Image Segmentation by Flexible Models Based on Robust Regularized Networks

Mariano Rivera[1,2] and James Gee[1]

[1] Department of Radiology, University of Pennsylvania
3600 Market Street, Suite 370, Philadelphia, PA, USA 19104
{mrivera,gee}@grasp.cis.upenn.edu
[2] Centro de Investigacion en Matematicas A.C.
Apdo. Postal 402, Guanajuato, Gto., Mexico 36020
mrivera@cimat.mx
http://www.cimat.mx/~mrivera

Abstract. The object of this paper is to present a formulation for the segmentation and restoration problem using flexible models with a robust regularized network (RRN). A two-steps iterative algorithm is presented. In the first step an approximation of the classification is computed by using a local minimization algorithm, and in the second step the parameters of the RRN are updated. The use of robust potentials is motivated by (a) classification errors that can result from the use of local minimizer algorithms in the implementation, and (b) the need to adapt the RN using local image gradient information to improve fidelity of the model to the data.

Keywords. Segmentation, Restoration, Edge-preserving Regularization.

1 Introduction

Image segmentation from different cues (such as gray level, local orientation or frequency, texture, and motion) is typically formulated as a clustering problem, where the consideration of spatial interactions among pixel labels provides additional, useful constraints on the problem. Thus, in order to classify a pixel, one takes into account the observed value (which is generally noisy) and the classification of neighboring pixels. Specifically, let g be the image that is to be segmented into K classes according to its gray level, then—in the context of Markov random fields [1]— the classification c and the means of the gray levels $\mu = \{\mu_1, \mu_2, ..., \mu_K\}$ can be computed by minimizing a cost functional of the general form:

$$U(c, \mu) = \sum_{r \in L} \left\{ V_1 \left(\mu_{c_r} - g_r \right) + \lambda \sum_{s \in N_r} V_2 \left(c_r, c_s \right) \right\}, \tag{1}$$

where r indicates the position of a pixel in the regular lattice L, $c_r \in C$ (with $C = \{1, 2, ..., K\}$) is the classification for pixel r, μ_q is the centroid of class q, the neighborhood $N_r = \{s : |s - r| = 1\}$. Functional (1) has two terms, their

A. Heyden et al. (Eds.): ECCV 2002, LNCS 2352, pp. 621–634, 2002.

contribution to the total cost controlled by the positive parameter λ. Within the context of Bayesian regularization, the first term corresponds to the negative of the log-likelihood and promotes fidelity of the segmentation to the observed data. Assuming Gaussian noise, for instance, $V_1(x) = |x|^2$. The second term, regularization term, encodes our *a priori* knowledge that the images contain spatially extended regions. To achieve this, the potential of the regularization term is constructed to penalize differences in the classification of pixel r with respect to the pixels within its neighborhood N_r, one of the most commonly used being the Ising potential [3]:

$$V_2(x, y) = 1 - 2\delta(x - y), \tag{2}$$

where $\delta(\cdot) \in \{0, 1\}$ is the Kronecker's delta function. The trouble with the Ising potential (2), however is that it leads to a combinatorial optimization problem. Stochastic algorithms that can minimize (1), such as simulated annealing [2][3], are extremely time consuming. In contrast, fast deterministic algorithms such as the Iterated Conditional Modes (ICM) [4], are easily trapped by "bad" local minima. Consequently, investigators have proposed fast segmentation schemes that produce "good results with reasonable computational cost" based on approximation of the cost functional (1) or variants of deterministic minimization algorithms; for example, the techniques as the Mean Field theory [5][6],[7] Multigrid-ICM [8], Relaxation Labeling [9][10], Annealing-ICM [11], or more recently Gauss-Markov Measure Fields [12].

In the early work of Darrell and Pentland [13], simple multi–layered models (based on polynomials) are used to represent and segment smooth piecewise gray scale images or optical flow fields for regions containing multiple motions. More recently, other authors have reported different approaches toward or applications of layered models. In the work of Black et al. [14], models were proposed for computing the optical flow from a pair of images under changes of form, illumination, iconic content (those produced by pictorial changes in the scene) and specular reflections. Specifically, affine transformations or linear combinations of basis displacement fields model the motion in the regions. The model parameters and the membership weights (that define the support region for each model) are computed by using the Expectation–Maximization (EM) algorithm. Recently, Marroquin et. al. [15] introduce the MPM–MAP algorithm, an EM like algorithm. In which the classification is computed by approximating the MPM estimator using Gauss–Markov measure field models [12] and applying parametric models based on a finite element basis functions. The resultant models have greater flexibility for capturing the inherent geometry of the images.

Samson et al. [16] have reported a variational approach to classification (segmentation) and restoration, in which robust multimodal potentials are used to perform piecewise smooth restoration of images. The segmentation is automatically obtained by detecting the valleys (local minima) within which restored pixels lays. The minimization of the corresponding energy functional is achieved by means of an efficient deterministic algorithm. However, it is not clear, how spatially varying classes can be introduced, which may be required for represent, for example, illumination gradients.

This work builds on the fore mentioned methods and recasts them in a general variational framework, We develops a formulation for the segmentation and restoration problem using flexible models within a robust regularized network (RRN). Without loss of generality, the technique (based on an iterative algorithm) is presented in the context of image segmentation based on gray level intensity. In the first step of the iterative algorithm, an approximation of the classification is computed by using a local minimization algorithm. In the second step, the parameters of the RRN are updated. The use of robust potentials is necessary for two reasons, which will be developed in full in the remainder of the text:

1. To account for errors in the classification step that result from the use of a local minimizer.
2. To improve the fidelity of the segmentation to the data, the flexibility of the RN must to be adapted according to the local image gradient.

The plan of the paper is as follows. In the next section, we present a general framework for adaptive flexible models. The formulation of our models is based on Regularized Networks (RN) [17] with a robust potential for the data term and a quasi–robust potential for the regularization term [18]. Then, in subsection 2.2, the complete cost functional is described. Next, a general iterative minimization algorithm is proposed. In subsection 2.4, details of the implementation are discussed. Numerical experiments that demonstrate the performance of the method are presented in section 3. Finally, our conclusions are given in section 4.

2 Segmentation and Restoration with Flexible Models

In order to motivate our cost functional, first let us assume that the classification is given and it contains some errors as a product of noisy data or inexact models. The formulation of the energy functional for a specific class must therefore take into account that the data may be corrupted by outliers. Then, based on the introduced robust models, the complete functional for classification and restoration is presented. For minimizing the cost functional (which by construction is bounded by zero), we propose a deterministic iterative algorithm which alternates between two steps: updates of the classification (by means of a local minimization); and updates of the restoration (by means of a half-quadratic minimization scheme [20][21][22][23]).

2.1 Flexible Models Based on Robust Regularized Networks (RRN)

Without loss of generality, the algorithm is presented in the context of image segmentation based on gray intensity level. First, we note that an image can be viewed as an ensemble of regions each with smooth intensity gradient (i.e., we are assuming a piecewise smooth image model), so the parametric model that we use for representing the data in each region must be consistent with this

assumption. For representing the restored data ϕ that belongs to the class $\hat{c}$, we choose a model based on a regularized network [17]. Given a class $c = \hat{c}$, the associated restored data ϕ is computed as the minimizer of the cost functional:

$$R(\theta) = \sum_{r \in L: c_r = \hat{c}} \rho_1 \left(\phi_r \left(\theta_{\hat{c}}, \hat{c} \right) - g_r; k_1 \right) + \gamma P \left(\phi \left(\theta_{\hat{c}}, \hat{c} \right) \right), \tag{3}$$

where P is a regularization term over the model whose contribution to the total cost is controlled by the parameter γ, and $\phi_r \left(\theta_{\hat{c}}, \hat{c} \right)$ is the computed value for pixel r (that belongs to class $\hat{c}$) with corresponding parameter $\theta_{\hat{c}}$:

$$\phi_r \left(\theta_{\hat{c}}, \hat{c} \right) = \sum_j \theta_{\hat{c},j} \Phi_j \left(r \right), \tag{4}$$

where $\Phi_j \left(r \right)$ represents the j^{th} interpolation function evaluated at the pixel r whose contribution is controlled by the parameter $\theta_{\hat{c},j}$. ρ_1 is a robust potential function for outlier rejection and k_1 is the associated positive scale parameter [20][22][23][24][18][25]. In section 2.4 (Implementations Details), we specify the robust potential ρ_1 and ρ_2 and the form of the interpolation functions Φ. Examples of models with the form (4) include Radial Basis Functions (RBF) or Additive Splines [17][26][27]. We use a robust potential in the data term because the classification field c_r may be corrupted by outliers as a product of noisy data or inexact models. The classical (L_2) regularized network of RBF corresponds to choosing $\rho_1(x) = x^2$.

For choosing the form of the regularization term P, we take into account that:

1. The data term in (3) is only evaluated at the pixels that belong to class $\hat{c}$. Therefore, the regularization term P allows us to interpolate over missing data (those sites that do not belong to $\hat{c}$).
2. Regions with an inhomogeneous smoothness.

Therefore, we propose the following regularization term P in (3):

$$\begin{aligned} P \left(\phi \left(\theta, \hat{c} \right) \right) = & \sum_{r \in L: c_r = \hat{c}} \sum_{s \in N_r: c_s = \hat{c}} \rho_2 \left(\phi_r \left(\theta_{c_r}, c_r \right) - \phi_s \left(\theta_{c_s}, c_s \right); k_2 \right) \\ & + \mu \sum_{r \in L} \sum_{s \in N_r} \left(\phi_r \left(\theta_{\hat{c}}, \hat{c} \right) - \phi_s \left(\theta_{\hat{c}}, \hat{c} \right) \right)^2, \end{aligned} \tag{5}$$

where the first term in (5) uses a robust potential function ρ_2 (with scale parameter k_2) over the discrete version of the gradient of ϕ at those sites that belong to $\hat{c}$. This allows the model to adapt its stiffness according to the data. The second term implements a quadratic potential over the gradient of ϕ on the whole domain. The quadratic potential has two effects: performs a smooth interpolation (or extrapolation), and avoids an over–relaxation of the stiffness of the adaptive model. The trade–off between the two terms is controlled by the positive parameter μ.

2.2 Cost Functional for Restoration and Segmentation Based on Robust Regularized Network (RNN) Models

Given the flexible models, (3), (4) and (5), we can specify the complete cost functional for restoration and segmentation:

$$U_R(c,\theta) = \sum_{r\in L}\left[H(c_r,\theta_{c_r};C_r,\Theta_r) + \sum_{s\in N_r}\sum_{q\in C}\lambda_3\left(\phi_r\left(\theta_q,q\right)-\phi_s\left(\theta_q,q\right)\right)^2\right], \quad (6)$$

where $H(c_r,\theta_{c_r})$ is a local energy functional:

$$\begin{aligned} H(c_r,\theta_{c_r};C_r,\Theta_r) = {} & \rho_1\left(\phi_r\left(\theta_{c_r},c_r\right)-g_r;k_1\right) \\ & + \sum_{s\in N_r}\left[\lambda_1(1-\delta(c_r-c_s))\right. \\ & \left. +\lambda_2\rho_2\left(\phi_r\left(\theta_{c_r},c_r\right)-\phi_s\left(\theta_{c_r},c_r\right);k_2\right)\delta(c_r-c_s)\right], \end{aligned} \quad (7)$$

$C_r = \{c_s : s \in N_r\}$ and $\Theta_r = \{\theta_{c_s} : s \in N_r\}$ and λ_1, λ_2 and λ_3 redefine the regularization parameters λ, γ and μ.

Functional (6) was derived by assuming that the classification was given. Note that in order to compute the minimization with respect to the classification c (keeping θ fixed), we need to take into account an additional regularization term [weighted by λ_2 in (7)] that does not appear in the original functional (1). We can understand the effect of this new term as follows: the first term [in (7)] penalizes unconditionally the granularity of the classification (an over–segmentation), while the second term [in (7)] promotes the change of data model only if the gradient of the model ϕ is large enough (via the robust potential ρ_2).

2.3 Minimization Algorithm

The computation of the global minimum of (6) (the segmentation c^* and the parameters of the restoration defined by the set of models θ^*) corresponds to

$$\{c^*,\theta^*\} = \arg\min_{c,\theta} U_R(c,\theta). \quad (8)$$

An effective approximation is to alternatively minimize (6) with respect to c and θ by iterating between following two steps until convergence (given an initial guess for the model parameters θ_r^t and setting $t = 0$):

1. Update the segmentation c, keeping fixed the restoration θ^t:

$$c^{t+1} = \arg\min_{c\in C} U_R(c,\theta^t).$$

2. Update the restoration θ, keeping fixed the segmentation c^{t+1}:

$$\theta^{t+1} = \arg\min_{\theta} U_R(c^{t+1},\theta),$$

and set $t = t + 1$.

The combinatorial optimization implied in step 1 can be performed by stochastic algorithms (SA) such as simulated annealing or Gibbs sampling . However, it is well known that stochastic methods are extremely time consuming, and for this reason, deterministic approaches have been preferred (e.g., Iterated Conditional Modes, Mean Field Annealing, Label Relaxation, and others [1]) in spite of the fact that only the computation of a local minimum is guaranteed. The use of heuristics, such as multigrid algorithms, graduated non-convexity or the provision of a good initial guess, reduce the risk of being trapped by "bad" local minimum. Thus, good quality results can be computed at a fraction of the time than with SA algorithms. With this in mind, we present a deterministic algorithm for minimizing (6).

First, we note that for a given classification c, (6) is half–quadratic with respect to the restoration parameters θ [20][21][22][23]. Thus, in order to perform this minimization, we can use efficient algorithms (as reported in [18][23][28]) that guarantee convergence to a local minimum. Our iterative algorithm consists of the followings two steps:

1. Update the segmentation c (keeping fixed θ) by performing a local optimization using ICM:

$$c_r^{t+1} = \arg\min_{c_r \in C} H(c_r, \theta_r^t); \quad \text{for all} \quad r \in L,$$

   where H is given in (7).
2. Update the parameters of the models θ (keeping c fixed) by using a Half–Quadratic minimization:

$$\theta^{t+1} = \arg\min_{\theta} U_R(c^{t+1}, \theta),$$

   and set $t = t + 1$.

In practice, step 1 is performed more that once (in our experiments we performed three iterations), which guarantee that

$$U_R(c^{t+1}, \theta^t) \leq U_R(c^t, \theta^t).$$

On the other hand, the minimization indicated in step 2 is not completely achieved, in order to guarantee convergence it is only necessary that

$$U_R(c^{t+1}, \theta^{t+1}) \leq U_R(c^{t+1}, \theta^t),$$

thus, we performed only three iteration the weighted Gauss-Seidel algorithm [18][23].

2.4 Implementation Details

The use of the ICM algorithm for computing the segmentation step results in a significant computational time reduction. However, it also increases the number

of errors in the classification when compared to the use of a stochastic algorithm or other deterministic approaches, such as mean field theory based algorithms. Therefore, we introduce a potential term for outlier rejection needs to be used in the data term. On the other hand, we chose a quasi–robust potential for the regularization term [18] since we wanted the model to adapt its stiffness to smooth changes in the data, while at the same time disallowing internal discontinuities within regions—strong discontinuities must result in a change of the model.

The quasi–robust potentials $\rho(t)$ correspond to (in general) non-convex potentials that grow at a slower rate than quadratic ones, and have the following characteristics [18]:

1. $\rho(t) \geq 0 \; \forall t$ with $\rho(0) = 0$.
2. $\rho'(t) \equiv \partial \rho(t(f))/\partial f$ exists.
3. $\rho(t) = \rho(-t)$, the potential is symmetric.
4. $\frac{\rho'(t)}{2t}$ exists.
5. $\lim_{t \to +\infty} \frac{\rho'(t)}{2t} = m$ and $\lim_{t \to 0^+} \frac{\rho'(t)}{2t} = M$, $0 \leq m < M < +\infty$.

(9)

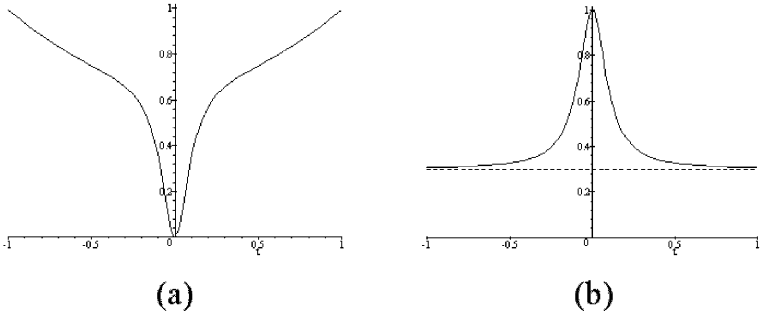

Fig. 1. (a) Potential function corresponding to the quasi–robust regularization potential $(1-\mu)t^2 + \mu\rho_2(t^2; k_2)$, and its (b) corresponding weight function.

Condition (9.1) establishes that a residual equal to zero produces the minimum cost. Condition (9.2) constrains ρ to be differentiable, so that one can use efficient deterministic algorithms for minimizing the cost function. Condition (9.3) constrains ρ to penalize equally positive and negative values of t. Finally, conditions (9.4)–(9.6) impose the outlier rejection condition (particularly, (9.5) corresponds to the quasi–robust condition [18], for $\mu = 0$ the potential is robust).

Specifically, our method implements the Geman–McClure robust potential [19]:

$$\rho(x;k) = \frac{(x/k)^2}{1+(x/k)^2}$$

for both the data and regularization potential (ρ_1 and ρ_2, respectively). We recall that the combination of a hard–redescending robust potential (such as ρ) and a quadratic one results in a quasi–robust potential (see figure 1).

We assume a smooth intensity variation for each class, so various parametric models are possible, as interpolation polynomials (for example, Hermite and Lagrange), radial basis functions (RBF), splines or as in [27][17], the function $|x|$. In this work, because of their computational efficiency, we use the finite support functions:

$$\Phi(x,y) = \begin{cases} \psi(x)\,\psi(y) \text{ for } |x|,|y| \leq 1 \\ 0 \qquad\qquad \text{otherwise,} \end{cases}$$

where two choices for ψ are employed:

1. First order Lagrange linear interpolation polynomials (that correspond to bilinear interpolation): $\psi_1(t) = 1 - |t|$, see figure 2(a).
2. The two–times differentiable Gaussian like function: $\psi_2(t) = 0.5 + 0.5\cos(\pi t)$, see figure 2(b).

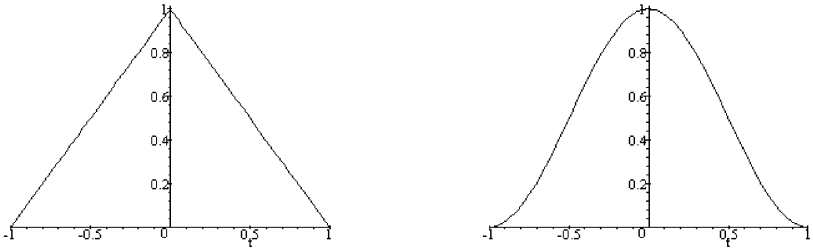

Fig. 2. Interpolation functions, *Phi*, corresponding to(a) $\psi_1(t) = 1 - |t|$ and (b) $\psi_2(t) = 0.5 + 0.5\cos(\pi t)$,.

3 Related Work

Segmentation algorithms are, in general, implemented, as iterative, two–step procedures of the following form:

1. Compute a classification keeping fixed the parameters of the models.
2. Compute the parameters of the models keeping fixed the classification.

The specific implementation of each step depends on the algorithm. It is evident that a more accurate classification is obtained when the models on which the classification is based are themselves accurately estimated. In this work we compute the model parameters using robust methods and adapt the values according to intensity variations in the image.

Ref. [15] presents an MPM–MAP segmentation method in which the classification step is computed by maximizing the posterior marginals using a Gauss–Markov random field (GMMF) [12]. A probability measure field representing the class membership of each pixel for every class is computed. The label field is then obtained by choosing the maximally probable class at each pixel. The advantage of the formulation is that the probability measure field can be computed using deterministic means, but at the cost of storing in memory the measure field for each class. The implementation in [15] uses a coarse membrane based on the finite basis functions (FE) to represents each class. Thus the parameters of the models are the nodal values of the FE model. The deformation energy of the membrane is used as the regularization term. The nodes of the FE–mesh are distributed in a regular lattice. Since linear interpolation functions are used, the model corresponds to the use of linear splines or ψ_1 in our RN formulation.

Specifically, the classification stage in the MPM–MAP was implemented as follows:

$$c_r = \arg\max_{q \in C} p_r^T(q), \tag{10}$$

where p^T is the result at the T^{th} iteration of the Gauss–Seidel homogeneous diffusion over the likelihood p^0 (see [12]):

$$p_r^{t+1}(q) = \frac{1}{|N_r|} \sum_{s \in N_r} p_s^t(q) \quad \text{for } t = 0, 1, ..., T-1; \tag{11}$$

where—by assuming Gaussian noise,—p^0 is the likelihood of the data g given the models ϕ:

$$p_r^0(q) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(\phi_r\left(\theta_q, q\right) - g_r\right)^2}{\sigma^2} \right), \tag{12}$$

where the variance, σ, is a parameter (that in general depends on the class, i.e., $\sigma = \sigma_q$). Then, in the MAP step the models $\phi_r\left(\theta_r, c_r\right)$ are updated by minimizing (3) with the regularization term:

$$P\left(\phi\left(\theta, c\right)\right) = \sum_{q \in C} \sum_{r \in L} \sum_{s \in N_r} \left(\phi_r\left(\theta_q, q\right) - \phi_s\left(\theta_q, q\right)\right)^2. \tag{13}$$

The parameters of the algorithm are the variance of the noise σ^2, the number of iterations T, and the smoothness of the FE-membrane γ.

In the next section, we present numerical experiments demonstrating that our formulation, based on a RRN, produces better results that the MPM–MAP method, even when the classification step is computed with a simple local optimization method (ICM algorithm).

4 Experiments

In this section, we show numerical experiments that demonstrate the performance of the presented method on the segmentation and restoration of real data.

The first set of experiments present a comparison of the proposed algorithm and the MPM–MAP method with flexible models. Specifically, the flexible model is identical for the two algorithms (but with a quasi-robust potential in the RN in our case). The task is to segment a noisy image containing bright and dark objects. We assume that the objects have a relatively large size and smooth variations in their gray level. Figure 3 shows the test data: (a) the original image (128×128 pixels and normalized to the interval $[0, 1]$), (b) the hand segmented mask (for comparison purposes); (c) the noisy test image generated by adding gaussian noise with σ equal to the 10% of the original dynamic range. Panel 3-(d) shows the "ideal" segmented data. Figure 4 shows the computed results, the top row shows the results of the MPM–MAP algorithm and the bottom row corresponds to the results of our method. The interpolation nodes of the RN were distributed in a regular 17×17 grid. For the case of the MPM–MAP the likelihood was computed using the true value of the variance of the noise ($\sigma = 0.1$) and the regularization parameter was set $\gamma = 15$. The smoothing for each measure field was performed by computing 10 iterations (T) of homogeneous diffusion. For the proposed algorithm, the parameters were $k_1 = 1$, $k_2 = 0.1, \lambda_1 = 0.24$, $\lambda_2 = 0.3$ and $\lambda_3 = 3$. Images in the first column of figure 4 correspond to the "restorations" computed with the two methods, while the second column shows the computed segmentation masks. The third column shows the masked data, and finally, the error maps for the computed segmentation masks are shown in the last column. The percent of misclassified pixels was 4.9% and 1.8% for the MPM–MAP and the proposed algorithm, respectively. Note that the error of the reported method is mainly located within a thin region close to the assumed "true border", where many of the pixels were difficult to classify manually. In contrast, MPM–MAP fails in areas where the gray levels of two regions are similar (for example, the bottom of the ball), while the proposed method takes better account of the gradients present in the image. Indeed, the error at the top of the ball is the result of a gradient in the gray level. Furthermore, the borders are not well located using MPM–MAP, because of the tendency of the algorithm to over–smooth shapes.

In order to explain the last result, we note that the Gauss–Markov random field (GMMF) approach to maximizing of the *a posteriori* marginals (MPM step in MPM–MAP) smooths the shape of regions. As a consequence, region borders do not coincide with the edges in the image. This effect is more noticeable for

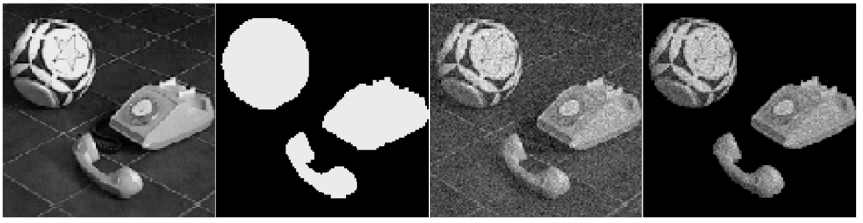

Fig. 3. Test data: (a) Original image, (b) manual segmentation mask, (c) noisy version of the original data, and (d) masked version of (c).

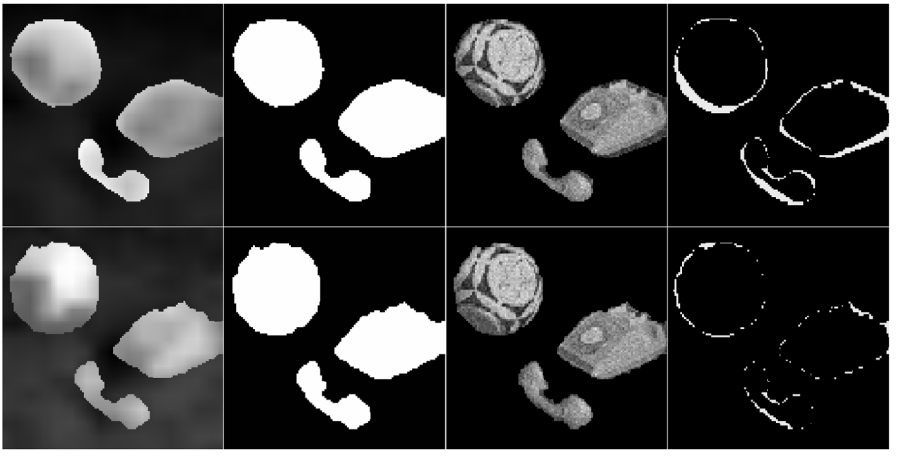

Fig. 4. Restoration–segmentation computed using two models. Results of the MPM–MAP (Top row): Restoration, Segmentation mask , Masked data using the segmentation mask and Errors in the mask. Second row: Results computed with the proposed method

large values of the regularization parameter in the GMMF algorithm. In figure 5, we illustrate this effect. The task consists of segmenting the image into three constant levels with simplified shapes (i.e., for a large value of the regularization parameter). Figure 5(a) shows the initial guess (maximum likelihood estimate for the MPM–MAP). Figure 5(b) shows the computed segmentation using MPM–MAP, and 5(c), the corresponding segmentation computed with the proposed method (specifically, functional (1) with a robust data term, given that we are using constant models).

Fig. 5. Segmentation with constant levels: (a) Initial guess (maximum likelihood estimate), (b) MPM–MAP and (c) proposed method.

5 Conclusion

We have presented a novel and general variational approach to segmentation and restoration based on the use of robust regularized networks (adaptive flexible models). The introduction of a robust potential in the data term allows us to use fast deterministic algorithms for estimating the segmentation (support region for each model), by compensating for outliers that typically confound such algorithm. Moreover, the quasi–robust potential in the regularization term allows us to adapt the stiffness of the flexible model according to the data image gradient, which results in more accurate segmentation along edges in the image.

Acknowledgments. This work was funded in part by the U.S.P.H.S. under grants NS33662, LM03504, MH62100, AG15116 and AG17586. M. Rivera was also supported in part by CONACYT, Mexico under grants 34575-A and Sabbatical stance.

References

1. Li, S.Z.: Markov Random Field Modeling in Image Analysis, Springer-Verlag, New York (2001)
2. Kirkpatrick, S., Gellatt, C. D., Vecchi, M.P.: Optimization by simulated annealing, Science, 220 (1983) 671–680
3. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions and Bayesian restoration of images, IEEE Trans. Pattern Anal. Machine Intell., 6 (1984) 721–741
4. Besag, J.: On the statistics analysis of dirty pictures (with discussions), Journal of the Royal Statistical Society, Series B, 48 (1986) 259–302
5. Peterson C., Soderberg, B.: A new method for mapping optimization problems onto neural networks, International Journal of Neural Systems, 1 (1989) 3–22
6. Geiger D., Girosi, F.: Parallel and deterministic algorithms for MRFs: Surface reconstruction and Integration, IEEE Trans. Pattern Anal. Machine Intell., 12 (1991) 401–412

7. Zhang, J.: Mean field theory in EM procedures for MRFs," IEEE Trans. Signal Processing, 40 (1992) 2570–2583
8. Heitz, F., Perez, P., Bouthermy, P.: Multiscale minimization of global energy functions in some visual recovery problems, Computer Vision, Graphics, and Image Processing: Image Understanding, 59 (1994) 125–136
9. Faugeras, O., Price, K.: Semantic description of aerial images using stochastic labeling, IEEE Trans. Pattern Anal. Machine Intell., 3 (1981) 638–642
10. Hummel, R.A., Zucker, S.W.: On the foundations of relaxation labeling process, IEEE Trans. Pattern Anal. Machine Intell., 5 (1983) 267–286
11. Li, S.Z., MAP image restoration and segmentation by constrained optimization, IEEE Trans. Image Processing, 9 (1998) 1134–1138
12. Marroquin, J.L., Velazco, F.A., Rivera, M., Nakamura,M.: Gauss–Markov Measure Field Models for low–level vision, IEEE Trans. Pattern Anal. Machine Intell., 23 (2001) 337–348
13. Darrell, T.J., Pentland, A.P.: Cooperative robust estimation using layers of support, IEEE Trans. Pattern Anal. Machine Intell., 17 (1995) 474–487
14. Black, M.J., Fleet D.J., Yacoob, Y.: Robustly estimating changes in image appearance, Comput. Vision Image Understand. 78 (2000) 8–31
15. Marroquin, J.L., Botello, S., Calderon F., Vemuri,B.C.: MPM-MAP for image segmentation, In Procc. ICPR 2000, Barcelona Spain, 4 (2000) 300–310
16. Samson, C., Blanc-Feraud, L., Aubert, G., Zerubia, J.: A variational model for image classification and restoration, IEEE Trans. Pattern Anal. and Machine Intell., 22 (2000) 460-472
17. Girosi, F., Jones, M., Poggio, T.: Regularization theory and neural networks architectures, Neural Computing, 7 (1995) 219–269
18. Rivera, M., Marroquin, J.L.: Efficient Half-Quadratic Regularization with Granularity Control, Technical Report: 05.06.2001, I-01-07 (CC), CIMAT, Mexico (2001)
19. Geman, S., McClure, D.E.: Bayesian image analysis methods: An application to single photon emission tomography, Proc. Statistical Computation Section, Amer. Statistical Assoc., Washington, DC, (1985) 12–18
20. Geman, D., Reynolds, G.: Constrained restoration and the recovery of discontinuities, IEEE Trans. Image Processing, 14 (1992) 367–383
21. Geman, D., Yang, C.: Nonlinear image recovery with half-quadratic regularization, IEEE Trans. Image Processing, 4 (1995) 932–946
22. Black, M.J., Rangarajan, A.: Unification of line process, outlier rejection, and robust statistics with application in early vision," Int. J. Comput. Vis., 19, (1996) 57–92
23. Charbonnier, P., Blanc-Féraud, L., Aubert, G., Barlaud, G.: Deterministic edge–preserving regularization in computer imaging, IEEE Trans. Image Processing, 6 (1997) 298-311
24. Rivera, M., Marroquin, J.L.: The adaptive rest condition spring model: an edge preserving regularization techinque, in Proc.Interational Conference on Image Processing, IEEE-ICIP 2000, Vol. II, Vancouver, BC, Canada (2000) 805–807
25. Rivera, M., Marroquin, J.L.: Adaptive Rest Condition Potentilas: Second Order Edge-Preserving Regularization, In Proc. European Conference on Computer Vision, ECCV 2002, Springer-Verlag, Copenhagen, Denmark, May (2002)
26. Girosi, F., Poggio, T., Caprile, B.: Extensions of a theory of networks for approximation of and learning: outliers and negative examples," In R. Lippmann, J. Moody and D. Tourezky, editors, Advances in Neural information processing systems 3, San Mateo, CA, Cambridge University Press (1991)

27. Girosi, F., Jones, M., Poggio, T.: Priors, stabilizers and basis functions: from regularization to radial, tensor and additive splines, MIT AI Lab Memo, No. 1430, June (1993)
28. Weickert, J.: Efficient image segmentation using partial differentiable equations and morphology, Pattern Recognition, 34 (2001) 1813–1824