Simulated Annealing for Restoration of NARMA Distorted Images

Michał Jaworski and Andrzej Pacut
Instytut Automatyki i Informatyki Stosowanej
Politechniki Warszawskiej*

1 Introduction

The image reconstruction problem arises in many applications. Each of them can be specific due to various mechanisms of distortion and different nature of the images. For instance, distortions of ultrasound images are related to a conical shape of the ultrasound ray, reflections, etc., while for vision images the distortions are caused by various imperfections of the vision systems. The presented method of image reconstruction can be applied to distortions described by a two-dimensional nonlinear ARMA model. Two related images: the original image and the distorted image, are represented here by random fields. For the image reconstruction we apply a methodology of simulated annealing which consists of creating a Markov chain of images convergent to the maximum a posteriori estimator which represents the reconstructed image.

The restoration problem is stated in Section 2. The basic algorithm is described in Sections 3 and 4, and the examples of image reconstruction for typical distortions are presented in Section 6.

2 Image Modeling and Maximum Aposteriori Estimation

The description of the image used in this paper is borrowed from [1]. We assume that each image x consists of a set of pixels $x_s, s \in S$ arranged in a finite rectangular lattice

$$S = \{(i,j) : 1 \leq i \leq M, 1 \leq j \leq N\} \tag{1}$$

Each pixel can take only finite number of values, $x_s \in \mathcal{X}_0$, the finiteness assumption being well justified for digital images. The images can be thus thought as points in $\mathcal{X} = \mathcal{X}_0^M \times \mathcal{X}_0^N$. Because of the random model of distortions, pixels may be regarded as random variables $X_s, s \in S$, taking values in $\mathcal{X}_0$ hence then the image $X = X_s, s \in S$ is a multidimensional random variable with the state space $\mathcal{X} = \mathcal{X}_s^M \times \mathcal{X}_s^N$. A strictly positive distribution Π defined on the set of all possible images $\mathcal{X}$ is called the *random field*. Each random image X is thus assigned a positive probability Π which is determined by its *local characteristics* i.e. by the collection of conditional probabilities of the form:

*This paper was supported by the State Committee for Scientific Research (KBN) grant 8T11A 023 11.

$$\Pi(X_s = x_s \mid \{X_r = x_r,\ r \in S,\ r \neq s\}), \quad s \in S \tag{2}$$

According to Hammersly-Clifford Theorem [1], every random field Π can by written in the *Gibbs form*, namely

$$\Pi(x) = \frac{\exp(-H(x))}{\sum_{z \in \mathcal{X}} \exp(-H(z))} \tag{3}$$

where the *energy function* H is strictly positive for each $x \in X$. The denominator of (3) is called the *partition function.* The Gibbs form for fields with only local dependencies between pixels (i.e. the dependencies bounded to small neighborhoods) is a starting point for various image processing algorithms. For each $s \in S$, a set $\partial(s)$ is called the *neighborhood* of s if

$$s \notin \partial(s) \tag{4}$$

$$s \in \partial(t) \Leftrightarrow t \in \partial(s) \tag{5}$$

The neighborhood of s completed with s is denoted by $\overline{\partial}(s) = \partial(s) \cup \{s\}$. The collection ∂ of neighborhoods $\partial(s)$ for each $s \in S$ is called *the neighborhood system*, namely

$$\partial = \{\partial(s), s \in S\} \tag{6}$$

One can express the local dependencies between points through the conditional probabilities of the form (2) i.e.

$$\Pi(X_s = x_s \mid \{X_r = x_r,\ r \in S,\ r \neq s\}) = \Pi(X_s = x_s \mid \{X_r = x_r,\ r \in \partial(s)\}) \quad s \in S \tag{7}$$

We use the Gibbs field on the local neighborhood system as the model of dependencies between the pixels. Properties of the Gibbs formula are exploited also to build the restoration algorithm. The energy function (3) is more convenient and natural mechanism for embodying the picture attributes than the local characteristics (2). Let $\mathcal{X}_C$ denotes the set of images which maximize Π (or minimize H). This set can be thought as a prior knowledge about the processed images. This prior can be embedded in Π before the restoration. The restoration procedure should also incorporate the observation of the distorted image y. In the approach described below we look for the maximum a posteriori (*MAP*) estimator of x defined as

$$\hat{x}_{MAP} = \arg\max_{x \in \mathcal{X}} \Pi(x \mid y) \tag{8}$$

Search for the images which maximize $\Pi(x \mid y)$ is equivalent to looking for the images which minimize the *aposteriori energy function* $H(x \mid y)$.

3 Simulated Annealing Algorithm

The restoration process can be considered as a recursive search for an image $x \in \mathcal{X}$ related to the lowest energy H. The use of the exhaustive search in this goal is computationally prohibitive because of the large dimension of X, many local minima, and flatness of the energy function. The partition function in (3) depends upon all possible images from $\mathcal{X}$ hence the direct usage of (3) is practically impossible.

On the other hand, one can easily compute the local characteristics (7) using the Gibbs expression. This property is used in the *simulated annealing algorithm* which consists of building a Markov chain of images Z^k, $k = 1, \ldots, \infty$ whose distributions Γ_k converge to the Gibbs distribution. Such a chain of images can be defined in various ways [1]. The main idea of *Metropolis Algorithm* used here in experiments is as follows:

The image $x(k+1)$ is obtained by a random update of a pixel $x_s(k)$ chosen randomly from the image $x(k)$. If $H(x(k+1)) \leq H(x(k))$ then $x(k+1)$ is accepted, otherwise it is accepted with probability $\exp(H(x(k)) - H(x(k+1)))$.

The subsequent pictures differ then only by a single pixel. The energy of consecutive images can locally increase to protect against settling at the local minima of H. To make the chain converge to the global minimum of $H(x)$ one introduces a Gibbs field equipped with a parameter T called the *temperature* or with a parameter $\beta = 1/T$ called the *inverse temperature* can be introduced by a modification of (3) to

$$\Pi(x) = \frac{\exp(-\beta H(x))}{\sum_{z \in \mathcal{X}} \exp(-\beta H(z))} \tag{9}$$

Suppose that the inverse temperature increases according to the *basic annealing schedule* given by

$$\beta(n) \leq \frac{1}{c\Delta(n)} \ln(n) \tag{10}$$

where n is the iteration number (during one iteration all pixels should be updated), $\Delta(n)$ is the biggest energy fluctuation during iteration $n-1$, and c is a parameter. One can prove [1], [4] that the limit distribution of such defined chain is of the form

$$\Gamma_\infty(x) = \lim_{\beta \longrightarrow \infty} \Pi^\beta(x) = \begin{cases} \frac{1}{|M|} & x \in M \\ 0 & x \notin M \end{cases}$$

where M is the set of global minima of H. Unfortunately, the basic annealing schedule makes the annealing process quite slow. We address this problem in Sec. 6.

4 Extended Field and Prior Constraints

We assume that the images have the following property:

The whole image can be divided into subsets of pixels which have "similar" values.

To express this prior in the energy terms one may introduce the family of *edge elements*. The edge elements surround each pixel (Fig. 1) and can take values 0 or 1 which correspond to the absence or the presence of the edge between the appropriate pixels. The set of edges for a single image can be thought as a vector b from the space of all possible picture edges. Similarly as for pixels, the edge elements are described as random variables B_r taking values in $\{0, 1\}$, and the whole edge is described by the multidimensional random variable B. To treat the pixels and the edges uniformly, one may introduce an *extended image* $x = \{x_s, s \in \bar{S}\}$ which consists of both the pixels

$$x_s, \quad s \in S = \{(\bar{\imath}, \bar{\jmath}) : \bar{\imath} = 2\,i,\ \bar{\jmath} = 2\,j,\ 1 \leq i \leq M,\ 1 \leq j \leq N\} \tag{11}$$

and the edge elements

$$b_s = x_s, \quad s \in R = \bar{S} - S \tag{12}$$

One may consider a field over the set of all possible extended images, called later the extended field. Let us consider the neighborhood system for the extended field. The neighborhood system ∂ for pixels is given by:

$$\partial(s) = \{r : 0 < d(s,r) \leq 2\sqrt{2}\}$$

where $d(s,t)$ denotes the distance between pixels s and t. Let the set ∂_b be the neighborhood system for edge elements, namely

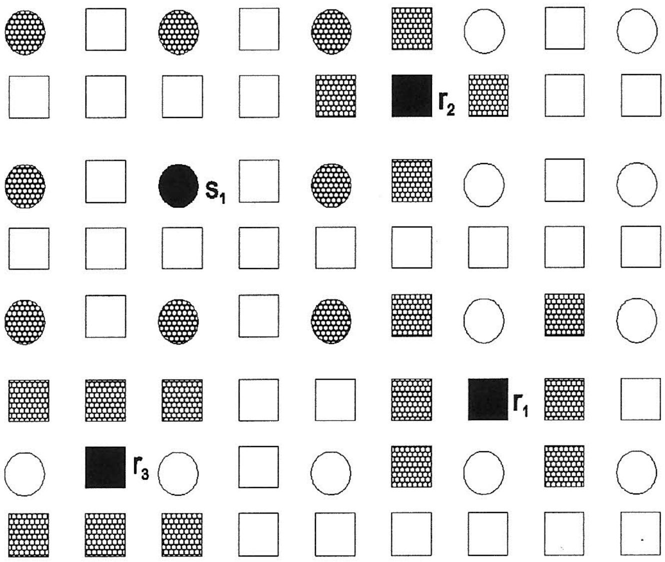

Figure 1: The extended field built from pixels (circles) and edge elements (squares). Dashed circles denote the neighborhood of pixel s_1. Dashed squares denote neighborhoods of the edge elements r_1, r_2, r_3

$$\partial_b(b) = \{a : 0 < d(b,a) \leq \sqrt{2}\}$$

where a, b denote edge elements (see Fig. 1). The energy function of the extended field describes the "degree of membership" of the image to the class $\mathcal{X}_C$

$$H(x,b) = \sum_s \sum_{t \in \partial(s)} (x_s - x_t)^2 (1 - b_{\{s,t\}}) \tag{13}$$

where $b_{\{s,t\}}$ is equal 1 if there is an edge element between s and t and is equal to 0 otherwise. The smaller values of the energy function the better is the fulfillment of the prior assumptions. As in the case of pixels, one can consider specific prior constraints to shape the edges properly. Those constraints may vary during the restoration, for instance in the early stage of the annealing the lack of edge continuity is natural but later the stress on that requirement can be increased. This leads to a nonstationary energy function. Let $\|\partial_b(r)\|$ be equal to the number of nonzero edge elements in the neighborhood of r, i.e.

$$\|\partial_b(r)\| = \sum_{t \in \partial_b(r)} b_t$$

We assume that an active edge element which has less than 2 neighbors which have value 1 is of small probability. On the other hand, an edge element having too many active edge neighbors would lead to many crossing lines. Thus we do not "penalize" only those active elements which have exactly two active edge neighbors. Let H_1 denote a term of the energy function describing the above dependencies i.e. for $b_r = 1$

$$H_1(b_r) = \begin{cases} 0 & \text{if } \|\partial_b(r)\| = 2 \\ \gamma_1 & \text{otherwise} \end{cases} \tag{14}$$

where γ_1 is a constant. Using the notation

$$\delta_r = \begin{cases} 1 & \text{if } \|\partial_b(r)\| = 2 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

one can express H_1 as

$$H_1(b_r) = \gamma_1(1 - \delta_r) \tag{16}$$

We put additional stress on the creation of closed lines by penalizing inactive edge elements having two active neighbors i.e. for $b_r = 0$

$$\begin{aligned} H_2(b_r) &= \begin{cases} \gamma_2 & \text{if } \|\partial_b(r)\| = 2 \\ 0 & \text{otherwise} \end{cases} \\ &= \gamma_2\, \delta_r \end{aligned} \tag{17}$$

where γ_2 is a constant. The edges energy varies during the restoration i.e. $\gamma_1 = \gamma_1(n)$ i $\gamma_2 = \gamma_2(n)$, where n is the iteration number. The total energy of extended field is the sum of terms (13), (16) and (17), namely

$$H(x, b) = \beta \sum_{s,t \in \partial(s)} (x_s - x_t)^2 \left(1 - b_{\{s,t\}}\right) + \sum_{r \in B} (\gamma_1(n)(1 - \delta_r)b_r + \gamma_2(n)\delta_r(1 - b_r)) \tag{18}$$

where β is a parameter and n is the iteration number.

5 Degraded Image

The restoration algorithms exploits the knowledge about the possible nature of distortions. This is often done by choosing a typical distortion structure and estimating its parameters on the base of $y \in \mathcal{Y}$, $x \in \mathcal{X}$ or $y \in \mathcal{Y}$ [6]. We consider two typical distortions: the first is random and is called the *noise*, and the second is deterministic and is called the *blur* [1]. We assume that both of them can be described by a function of "local" domain

$$y_s = \Xi(\Psi(\overline{\partial}(s)), \eta_s) \tag{19}$$

where

- y_s, $s \in S$ and x_s, $s \in S$ are pixels of the degraded and the original images, resp.,

- Ψ is the blurring function whose domain x is restricted to the neighborhood $\overline{\partial}(s)$ of s.
- The field η represents the random distortions with moments μ and σ^2. Random variables η_s, $s \in S$ are i.i.d. and are independent from the field X,
- Function Ξ is invertible with respect to η_s and the inverse $\eta_s = \Xi^{-1}(y_s, \Psi(\overline{\partial}(s)))$ is smooth.

The two-dimensional nonlinear ARMA transformation of the image satisfies the above conditions and is used as the basic model of image distortions. As it is known from Geman and Geman Theorem ([4]), the aposteriori energy $H(x, b \mid y)$ is related to the prior energy $H(x, b)$ by

$$H(x,b \mid y) = H(x,b) + \sum_{s \in S} \left(\frac{\left\| \mu - \Xi^{-1}(y_s, \Psi(\overline{\partial}^x(s))) \right\|}{2v} - \ln \frac{\partial}{\partial y_s} \Xi(y_s, \Psi(\overline{\partial}(s))) \right) \tag{20}$$

In the case of additive noise η the last factor in (20) vanishes. The minimization of (20) leads to the *MAP* estimator of the original image given the observation y.

6 Acceleration Methods

The classical simulated annealing described in Section 3 converges to the global minimum of energy but is quite slow. There are several ways to speed up the annealing which lead to suboptimal solutions. Fast cooling is the most popular one. The typical fast cooling schedules can be found in [1] and [2]. The second way is the parallel processing, which consists of updating the set of pixels or even the whole image at one time instant. The chain of images obtained in this way has its limit distribution but it may substantially differ from the Gibbs distribution. Parallel algorithms are difficult in theoretical analysis but the development of parallel computers stimulates the attention paid to these problems [1].

In the experiments described below we assured a balance between the parallel and the serial processing by gradual switching from parallel to serial processing. Early in the reconstruction process the whole image is processed parallelly, then the image is split into four subsets. The subsets are processed serially but its elements are processed parallelly. The split procedure is repeated and eventually all pixels are processed serially.

One can also extract from the whole image the sets of "interesting" pixels. The sets which do not fulfill the assumed criteria, for instance do not contain the edge elements, can thus be excluded from subsequent processing.

In the experiments described below we use all the acceleration methods described above. At one instant the set of sites $P \subset (S, R)$ was updated parallelly. Early in the reconstruction process the set P contained all sites i.e. the whole image was processed at one instant. After a few iterations, P was split into four subsets, and then each new set was processed independently. That rule was applied many times and at the end P contained only one site. After each split, a verification of the new set was performed. The set which did not fulfill the assumed criteria, namely it did not contain the edge elements, was excluded from subsequent processing. In the case of simple pictures, this policy results in decreasing the amount of sites taken into account during the reconstruction. The usage of the criteria described above results in the processing of a set containing the pixels placed in the neighborhood of edge elements. When the processed area was small enough, we used the extended constraints connected with the shape of the edges.

Figure 2 shows the results of the restoration of a simple image of dimension 32×32 pixels and 16 levels of grayness. We performed in each case three hundred iterations which took about 20 sec. on a PC computer (P133, 32 Mb RAM). In the first experiment, the original image was corrupted by an additive noise of ± 5 levels of grayness. In the second experiment, the image was blurred and than corrupted by an additive noise. A simple linear blurring mask was used. The general rule of Eq. (19) was specified to

$$y_q = \frac{1}{2} x_q + \frac{1}{16} \sum\nolimits_{r \in \partial(q)} x_r + \eta_q$$

7 Conclusions

The experiments indicate usefulness of the parallel-serial processing and call for further theoretical analysis. In particular, it seems prospective to use a space-scale transformation, for instance the wavelet transform. The scale domain is useful for both building the prior constraints and speeding up the algorithms. Some interesting ideas in this direction can be found in [7] and [8].

References

[1] G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods,* Springer-Verlag, Berlin, 1995

[2] A. Katsaggelos, *Digital Image Restoration*, Springer-Verlag, Berlin, 1991

[3] R. Chellapa, "Digital Image Restoration Using Spatial Interaction Models," *IEEE Trans. on Acoustics, Speech and Signal Processing,* vol. ASSP-30, no. 3, Jun. 1982, pp. 461-472

[4] S. Geman and D. Geman "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. PAMI-6, no. 6, Nov. 1984, pp. 721-741

[5] A. Jain, "Advances in Mathematical Models for Image Processing," *Proc. of the IEEE,* vol. 69, no. 5, May 1981, pp. 502-528

[6] A. Tekalp, "Identification of Image and Blur Parameters for the Restoration of Noncausal Blurs," *IEEE Trans. on Acoustics, Speech and Signal Processing,* vol. ASSP-34, no. 4, Aug. 1986, pp. 963-971

[7] M. Luettgen "Multiscale Representation of Markov Random Fields," *IEEE Trans. on Signal Processing,* vol. SP-41, no. 12, Dec. 1993, pp. 3377-3395

[8] M. Malfait "Wavelet Based Image Denoising Using a Markov Random Field a Priori Model," *IEEE Trans. on Image Processing,* vol. IP-6, no. 4, April 1997

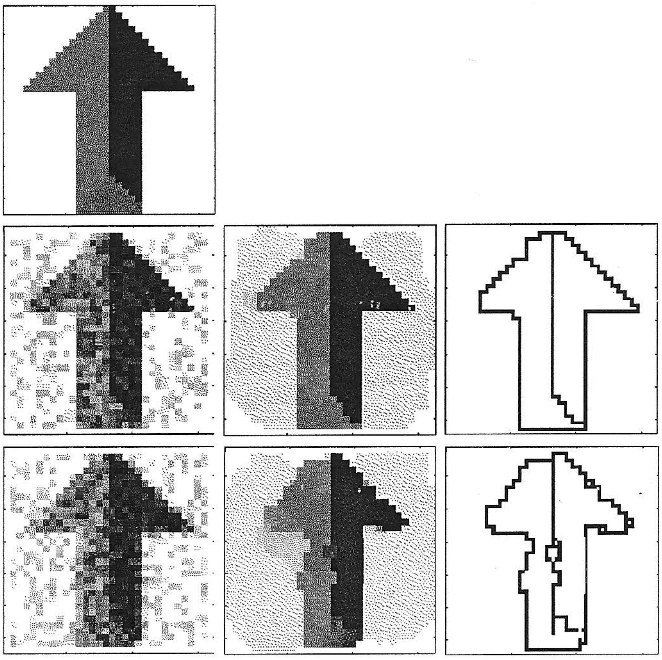

Figure 2: Results of experiments. Upper row: the original version is of the image. Middle row, from the left: the image corrupted by additive noise, the restored version and the edge elements obtained during restoration. Lower row, from the left: the image corrupted by noise and blur, the restored version and the edge elements obtained during restoration.