation for simultaneously restoring and segmenting piecewise smooth images. This implies estimation of the associated parameters of the classes within an image, the class label for each image pixel and the number of classes. The intensity image is modelled by parametric models based on regularized networks. The method fits the regions (or classes) with complex spatial intensity distributions with an identifiable group of simple models. Prior information is introduced in form of a two-level Markov random field (MRF). The low–level MRF models the information required to recover piecewise restorations, while the high-level MRF constraints the segmentation. The high–level MRF supports a merging process of simple intensity models into classes.

1   Introduction

In this paper, we present a unified method for image segmentation and restoration formulated using Bayesian statistics [1]–[3]. Before establishing the restoration–segmentation problem, we introduce our image observation model. We assume a two steps image model, this model is illustrated in figure 1–(a). In the first step, the hidden image $\hat{f}$ is generated by parametric intensity models and then, in a second step, the observed data g are produced. These steps are followed detailed:

I) Generation Step. We consider that the real image $\hat{f}$, in the regular lattice L, is an assemble of K disjoint regions, $R = \{R_1, R_2, ..., R_K\}$. Then, we suppose that the intensity value of the image $\hat{f}$ within each region R_q is generated by means of a parametric model $\hat{\Phi}_q$. Therefore, we have a set of parametric *intensity models* (IM), $\hat{\Phi} = \{\hat{\Phi}_1, \hat{\Phi}_2, ..., \hat{\Phi}_K\}$, with the corresponding parameters $\Theta = \{\Theta_1, \Theta_2, ..., \Theta_K\}$. Thus, let r a pixel in the regular lattice, L, then the pixel value, $\hat{f}_r$, is generated by $\hat{f}_r = \hat{\Phi}_{\hat{c}_r}(\Theta_{\hat{c}_r}, r)$, where $\hat{c}_r$ is the index to the region that corresponds to the pixel r (i.e. $\hat{c}_r \in \{1, 2, ..., K\}$). For instance, a piecewise constant $\hat{f}$ results of assuming the simple IM: $\hat{\Phi}_q(\Theta_q, r) = m_q$, where m_q is the intensity value for all the pixels $r \in R_q$. More practical parametric IM are: polynomials [5], basis models [6], splines [8] or radial basis functions [9].

II) Observation Step. Now, as we said, $\hat{f}$ is not directly observed, but the data g is. The relationship between g and $\hat{f}$ is expressed by $g_r = [H\hat{f}]_r + \eta_r$, where H is a known linear operator and η represents independent additive noise.

Without lost of generality, we only consider the restoration case of denoising, i.e. we assume $H = Id$. Therefore, the image generation–observation model (GOM) is

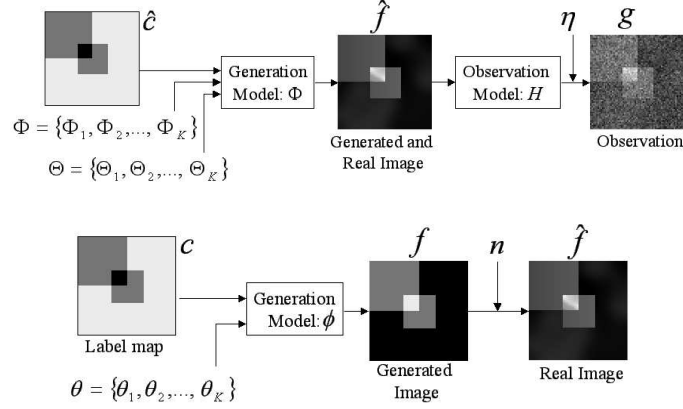$$g_r = \hat{\Phi}_{\hat{c}_r}(\Theta_{\hat{c}_r}, r) + \eta_r. \tag{1}$$

Figure 1: (Top) Generation-observation image model with exact intensity models $\hat{\Phi}$ and (Bottom) generation step with approximated intensity models ϕ. See text for more details.

The segmentation task can be seen as the solution of the inverse problem expressed by (1). This is, the computation of the label map $\hat{c}$. However, this task requires of the exact knowledge of the IM's, $\hat{\Phi}$, and of the respective parameter values, Θ. In the general case, the IM and their parameters are not known and they need be estimated. This is a difficult estimation task; one must consider very variate forms of models. For example, in the case of natural images, intensity models for the sky, walls of bricks, paths, forest, etc. Fortunately, the task can be simplified if one approaches the GOM (panel 1–(a)) by substituting the generation stage by the one shown in panel 1–(b). In the new GOM, the true unknown set of IM, $\hat{\Phi}$, is approximated by a set $\Phi = \{\phi(\theta_1), \phi(\theta_2), ..., \phi(\theta_K)\}$, where ϕ is a generic parametric and θ_j is the parameter set corresponding to the jth region. Then, the image pixel f_r is generated with:

$$f_r = \phi\left(\theta_{c_r}, r\right), \tag{2}$$

where the label c_r selects the appropriated parameter set. Note that c, in general, differs from the true segmentation, $\hat{c}$. Now, we introduce the residual, $n = \hat{f} - f$, that represents superfluous details in each class that do not provide useful information. Note that, by depending on the application: an image detail is spurious or it is not. For example, moles in the face are important features in the face recognition task, however, moles may introduce ambiguity in the face detection task. Other example of n is the texture in objects, as the bricks in a wall. Then the image generation model (1) is rewritten as:

$$g_r = \phi\left(\theta_{c_r}, r\right) + n_{r,c_r} + \eta_r. \tag{3}$$

In the current work, we present a parametric restoration–segmentation method based on Bayesian regularization and Markov Random Fields (MRF) [3]. The method is a generic edge–preserving regularization method and at the same time, the method produces a segmentation of the feature processed (intensity in this case). However, in order to motivate the segmentation method, we first present and discuss the case of parametric filtering. The parametric filtering method is based on the minimization of an energy functional associated to a MAP criterion. This energy function is close related with others previously reported in nonparametric restoration, particularly with robust or edge–preserving regularization [10]-[17]. However, instead of estimate the restoration, f, we compute the

parameters set of the IM ϕ and the label map, c. The label field c, computed by this way, is consistent with the prior that promotes piecewise smooth restorations (i.e. an edge–preserving filter). This low-level (pixel–level) prior models intra–pixel interactions and it does not necessarily promotes smooth segmentations. Thus c can be different from the real segmentation field, $\hat{c}$. This is better understood with an example: we suppose that our IMs are too rigid (for instance we use constant models) for adjust a well defined region without observable discontinuities (as the sky in the image of figure 2), then a parametric filtering produces a smooth quantization of the image and, at the same time, such region is over–segmented by effect of the quantization. Therefore, we introduce a additional MRF that codifies the intra–label (or intra–region) interaction and allows to group subregions, c's, into a region, $\hat{c}_i$. The grouping process, based on the high–level MRF is inspired on techniques of grouping based on graph theory [18][19].

In the seminal work of Darrell and Pentland [5], simple multi–layered models (based on polynomials) were used for representing and segmenting smooth piecewise gray scale images or optical flow fields for regions containing multiple motions. More recently, other authors have reported different approaches or applications of layered models. For example, layered models were proposed for: computing the optical flow [6] and tracking [7]. Differently to those schemes, the feature of interest inside the jth region is only generated with the model ϕ_j determined by c; the final segmentation, $\hat{c}$, is estimated by grouping the over–segmented regions defined by c. It is important to note that, in our formulation, the computation of c is not a critical stage and can be achieved with fast greedy optimization algorithms.

The rest of the paper is organized as follows: section 2 presents the MAP parametric restoration method, section 3 presents the extension of the parametric restoration formulation for effectively segmenting complex image, a practical implementation is presented in section 4 and finally, the conclusions are given in section 5.

2 MAP Parametric Restoration

In this section, we present the MAP criterion for estimating a parametric representation of the image f (see figure 1-(b)). The objective is to compute the label field, c, and the parameters, θ such that maximize the posterior distribution of the restoration $P(f(\theta,c) \mid g)$ instead of $P(c,\theta \mid g)$, as in a segmentation problem [1]–[3].

The MAP estimator is computed by maximizing the posterior probability distribution: $P(f \mid g) \propto P(g \mid f)P(f)$. The likelihood, $P(g \mid f)$, promotes fidelity of the restoration to the data. On the other hand, the prior distribution, $P(f)$, is modelled by a low–level MRF (at pixel pairs level) with a Gibbs distribution. Then we obtains the posterior probability distribution: $P(f \mid g) \propto \exp(-U_f(f;g))$. Where U_f is a, well-known, robust regularized cost function [11, 12]:

$$U_f(f;g) = \sum_{r\in L}\rho_1\left(f_r,g_r\right) + \lambda \sum_{\langle r,s\rangle}\rho_2\left(f_r,f_s\right). \tag{4}$$

where $\langle r,s \rangle = \{(r,s) : (r,s) \in L, |r-s| < 2\}$ is the set of first horizontal, vertical and diagonal neighbor pixels and λ is a "lumped" hyper–parameter of the Gibbs distribution. The robust potentials, ρ_1 and ρ_2, are used in order to reject outliers and preserves intensity edges, respectively. [10]–[17].

Differently to the robust filtering case of (4), in our case, the restoration f is expressed by the parametric intensity model (IM), ϕ, with specific parameters θ_{c_r} for the support

region, c_r [see (2)]. The selection of the generic intensity model ϕ is important for the success of the multi–layer strategy. Too simple intensity models fail to represent regions with complex spatial distributions of the feature of interest, while more realistic models have a large number of parameters that need to be estimated. We use a regularized network (RN) [20], with linear interpolation functions, as the generic intensity model, ϕ, see [9]. In the RN formulation, the intensity value of the pixel r is computed with:

$$f_r = \phi\left(\theta_{c_r}, r\right) \overset{def}{=} \sum_j \theta_{c_r, j} \Psi_j(r), \tag{5}$$

where $\Psi_j(r)$ represents the jth interpolation function evaluated at the pixel r whose contribution is controlled by the parameter $\theta_{c_r, j}$—control point in the RN formulation. Now, the prior distribution, $P(f)$ should express our *a priori* knowledge: a) The restoration, f, is piecewise smooth and b) each layer, $q \in \{1, 2, ..., K\}$ defined by $\sum_j \theta_{q, j} \Psi_j(r)$, is a smooth surface. Last two priors can be expressed in terms of MRF theory. As result, a regularization membrane potential is added to the parametric version of (4):

$$\begin{aligned} U_\phi(\theta, c; g) &= \sum_{r \in L} \rho_1\left(\phi\left(\theta_{c_r}, r\right), g_r\right) + \lambda \sum_{\langle r, s \rangle} \rho_2\left(\phi\left(\theta_{c_r}, r\right), \phi\left(\theta_{c_s}, s\right)\right) \\ &\quad + \gamma \sum_q \sum_{\langle r, s \rangle} \left(\phi\left(\theta_q, r\right) - \phi\left(\theta_q, s\right)\right)^2, \end{aligned} \tag{6}$$

where λ and γ are "lumped" hyper–parameters of the Gibbs distributions. The robust potential ρ_1 has a similar behavior than in (4). However ρ_2 controls the number of regions (the granularity) of the restoration, see experiment of Figure 2. In the experiments reported in this work: the control points are distributed in a coarse grid. The parametric flexible IM (5) can represent the smooth variation of complex regions with a few number of control points (parameters). If the number of control points is increased, then the IM can better adjust the local smooth variations in the region. One can use more control points such that that each pixel in the image is represented by a "parameter": $\phi\left(\theta_r, r\right) \overset{def}{=} \theta_{c_r, r}$. The implication of this case is studied in section 4.

In the current work, we implement the same robust potential for both the likelihood and prior potentials: $\rho_1(x, y) = \rho(x - y; k_1)$ and $\rho_2(x, y) = \rho(x - y; k_2)$, where ρ can be either the Andrew function, $\rho(z; k) = 1 - \exp\left(-(z/k)^2\right)$, or the Geman-McClure function, $\rho(z; k) = (z/k)^2 / \left(1 + (z/k)^2\right)$. In the case of non-parametric restoration (4), the robust potentials determine the convexity of the cost function and hence the convergence to a global minima [11, 12]. In our experiments with parametric restoration, we do not distinguish a significant difference in the performance by using any of the mentioned potentials.

The MAP label field, c^*, and the model parameters, θ^*, are computed by iteratively performing alternated minimizations of (6) with deterministic algorithms in each step. Given that $U(\theta, c; g)$ is bounded by zero and it is decreased in each step, thus, the convergence to a local optimum is guaranteed. A good initialization is thus required to steer the algorithm toward the desired solution. The model parameters are updated by performing a few iterations of a half-quadratic minimization algorithm. Specifically, we use three iterations of the weighted version of the Gauss-Seidel algorithm detailed in Ref. [12]. Note, however, that the labelling step corresponds to a combinatorial problem. In this work, we implement such minimization with either of the next two algorithms:
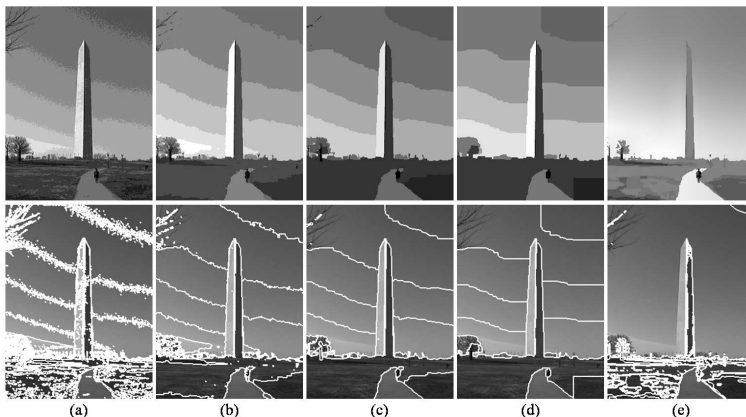
Figure 2: Edge preserving filtering with (a)-(d) flat and (e) flexible intensity models.

I) Iterated conditional modes (ICM) algorithm[2]. Such approach involves local computations and is thus computationally fast. The ICM tends toward undesirable classifications is mitigated here by our use of robust potentials in the data term.

II) Multiway graph cut [21, 22, 23]. The energy function (6) satisfies the requirements to be minimized by a graph cut scheme. We implement the α–β multiway cut algorithm proposed in [23], in such case, the graph capacities between pixel-nodes are computed with $w_{rs} = W - \rho_2\left(\phi(\theta_{c_r}, r), \phi(\theta_{c_s}, s)\right)$ and capacities to the sources with $w_{qs} = W - \rho_1\left(\phi(\theta_q, r), g_r\right)$; where W is a large constant that guaranty positive capacities (in our implementation $W = 1$).

Note that the exact MAP estimator for c can, however, be computed in each iteration with more computationally expensive algorithms as simulated annealing [3].

2.1   Experiments of parametric edge–preserving filtering

Figure 2 shows the result of an experiment designed for illustrating the performance of the cost function (6) in the edge preserving filtering task and the behavior of the method to changes in the parameter k_2, the one that controls the robust interaction between restored pixels. First row shows the restorations and second row shows the respective edges between subregions. Column 2a shows the maximum likelihood estimation used as as initial guess. Columns 2b to 2d show the results computed by increasing the robustness of the intra-pixel potential, i.e. with k_2 equal to 0.01, 0.1 and 1.0, respectively. The other parameters, $k_1 = 0.2$, $\lambda = 1.5$ and $K = 10$ were not modified. The algorithm for labelling was the multiway graph cut with Andrew's potential functions. The method preserves the real intensity edges in the filtered image and the number of subregions is effectively controlled by the parameter k_2: a small k_2 value preserves details while a large k_2 value reduces the number of regions and promotes large regions.

Flat intensity models have been extensively used in image segmentation because of its simplicity and ease of computation. However, intensity gradients produce a "staircase" effect and the restoration appear over-segmented. On the other hand, flexible intensity layers are proper for regions with intensity gradients [column 2e], however edges with low contrast are over–smoothed. Then, in order to effectively compute a segmentation, additional work should be done either with flat or flexible intensity models. We will focus such task in section 3, meanwhile we present an experiment with flexible intensity model

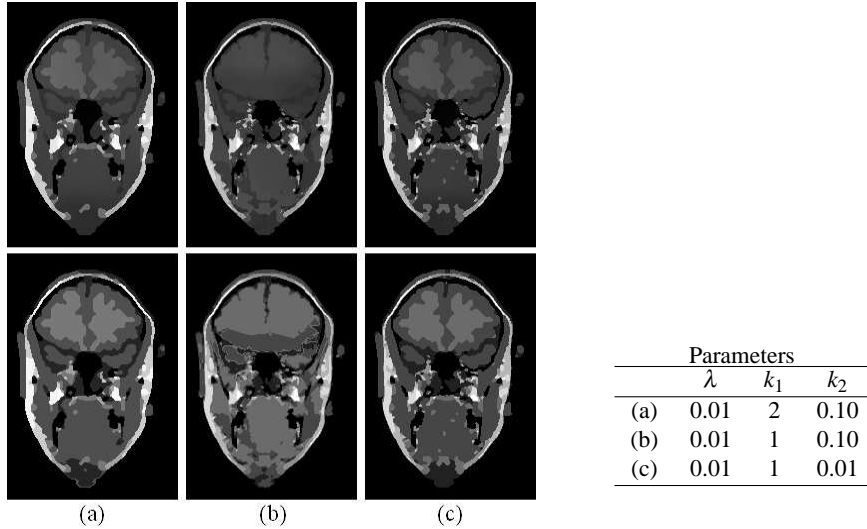|	Parameters		
	λ	k_1	k_2
(a)	0.01	2	0.10
(b)	0.01	1	0.10
(c)	0.01	1	0.01

| (a) | (b) | (c) |

Figure 3: Effect of the scale parameter in the Edge-preserving filtering task with flexible intensity models, see text. Filtered images (first row) and label fields (second row).

and analyzes the effect of the scale parameters, k_1 and k_2. Figure shows the results computed 10 flexible models each one with 9×9 control points (parameters), and a stiffness $\gamma = 2$ (see [17] for details). Ψ was a bilinear interpolator the potentials were the Geman–McClure function and the labelling step was achieved with the ICM algorithm. The data corresponds to a magnetic resonance of a coronal slice of the head. First row, panels 3-(a) to 3-(c), shows the filtered images and second row the corresponding label fields. The left side show a summary of the parameters used. We select by hand the parameters in column 3-(b) (central column) such that a good subjective result was obtained; this subjective criteria was to preserve the main structures. Now, column 3-(a) shows the effect of reducing the data term robustness. So that, more details are preserved and the granularity is low. On the other hand, if the edge detection is increased (keeping fixed the data term robustness) small details are also preserved but the solutions appear more granular, column 3-(c)).

3 MAP Restoration and Segmentation

The experiments on subsection 2.1 illustrated the performance of the parametric restoration method. In those experiments, we appreciated that edges can effectively be preserved in the "restorations" or filtered images. Those results are comparable with non-parametric edge preserving regularization methods as the reported in Refs. [14]-[17]. However, it is important to note that the computed label maps (see second row in figure 3) need be refined in order to effectively estimate the segmentation map, $\hat{c}$, the real GOM (figure 1-(a)). For example, in panel 3–(b), the brain region is smooth, but there were needed at least two intensity models. So these subregions should be merged.

In this section, we extend the parametric restoration method presented in section 2 for effectively compute a segmentation. A graph–based interpretation motivates an approach for weighting the edges of the graph that represent the relationship (or similarity) between subregions. This more general approach is derived in the same form as the original formulation: with base on MRF and Bayesian Regularization. In the graph paradigm,
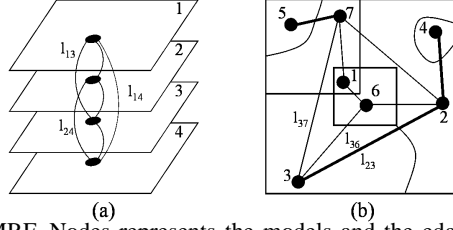
Figure 4: High-level MRF. Nodes represents the models and the edges, l, represent the intra–models interaction. Models pairs with high interaction are candidates to be grouped. (a) Global models (GM) and (b) local models (LM).

the intensity models are represented by nodes and the inter–model interaction processes is modelled by the weights, l, see figure 4. We estimate the field l as a hidden variable of our system and establish a formal mechanism for introducing prior knowledge in its computation.

The approach implements the segmentation by merging labels with high affinity. In the new formulation, the edges, $l_{ij} \in [0,1]$, need be estimated. Differently to standard graphics–based region grouping algorithms [18, 19, 24] or heuristics merging procedures [25], we incorporate l_{ij} in our MRF formulation to explicitly model the degree of interaction or affinity between the pair of models i and j. The join posterior probability distribution for the restoration, $f(c, \theta)$, and the interaction coefficients l, corresponds in this case to:

$$P(f, l \mid g) \propto P(g \mid f, l) P(f \mid l) P(l), \tag{7}$$

where we propose to use a conditional probability, $P(f \mid l)$, such that its negative–log, $\sum_{\langle r,s \rangle} \rho_2 \left(\phi \left(\theta_{c_r}, r \right), \phi \left(\theta_{c_s}, s \right) \right) l_{c_r c_s}^2$, expresses the relationship between the low-level and high-level MRF's and l_{pq} measures of the interaction between two given regions or models $\langle\langle p, q \rangle\rangle$ (nodes), where $\langle\langle, \rangle\rangle$ denotes a high-level "clique." Therefore, the local energy of the particular low-level clique $\langle r, s \rangle$ results from the potential of the clique $\langle r, s \rangle$ and the interaction degree between the models c_r and c_s. The prior distribution for l, penalizes the merging process of the given pair of regions (IM) $\langle\langle p, q \rangle\rangle$: $-\log P(l) = \mu\lambda \sum_{\langle\langle p,q \rangle\rangle} (1 - l_{pq})^2$; where μ is a positive parameter. Thus, (7) is of the form: $P(f, l \mid g) \propto \exp(-U_l(c, \theta, l; g))$ and l is computed as the minimizer of $U_l(c, \theta, l; g)$, for a given c and θ in an additional step of our iterative algorithm, with the closed formula:

$$l_{pq} = \frac{\mu \left| \langle\langle p, q \rangle\rangle \right|}{\mu \left| \langle\langle p, q \rangle\rangle \right| + \sum_{\langle c_r, c_s \rangle = \langle\langle p,q \rangle\rangle} \rho_2 \left(\phi \left(\theta_{c_r}, r \right) - \phi \left(\theta_{c_s}, s \right) \right)}, \tag{8}$$

where $\left| \langle\langle p, q \rangle\rangle \right|$ is the cardinality of the set of first neighboring pixels with labels i and j—note that $\langle\langle \cdot \rangle\rangle$ is a symmetric operator. Examining (8), one can see that $l_{pq} \in [0, 1]$: $l_{pq} \approx 1$ if the sum that ranges over all the cliques $\langle c_r, c_s \rangle = \langle\langle p, q \rangle\rangle$ is small (i.e., the affinity between models p and q is small) and $l_{pq} \approx 0$ if the number of highly interacting low–level cliques increases. The availability of the interaction processes, l, thus enables an alternative segmentation approach in which models are grouped.

Figure 5 compares minimization algorithms for the MAP parametric segmentation. First column shows the restoration and second column shows the edges between models superposed on the original synthetic noisy data, a piecewise constant (of image 256×256 pixels) with 3 gray levels corrupted with gaussian noise). Row 5a shows the results computed (after 1500 sec.) with the simulated annealing (SA) algorithm using the exact
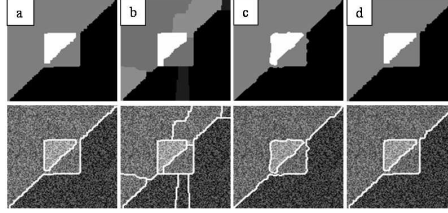
Figure 5: MAP parametric restoration (first row) and segmentation (second row). (a) Simulated annealing (SA) with exact IM, (b) SA with $K = 10$. (c) Proposed method with ICM labelling and merging. (d) Proposed method with multiway cut labelling.

3 intensity models (IM). Row 5b shows the partial result of one labelling step computed with SA using 10 IM after 20000 sec. Row 5c shows the result computed with 10 IM using ICM labelling and then by grouping in 3 classes the IM according to the estimated interaction processes, l. The intra–model graph corresponds to the illustrated in Fig. 4a. Row 5d shows the result directly computed using the multiway cut labelling (with 10 IM and with parameter as in Fig. 5d). The computational time for 5c and 5d were similar (5 sec. approximately). In general, that multiway cut labelling produces better results than ICM labelling because it is not easy trapped by a "bad" local minima and the parameter k_2 controls effectively the size (and number) of the regions. This is an important feature for merging based segmentation because it efficiency depends on the number of regions.

4   Practical Algorithm for Restoration and Segmentation

Now, we present a practical implementation for segmenting complex scenes, i.e. images with regions with large intensity gradients and gaps in edges. The keystone of this iterative procedure is to assume that:

I) The label map c is given. This means that an initial over–segmentation of the image is available. It can be computed very efficiently with the parametric restoration method presented in section 2 (for instance, with the multiway cut implementation for the labelling step).

II) The image intensity is modelled with dense membranes (i.e. $f_r = \theta_{c_r,r}$). In such case, the parameters, of the layer q, that need be updated and stored in memory are the ones such that $c_r = q$. Therefore, the whole set of useful parameters can be stored in a single image with the same dimension of the data, g. We define such useful parameters as $\theta_r \stackrel{def}{=} \theta_{c_r,r}$.

Then, we propose the cost function:

$$\hat{U}(\theta, l; c, g) = \sum_{r \in L} \rho_1(\theta_r, g_r) + \lambda \sum_{\langle r, s \rangle} \left[\rho_2(\theta_r, \theta_s) l^2_{c_r c_s} + \mu (1 - l_{c_r c_s})^2 \right], \qquad (9)$$

where c is a given over–segmentation. The algorithm is following described: in a first step the alternated minimization of (9) w.r.t. θ, and l is performed and then in a second step the two subregions (labels) with largest affinity are merged. We continue the minimization-merging iterations while the highest affinity is larger than a given threshold, τ. This threshold can be estimated as: $\tau = \min_{qq} l_{qq}$. Differently to the region competition algorithm[26], the merging process is controlled by the affinity measure l.

Figure 6 illustrates the segmentation process of real images with regions within large gradient in the gray level. Column 6-(a) shows the initial over–segmentation. The final

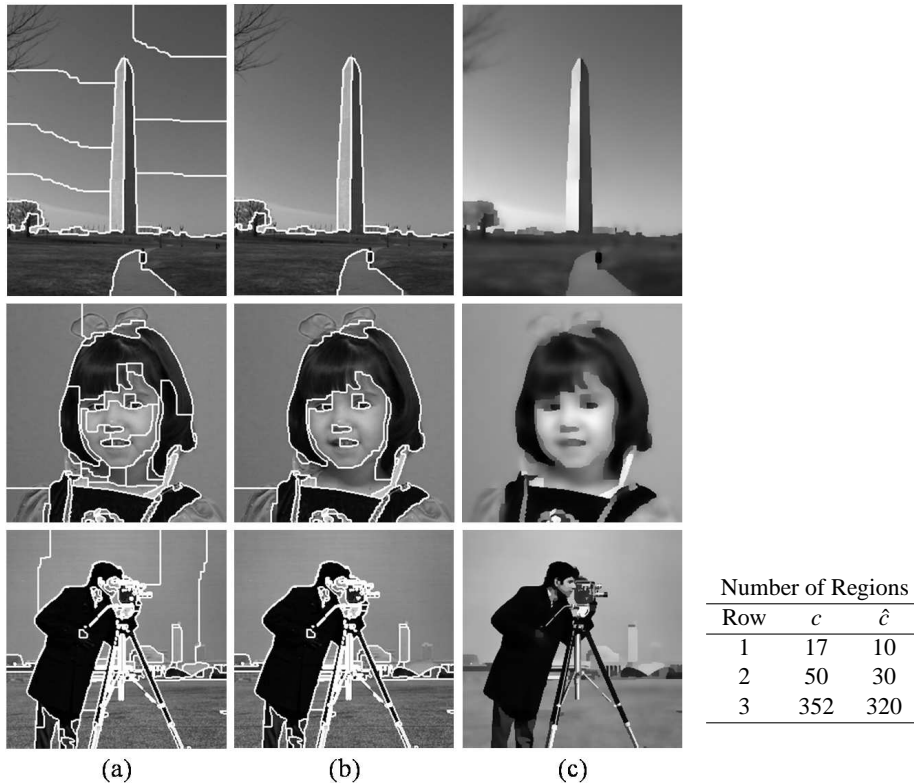|	Number of Regions	
Row	c	$\hat{c}$
1	17	10
2	50	30
3	352	320

(a)   (b)   (c)

Figure 6: Segmentation of real images, rows: 1) obelisk, 2) Snow White, and 3) cameraman

segmentations and restorations are shown in columns 6-(b) and 6-(c), respectively. The results correspond to stop by hand a systematic merging of regions, i.e. at each iteration the pair of IM with largest affinity are merged. The experiment details are resumed in the left side of figure 6. The parameters were fixed in the set of experiments: $K = 10, \lambda = 10, k_1 = 0.2, k_2 = 1, \mu = 0.1$.

5   Conclusions

We presented a new general Bayesian formulation for simultaneously restoring and segmenting images. We derived a set of methods suitable for tasks in image processing and low-level computer vision. The presented methods are associated with the MAP estimation criterion, where prior information is introduced in form of a two-level MRF. The low–level MRF models the information required to recover piecewise smooth restorations. The label map computed with the low-level MRF corresponds to an over–segmentation of the image. The high–level MRF models affinity or interaction between simple intensity models or labels (represented by regions or layers). In this way, the correct segmentation is obtained by taking into account the estimated information on inter-models affinity. This can be understood as a merging process of subregions into classes. We proposed efficient minimization algorithms in terms of computational time and memory requirement that guarantee convergence, at least, to a local minima.

References

[1] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and Bayesian restoration of images," *IEEE–PAMI*, 6, 721–741, 1984.

[2] J. Besag, "On the statistics analysis of dirty pictures," *J. Royal Statistical Soc. B* , 48, 259–302, 1986.

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

[4] Z. Tu, S.C. Zhu and H.Y. Shum, "Image Segmentation by Data Driven Markov Chain Monte Carlo," in *ICCV'01*, Vancouver, BC, Canada, 131–138, 2001.

[5] T.J. Darrell and A.P. Pentland, "Cooperative robust estimation using layers of support," *IEEE–PAMI*, 17, 5, 474–487, 1995.

[6] M.J. Black, D.J. Fleet, Y. Yacoob, " Robustly estimating changes in image appearance," *CVIU*, 78, 8–31, 2000.

[7] H. Tao, H.S. Sawhney, R. Kunar, "Object tracking with Bayesian estimation dynamic layer representations," *IEEE–PAMI*, 1, 75–89, 2002.

[8] J.L. Marroquin, S. Botello, F. Calderon and B.C. Vemuri, "The MPM-MAP algorithm for image segmentation," in *ICPR'00, 2000.*

[9] M. Rivera and J.C. Gee, "Image segmentation by flexible models based on robust regularized networks," *ECCV 2002*, vol. III, 621–634, 2002.

[10] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization and FFTs," *IEEE–PAMI*, 14, 367–383, 1992.

[11] M.J. Black and A. Rangarajan, "Unification of line process, outlier rejection, and robust statistics with application in early vision," *IJCV*, 19, 57–92, 1996.

[12] P. Charbonnier, L. Blanc-Féraud, G. Aubert and M. Barlaud, "Deterministic edge-preserving regularization in computer imaging," *IEEE Trans. Image Process.,* 6, 298–311, 1997.

[13] S. Geman and D.E. McClure, "Bayesian image analysis methods: An application to single photon emission computed tomography," in *Proc. Statistical Computation Section*, Amer. Statistical Assoc., 12–18, 1985.

[14] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Image Process.,* 14, 367–383, 1992.

[15] T. Kubota and T. Huntsberg, "Adaptive anisotropic parameter estimation in the weak membrane model," *EMMCVPR'97*, 179–194, 1997.

[16] S. Z. Li, "Close-form solution and parameter selection for convex minimization-based edge-preserving smoothing," *IEEE–PAMI*, 20, 916–932, 1998.

[17] M. Rivera and J.L. Marroquin, "Adaptive rest condition potentials: first and second order edge–preserving regularization," *CVIU*, 88, 76–93, 2002

[18] Z. Wu and R. Leaby, "An optimal graph theoretical approach to data clustering: Theory and its applications to image segmentation," *IEEE–PAMI*, 11, 1101—-1113, 1993.

[19] J. Shi and J. Malik, "Normalized cuts and image segmentation," in *IEEE–PAMI*, 22, 888–905, 2000.

[20] F. Girosi, M. Jones and T. Poggio, "Regularization theory and neural networks architectures," *Neural Computing*, 7, 219–269, 1995.

[21] Y. Boykov, O. Veksler, and R. Zabih "Fast approximate energy minimization via graph cuts." *IEEE–PAMI* , 2001.

[22] V. Kolmogorov and R. Zbih, "What energy functions can be minimized via graph cuts," *ECCV'02*, III, 65–81, 2002

[23] Y. Boykov, O. Veksler, and R. Zabih "Energy minimization with disontinuities,"*submitted to IJCV*, 2003.

[24] M. Sonka, S.K. Tadikonda and S.M. Collins, "Knowledge–based interpretation of MR brain images," *IEEE Trans. Medical Imaging*, 15, 443–452, 1996.

[25] S. Birchfield and C. Tomasi, "Multiway Cut for Stereo and Motion with Slanted Surfaces", *ICCV'09*, 489-495, 1999.

[26] S.C. Zhu and A. Yuille, "The region competition: Unifying snakes, region growing and Bayes/MDL for multiband image segmentation," *PAMI* 18, 884-900, 1996.