ADAPTIVE DISCONTINUITY LOCATION IN IMAGE RESTORATION

Mário A. T. Figueiredo and *José M. N. Leitão*

Instituto de Telecomunicações, and
Departamento de Engenharia Electrotécnica e de Computadores.
Instituto Superior Técnico, Lisboa, PORTUGAL

ABSTRACT

Discontinuity-preserving Bayesian image restoration, based on Markov random fields (MRF), involves an intensity field, representing the image to be restored, and an edge (discontinuity) field. The usual strategy is to perform joint *maximum a posteriori* (MAP) estimation of the intensity and discontinuity fields, this requiring the specification of Bayesian priors. Departing from this approach, we interpret the discontinuity locations as deterministic unknown parameters of the intensity field. This leads to a parameter estimation problem with the important feature of having an unknown number of parameters. We introduce a discontinuity-preserving image restoration criterion (and an algorithm to implement it) based on the *minimum description length* (MDL) principle and built upon a *compound Gauss-Markov random field* (CGMRF) model; the proposed formulation does not involve the specification of a prior for the edge field which is adaptively inferred from the data.

1. INTRODUCTION

The incorporation of discontinuity/edge detection into image restoration/reconstruction techniques has been an important research area since edges are key features of visual perception [1], [2], [3]. Following previous work (e.g. [3], [4], [5]) on Markov random fields (MRF) for Bayesian image restoration, Jeng and Woods introduced the compound Gauss-Markov random field (CGMRF) model [6], [7]. This model (briefly reviewed in Section 2.1) allows for edge-preserving Bayesian restoration using a continuous (Gauss-Markov) *a priori* statistical model for the intensity field together with a discrete (binary) hidden edge field signaling discontinuities [6], [7]. To perform joint Bayesian estimation (e.g. MAP) of both the image and its edges, some prior model has also to be specified for the edge field. This prior is usually not explicitly stated; instead, a joint intensity-edge prior is directly considered [6], [7], [8], [9].

Our approach avoids the specification of the edge prior by adopting the following perspective: the location of the discontinuities are deterministic parameters of the intensity field probability density function. Locating discontinuities is then a parameter estimation problem exhibiting the important feature of having an unknown number of parameters (discontinuities). This places the problem in a class to which the *minimum description length* (MDL) principle (briefly reviewed in Section 2.2) has been successfully applied [10], [11], [12].

Pioneer work using MDL for image segmentation was done by Leclerc [14]; our work departs from Leclerc's in that we model the true image as a sample of an MRF while in [14] it is deterministically modelled as piecewise polynomial. The work [15] also uses MDL together with an MRF-type model; it pursues, however, a different goal in which MDL is used to build priors based on the *"complexity"* of the discontinuity configurations.

This paper introduces an MDL-based discontinuity-preserving image restoration criterion built upon a CGMRF model and not requiring the specification of a Bayesian prior for the edge field.

To implement the proposed criterion we apply a method related to the *expectation-maximization* (EM) algorithm [13]. Beyond doing without discontinuity related parameters (such as detection thresholds or penalties) the algorithm also autonomously estimates the variance of the observed image noise.

2. BACKGROUND

2.1. Compound Gauss-Markov Random Fields - CGMRF

Let $\mathbf{x} = \{x_{ij} : i = 1, 2, ..., M; j = 1, 2, ..., N\}$ be a realization of a *Gauss-Markov random field* (GMRF) $\mathbf{X} = \{X_{ij} : i = 1, 2, ..., M; j = 1, 2, ..., N\}$ defined on a $M \times N$ pixels image; its probability density function (pdf) is

$$p(\mathbf{x}) = \frac{|\det \mathbf{A}|^{1/2}}{(2\pi)^{(MN)/2}} \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}\right\}, \qquad (1)$$

where $\mathbf{x}$ here stands for a vector containing the lexicographically ordered pixel values, and $\mathbf{A}$ is the inverse of the covariance matrix, termed *potential matrix* [16]. Being Markovian, the conditional probabilities verify

$$p(x_{ij}|\{x_{kl},\ (k,l) \neq (i,j)\}) = p(x_{ij}|\{x_{kl}:\ (k,l) \in N_{ij}\}),$$

where N_{ij} is the neighborhood of pixel (i,j) [3]. In (1), the Gibbsian joint pdf of the MRF $\mathbf{X}$ can be recognized [5].

0-8186-6950-0/94 $4.00 © 1994 IEEE

Assume that only a noisy version $\mathbf{y} = \mathbf{x}+\mathbf{n}$ is observed, where $\mathbf{n}$ is a sample of a white Gaussian homogeneous noise field with variance σ^2, i.e.

$$p(\mathbf{y}|\mathbf{x},\sigma^2) = (2\pi\sigma^2)^{-MN/2} \exp\left\{-\frac{1}{2\sigma^2} \parallel \mathbf{y}-\mathbf{x} \parallel^2\right\}. \quad (2)$$

Compound Gauss-Markov random fields (CGMRF) are a special class of GMRF's in which the potential matrix is parametrized by a collection of binary (0 or 1) edge variables; these variables, when equal to 1, break the correlation between neighbor pixels thus allowing for edge preserving Bayesian restoration [6], [7], [8], [9]. This is in contrast with homogeneous GMRF's which do not take edges into account. Let $\mathbf{h}$ and $\mathbf{v}$ be the edge fields signaling, respectively, horizontal and vertical discontinuities. The pdf of a first order (N_{ij} is the set of four nearest neighbors of (i,j), i.e. $N_{ij} = \{(i,j-1),(i,j+1),(i-1,j),(i+1,j)\}$) CGMRF $\mathbf{X}$, given an edge configuration $(\mathbf{h},\mathbf{v})$, is

$$p(\mathbf{x}|\mathbf{h},\mathbf{v}) = \frac{1}{Z(\mathbf{h},\mathbf{v})} \exp\{-\frac{\mu}{2}\sum_{i,j}[\omega_v(1-v_{ij})(x_{ij}-x_{ij-1})^2$$
$$+\omega_h(1-h_{ij})(x_{ij}-x_{i-1j})^2 + (1-2(\omega_v+\omega_h))\, x_{ij}^2]\}, \quad (3)$$

where μ, ω_h, and ω_v are parameters; roughly, μ is the inverse of a "*global*" variance of the CGMRF, while ω_h and ω_v control the relative vertical and horizontal "*smoothness*" of the field.

Expression (3) can be written in vector notation as

$$p(\mathbf{x}|\mathbf{h},\mathbf{v}) = \frac{|\det \mathbf{A}(\mathbf{h},\mathbf{v})|^{1/2}}{(2\pi)^{(MN)/2}} \exp\left\{-\frac{1}{2}\,\mathbf{x}^T\mathbf{A}(\mathbf{h},\mathbf{v})\,\mathbf{x}\right\}, \quad (4)$$

where the dependence of matrix $\mathbf{A}(\mathbf{h},\mathbf{v})$ on parameters ω_h, ω_v and μ is not explicitly indicated. The factor multiplying the exponential in (4) is the normalizing constant, reciprocal of the *partition function* $Z(\mathbf{h},\mathbf{v})$ [3]; in this case, it depends on the edge configuration $(\mathbf{h},\mathbf{v})$.

Higher order models are obtained by assuming neighbors other than just the four nearest ones; here, without loss of generality, we only consider the first order case for which $\mathbf{A}(\mathbf{h},\mathbf{v})$ is block tridiagonal with tridiagonal blocks [20]. Equation (3) requires corrections at the image boundaries which we do not mention here.

The condition $(1-2(\omega_v+\omega_h)) > 0$ is sufficient for having $\det \mathbf{A}(\mathbf{h},\mathbf{v}) \neq 0$, for any $(\mathbf{h},\mathbf{v})$. In isotropic models, the only ones considered here, $\omega_v = \omega_h = \omega$ and the above condition is simply $\omega < 1/4$. Weak membrane type models (see e.g. [17]), lacking the $(1-4\omega)x_{ij}^2$ term, are non-normalizable since for most discontinuity configurations matrix $\mathbf{A}(\mathbf{h},\mathbf{v})$ is singular; $p(\mathbf{x}|\mathbf{h},\mathbf{v})$ becomes a so called *improper* prior [18]. For MAP estimation this is not a difficulty; however for parameter estimation, requiring the explicit use of $Z(\mathbf{h},\mathbf{v})$, they can not be used.

The joint MAP estimate of $\mathbf{x}$, $\mathbf{h}$ and $\mathbf{v}$, given $\mathbf{y}$, is

$$\left(\widehat{\mathbf{x}},\widehat{\mathbf{h}},\widehat{\mathbf{v}}\right)_{\text{MAP}} = \arg\max_{\mathbf{x},\mathbf{h},\mathbf{v}} \{p(\mathbf{x},\mathbf{h},\mathbf{v}|\mathbf{y})\}, \quad (5)$$
$$= \arg\max_{\mathbf{x},\mathbf{h},\mathbf{v}} \{p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x}|\mathbf{h},\mathbf{v})\, p(\mathbf{h},\mathbf{v})\}, (6)$$

where $p(\mathbf{h},\mathbf{v})$ is the prior probability function of the edge fields which has to be specified.

2.2. The Minimum Description Length Principle

The MDL principle, an information-theoretical concept proposed by Rissanen, allows the estimation of parameters along with their number [10], [11], [12]. It generalizes *maximum likelihood* (ML) estimation to cases where not only the parameters but also their number are unknown.

The ML estimate of a k-dimensional parameter vector $\Theta_{(k)}$, given observation $\mathbf{x}$, is defined as $\widehat{\Theta_{(k)}}_{\text{ML}} = \arg\max\{p(\mathbf{x}|\Theta_{(k)})\}$ (subscript (k) indicates that a vector is k-dimensional). Conventional ML is inadequate when k is unknown; in this case, the MDL principle stipulates that one should look for the shortest description (code length) of the data, which also includes the parameters themselves. The total length of the optimal code for $\mathbf{x}$, given a certain k-dimensional $\Theta_{(k)}$, is

$$L(\mathbf{x},\Theta_{(k)},k) = L(\mathbf{x}|\Theta_{(k)}) + L(\Theta_{(k)}) + L(k),$$

where $L(\mathbf{x}|\Theta_{(k)}) = -\log p(\mathbf{x}|\Theta_{(k)})$ (according to coding theory [11], [19]), $L(\Theta_{(k)})$ is the code length for a k-dimensional $\Theta_{(k)}$, and $L(k)$ is the code length for k itself (usually a constant). The MDL estimate of k and $\Theta_{(k)}$ is then (after dropping $L(k)$)

$$(\widehat{k},\widehat{\Theta}_{(k)})_{\text{MDL}} = \arg\min_{k,\Theta_{(k)}} \left\{-\log p(\mathbf{x}|\Theta_{(k)}) + L(\Theta_{(k)})\right\} \quad (7)$$

Notice that, if $L(\Theta_{(k)})$ only depends on k, then for fixed k the MDL and ML estimates coincide.

3. Proposed Formulation

3.1. Discontinuity Locations as Parameters

We interpret discontinuity locations as deterministic but unknown parameters of $p(\mathbf{x})$ rather than as a random field. This does not involve the specification of a Bayesian prior for the discontinuities since we are now dealing with a parameter estimation problem. A difficulty arises: not only the locations of the discontinuities, but also their number, are unknown. To solve this we adopt the MDL principle.

Notice that we can focus on estimating the discontinuities alone. With fixed $\mathbf{h}$ and $\mathbf{v}$, the joint *a posteriori* probability density function $p(\mathbf{x},\mathbf{h},\mathbf{v}|\mathbf{y}) \propto p(\mathbf{x}|\mathbf{h},\mathbf{v})\, p(\mathbf{y}|\mathbf{x})$ (considering (2) and (4)) is convex with respect to $\mathbf{x}$ and its maximizer is simply

$$\widehat{\mathbf{x}}(\mathbf{h},\mathbf{v}) = \left(\sigma^2\mathbf{A}(\mathbf{h},\mathbf{v}) + \mathbf{I}\right)^{-1}\mathbf{y}, \quad (8)$$

which can be obtained by, e.g., deterministic relaxation [17], [20]. This shows that the only difficulty lies in estimating the discontinuities.

Let $\Theta_{(k)} = [\theta_1,\theta_2,...,\theta_k]^T$ be a k-dimensional parameter vector expressing the locations of the discontinuities; $\Theta_{(k)}$ is just a compact description of $\mathbf{h}$ and $\mathbf{v}$; if there are few discontinuities, when compared to the image dimension, it is shorter to specify their locations than to use $\mathbf{h}$ and $\mathbf{v}$ explicitly. Each component of $\Theta_{(k)}$ is a triplet (i,j,b) reporting that $h_{ij}=1$, if $b=1$, or that $v_{ij}=1$, if $b=0$:

$$\begin{cases} (h_{ij}=1) & \Leftrightarrow\quad \exists_{\theta_n=(i,j,1)},\quad n=1,2,...,k \\ (v_{ij}=1) & \Leftrightarrow\quad \exists_{\theta_n=(i,j,0)},\quad n=1,2,...,k. \end{cases}$$

Since $i \in \{1, ..., M\}$, $j \in \{1, ..., N\}$, and $b \in \{0, 1\}$,

$$L(\Theta_{(k)}) = k(\log M + \log N + \log 2) = k \log(2MN) \quad (9)$$

is the natural choice.

3.2. An MDL Criterion with Incomplete Data

If $\mathbf{x}$ was observed, MDL could be directly used by introducing (4) into (7). Since only $\mathbf{y}$ is observed we use $p(\mathbf{x}, \mathbf{y}|\Theta_{(k)})$ instead of $p(\mathbf{x}|\Theta_{(k)})$ and interpret the true image $\mathbf{x}$ as missing data [13]. The obtained criterion is (with explicit dependence on σ^2, which we also wish to estimate)

$$(\widehat{k}, \widehat{\Theta}_{(k)}, \widehat{\sigma^2})_{\text{MDL}} = \arg \min_{k, \Theta_{(k)}, \sigma^2} L(\mathbf{x}, \mathbf{y}, \Theta_{(k)}, \sigma^2), \quad (10)$$

where

$$\begin{aligned} L(\mathbf{x}, \mathbf{y}, \Theta_{(k)}, \sigma^2) &= -\log p(\mathbf{x}, \mathbf{y}|\Theta_{(k)}, \sigma^2) + L(\Theta_{(k)}) \\ &= -\log p(\mathbf{y}|\mathbf{x}, \sigma^2) - \log p(\mathbf{x}|\Theta_{(k)}) \\ &\quad + L(\Theta_{(k)}). \end{aligned} \quad (11)$$

The length $L(\sigma^2)$ is not included in (11) since it is constant (with respect to k); in other words, relatively to σ^2, (10) is an ML criterion. Following the equivalence between $(\mathbf{h}, \mathbf{v})$ and $\Theta_{(k)}$, function $p(\mathbf{x}|\Theta_{(k)})$ is as given by (4)

3.3. The Pseudo-likelihood Approximation

Before addressing the issue of how to find the minimum of the description length function (11), notice that there is a huge difficulty in computing it. This obstacle lies in $p(\mathbf{x}|\Theta_{(k)})$, as given by (4), which involves the determinant of an enormous matrix ($MN \times MN$ for a $M \times N$ image). This is the well known problem of computing the partition function, arising in parameter estimation with MRF models. Here we resort to Besag's pseudo-likelihood approximation [5] which, omitting the parameters, states that

$$p(\mathbf{x}) \simeq \prod_{ij} p(x_{ij}|\{x_{kl} : (k, l) \in N_{ij}\}); \quad (12)$$

recall we are dealing with an MRF.

Since X is a Gaussian MRF the conditional probabilities are also Gaussian,

$$p(x_{ij}|\{x_{i\,j-1}, x_{i\,j+1}, x_{i-1\,j}, x_{i+1\,j}\}, \Theta_{(k)}) = \mathcal{N}(\eta_{ij}, \lambda_{ij}^2) \quad (13)$$

where $\mathcal{N}(\nu, \psi^2)$ denotes a Gauss function with mean ν and variance ψ^2. The local mean and variance which depend on $\Theta_{(k)}$ (i.e. on $\mathbf{h}$ and $\mathbf{v}$) are:

$$\begin{aligned} \eta_{ij}(\Theta_{(k)}) = \lambda_{ij}^2(\Theta_{(k)})\, \mu\omega \big[&\overline{h}_{ij}\, x_{i-1\,j} + \overline{v}_{ij}\, x_{i\,j-1} + \\ &\overline{h}_{i+1\,j}\, x_{i+1\,j} + \overline{v}_{i\,j+1}\, x_{i\,j+1}\big], \end{aligned} \quad (14)$$

where $\overline{h}_{ij}$ stands for $(1 - h_{ij})$;

$$\lambda_{ij}^2(\Theta_{(k)}) = \frac{1}{\mu[1 - \omega(h_{ij} + h_{i+1\,j} + v_{ij} + v_{i\,j+1})]}. \quad (15)$$

Finally, introducing (9), (2), and the pseudo-likelihood approximation (12) (together with (13), (14) and (15)) into (11) leads (dropping additive constants) to

$$\begin{aligned} L(\mathbf{x}, \mathbf{y}, \Theta_{(k)}, \sigma^2) &\simeq \\ &2k \log(2MN) + \mu \sum_{ij} \frac{(x_{ij} - \eta_{ij}(\Theta_{(k)}))^2}{\lambda_{ij}^2(\Theta_{(k)})} \\ &+ \sum_{ij} \log \lambda_{ij}^2(\Theta_{(k)}) + \frac{1}{\sigma^2} \sum_{ij} (x_{ij} - y_{ij})^2. \end{aligned} \quad (16)$$

4. Algorithm

To minimize the description length function (16) we have devised an EM-type scheme. The EM nature of the algorithm is justified by the fact that $\mathbf{x}$ is missing and so we are facing a problem of parameter estimation from incomplete data [13]. The fast (albeit suboptimal) algorithm herein considered restricts the solutions, with respect to the discontinuities, to those obtained by comparing all the pairwise differences with a varying threshold. A range of thresholds is then swept and the MDL criterion is used to select the optimal one. The algorithm runs as follows:

Step 0 Initialization: set $k = 0$ and an initial value for the noise variance estimate $\widehat{\sigma^2}$; get an initial estimate $\widehat{\mathbf{x}}$ using (8). From these, compute and store the corresponding description length value. Also, initialize a threshold parameter denoted γ to some high value.

Step 1 From the current estimate $\widehat{\mathbf{x}}$, and the observed image $\mathbf{y}$, update the ML estimate $\widehat{\sigma^2}$ according to

$$\widehat{\sigma^2} = \frac{1}{MN} \sum_{ij} (x_{ij} - y_{ij})^2.$$

Step 2 From the current estimate $\widehat{\mathbf{x}}$, update the discontinuity estimates $(\widehat{\mathbf{h}}, \widehat{\mathbf{v}})$ according to

$$\begin{cases} \widehat{h}_{ij} = 1 & \Leftrightarrow \quad |x_{ij} - x_{i-1\,j}| > \gamma \\ \widehat{v}_{ij} = 1 & \Leftrightarrow \quad |x_{ij} - x_{i\,j-1}| > \gamma, \end{cases}$$

where γ is the current value of the threshold parameter. Update $\widehat{k}$ by counting the number of signaled discontinuities.

Step 3 From the recently updated $(\widehat{\mathbf{h}}, \widehat{\mathbf{v}})$ and $\widehat{\sigma^2}$, obtain a new image estimate according to (8).

Step 4 From $(\widehat{\mathbf{h}}, \widehat{\mathbf{v}})$, $\widehat{k}$, $\widehat{\mathbf{x}}$, and $\widehat{\sigma^2}$, compute and store the description length (16).

Step 5 Decrease parameter γ by some predefined amount; if it has not reached a (also predefined) lower limit go back to **Step 1**, otherwise go to **Step 6**.

Step 6 Final step: from the stored list of description lengths (see **Step 4**) find the minimum value and elect the respective image and discontinuity estimates as the final results.

Observe the similarity of **Step 3** to the E-step of the EM algorithm (since **X** is Gaussian and the observation mechanism considers additive Gaussian noise the MAP estimate coincides with the conditional expectation). The role of the M-step is here played by **Step 1** and **Step 2**.

Since, in practice, we deal with digital images, i.e. their pixels only assume integer gray levels, say $0, 1, ..., 255$, an important fact can be exploited: the smallest possible non-zero difference is 1 and the largest is 255. So, threshold γ can be swept from 255 down to 1 in unit steps. Moreover, a rough upper bound on the pixel differences (much better than 255) is easy to obtain and use as a starting point.

Finally, we call the attention to the fact that the developed algorithm is a first approach, putting in evidence the global features of the criterion introduced. It searches only a small subset of all possible configurations and a more sophisticated scheme will be necessary to fully exploit the potential of the proposed formulation.

5. Experimental Examples

Two experimental examples illustrate the behavior of the proposed technique. Both use 256 levels digital images and assume $\mu = 2.0$ and $\omega = 0.249$.

5.1. Synthetic Image

This example considers the synthetic image of Fig. 1(a) and its noisy version ($\sigma^2 = 30^2$) of Fig. 1(b). Fig. 2 shows a sequence of discontinuity configurations corresponding to the iteration numbers indicated, i.e. obtained for successively lower values of discontinuity detection threshold γ. Fig. 3 plots the evolution of the description length based on which the configuration of iteration 30 (see Fig. 2) was chosen as the solution. At this iteration, the noise standard deviation estimate was $\widehat{\sigma} = 30.8$.

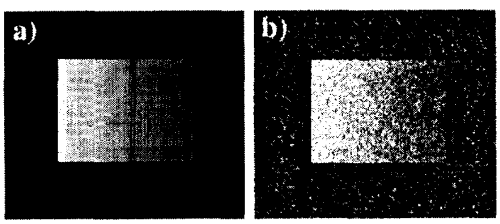

Figure 1: **Synthetic image (a); noisy ($\sigma^2 = 30^2$) version (b).**

5.2. Natural Image

A natural image (Fig. 4(a)) is now considered; its noisy version ($\sigma^2 = 40^2$) is shown in Fig.4(b). A sequence of discontinuity configurations is presented in Fig. 5, while Fig. 6 plots the evolution of the description length based on which configuration 32 was elected. At this iteration, the noise standard deviation estimate was $\widehat{\sigma} = 41.4$.

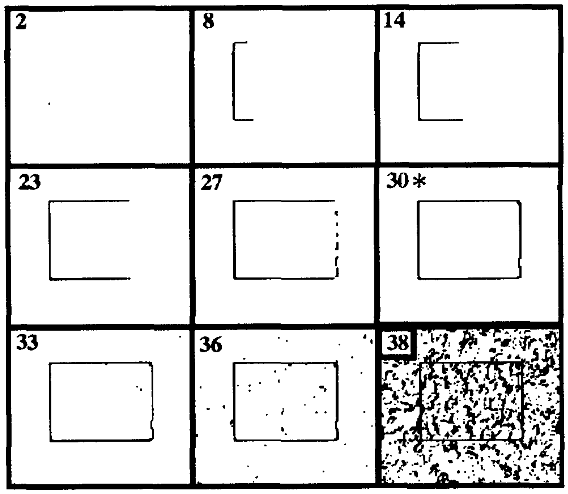

Figure 2: **Edge configurations at the iterations indicated by the numbers; the chosen one (see Fig. 3) is signaled by "*".**

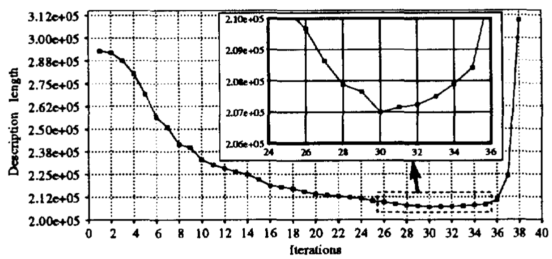

Figure 3: **Evolution of the description length; the minimum is achieved in iteration 30 as is clear in the detailed view contained in the inset.**

References

[1] A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, M.A., 1987.

[2] D. Terzopoulos, "Regularization of Inverse Visual Problems Involving Discontinuities", *IEEE Trans. on Patt. Anal. and Mach. Intell.*, vol. 8, pp. 413–424, 1986.

[3] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images", *IEEE Trans. on Patt. Anal. and Mach. Intell.*, vol. 6, pp. 721–741, 1984.

[4] J. Woods, "Two-Dimensional Discrete Markovian Fields", *IEEE Trans. on Inf. Theo.*, vol. 18, pp. 232–240, 1972.

[5] J. Besag, "On the statistical analysis of dirty pictures", *Jour. of the Royal Stat. Soc. (B)*, vol. 48, pp. 259–302, 1986.

[6] F. Jeng and J. Woods, "Image estimation by stochastic relaxation in the compound Gaussian case", *Proceedings of ICASSP'88*, pp. 1016–1019, New York, 1988.

[7] F. Jeng and J. Woods, "Compound Gauss-Markov random fields for image estimation", *IEEE Trans. on Sig. Proc.*, vol. 39, pp. 683–697, 1991.

[8] T. Simchony, R. Chellappa, and Z. Lichtenstein, "Graduated nonconvexity algorithm for image estimation using compound Gauss Markov fields", *Proceedings of ICASSP'89*, pp. 1417–1420, Galsgow, 1989.

[9] J. Zerubia and R. Chellappa, "Mean field approximation using compound Gauss-Markov random field for edge detection and image restoration", *Proceedings ICASSP'90*, pp. 2193–2196, Albuquerque, 1990.

[10] J. Rissanen, "Minimum-description-length principle", *Encyclopedia of Statistical Sciences, Vol. 5*, pp. 523–527, John Wiley & Sons, New York, 1987.

[11] J. Rissanen, *Stochastic Complexity in Stastistical Inquiry*, World Scientific, Singapore, 1989.

[12] J. Rissanen, "Universal coding, information, prediction, and estimation", *IEEE Trans. on Inf. Theo.*, vol. 30, pp. 629–636, 1984.

[13] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood estimation from incomplete data via the EM algorithm", *Jour. of the Royal Stat. Soc. (B)*, vol. 39, pp. 1–38, 1977.

[14] Y. Leclerc, "Constructing simple stable descriptions for image partitioning", *Intern. Jour. of Comp. Vision*, vol. 3, pp. 73–102, 1989.

[15] K. Keeler, "Map representations and coding-based priors for segmentation", *Proceedings of CVPR'91*, pp. 420–425, Hawaii, 1991.

[16] J. Moura and N. Balram, "Recursive structure of noncausal Gauss-Markov random fields", *IEEE Trans. on Inf. Theo.*, vol. 38, pp. 334–354, 1992.

[17] D. Geiger and F. Girosi, "Parallel and deterministic algorithms from MRF's: Surface reconstruction", *IEEE Trans. on Patt. Anal. and Mach. Intell.*, vol. PAMI-13, pp. 401–412, 1991.

[18] J. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York, 1980.

[19] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 1991.

[20] M. Figueiredo and J. Leitão, "Simulated tearing: an algorithm for discontinuity preserving visual suraface reconstruction", *Proceedings of CVPR'93*, pp. 28–33, New York, 1993.

Figure 4: **Natural image (a); noisy ($\sigma^2 = 40^2$) version (b).**

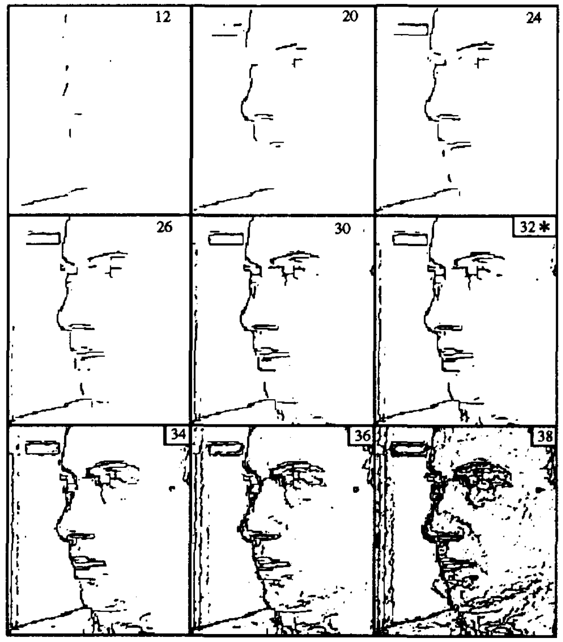

Figure 5: **Edge configurations at the iterations indicated by the numbers; the chosen one (see Fig. 6) is signaled by "*".**

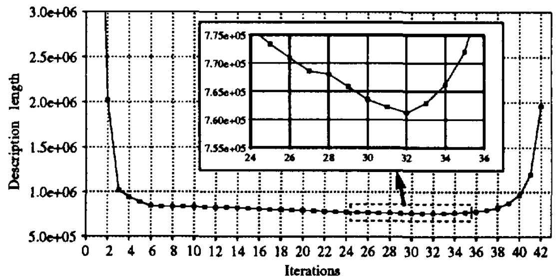

Figure 6: **Evolution of the description length; the minimum is achieved in iteration 32 as is clear in the detailed view contained in the inset.**

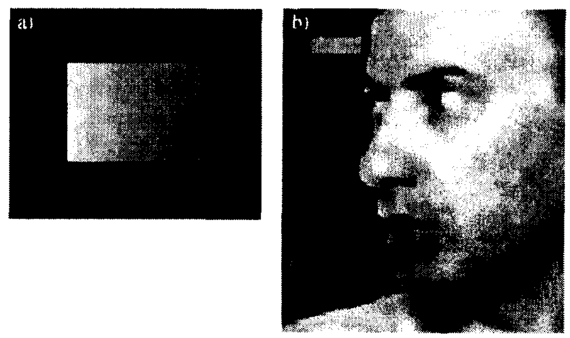

Figure 7: **Restored images: (a) from the synthetic image, and (b) from the natural image.**