International Journal of Computer Vision, 16, 135–145 (1995)
© 1995 Kluwer Academic Publishers, Boston. Manufactured in The Netherlands.

Markov Fusion of a Pair of Noisy Images to Detect Intensity Valleys

R. AZENCOTT
SUDIMAGE, 47 av. Carnot, Cachan 94230 France; and DIAM/CMLA, Ecole Normale Supérieure, Cachan, France

B. CHALMOND
SUDIMAGE, 47 av. Carnot, Cachan 94230 France; and University of Cergy-Pontoise, Physics Department, France

F. COLDEFY
SUDIMAGE, 47 av. Carnot, Cachan 94230 France; and University of Paris-Sud, Statistiques appliquées, Orsay, France

Received October 22, 1993; Accepted February 11, 1994

Abstract. Our presentation is related to a non-destructive control industrial task: the detection of defects on pairs of γ-radiographic images. The images are very noisy and have a strong luminosity gradient. Defects are identified with intensity valleys. First we present a Bayes-Markov model in order to estimate the noise, the gradient and the valley bottom lines of defects for a single image. Then, we define a Markov fusion model for a pair incorporating a criterion of similarity between matched images. The proposed Markov models are general and can be used in other situations for detecting valley bottoms in noisy images.

1 Introduction

Our study is concerned with a non-destructive control industrial task (Azencott et al. 1992; Coldefy 1993). The data consist of a pair of γ-radiographic images of the same view of a pipe in a nuclear power station, on which we have to detect defects. The defects appear as locally dark and oblong areas (Figs. 3a and 3b), in very noisy images (Figs. 3d and 3e) with a luminosity gradient (Fig. 3c). They are well-known to specialized experts and are classified in an atlas of American origin. When a γ-ray strikes the film, its energy is strongly attenuated. Thus, the defects are hardly visible and are often similar to acquisition artefacts. Furthermore, a same defect appears quite differently in form and in intensity on both images (Figs. 3d and 3e) although the acquisition process is rigorously identical for both films.

The experts perform the detection and the discrimination between defects and artefacts by using two stacked radiographies they slide one over the other until they detect a local contrast enhancement. This contrast enhancement corresponds to defects area and appears when both films are exactly matched. When defects are detected by this way, the films are digitized in order to be treated by the method described in this paper (This method is actually a preprocessing that will be used later to perform a 3D reconstruction of the defects (Chalmond et al. 1993)). Let us emphasize that in our presentation, the two images are assumed to be matched (except in the last section 8).

Defect and artefact areas on a given image can be regarded as valleys in a very uneven intensity landscape. We do not attempt to precisely estimate the valleys themselves but rather to accurately detect their *valley bottom lines* (VBLs for short). The ground idea justifying this method is that the Signal to Noise Ratio (SNR) is higher at the valley bottoms sites than anywhere else inside the valley.

Detection of valleys on a single image has already been developed by several authors (see Gauch and Pizer 1993 and the references therein). David and Zucker (1990) proposed a variational model of valley covering by continuous curves from intermediate data called tangent field. Tan et al. (1991) presented a Markov model for edge detection using a dissimilarity region enhancement. Their method is based on a cost minimization and can be viewed as a detection of VBL for a suitable energy. Both methods incorporate a combination of global criteria such as curves regularity and local information in order to counter the perturbation due to noise during the reconstruction. However, for low SNR, as in our application, the tangent field or the region dissimilarity enhancement are strongly degraded,

and these previous algorithms much less reliable. The discrimination between valleys created by the noise and those due to actual defects is crucial.

First we present a Bayes-Markov model for discrete VBL detection on a single image. We define the discrete VBL as one pixel wide and connected chains. This definition can be compared to Tan et al.'s model. Nevertheless we are dealing with non-intersecting and isolated valleys. We pay a particular attention to the noise model. Discrimination between defect valleys and valleys due to the noise relies on an estimate whose sensitivity depends on statistical criteria. More precisely, we define an energy controlled by a parameter representing the minimal level of SNR acceptable for valley detection.

Secondly, we define a joint Markov model for a pair of matched images. As a same valley area differs in form and in intensity from an image to the other, our model is based on fusion criteria. The problem does not consist in matching fields from primitives as in (Yuille et al. 1990) but in performing a fusion of two different fields of primitives extracted from matched images. The model takes into account the information present on both films in order to perform a more accurate detection from which acquisition artefacts are removed.

This article is organized as follows: in the next five sections, we describe the Bayes-Markov model for a single image. After a global overview of the method in section 2, the following sections present the prior distribution and the reconstruction algorithm. The section 7 is concerned with the Markov fusion model for a couple. Finally, the detection results are presented in section 8.

2 Detection on a Single Image

Definitions. Let y denote the digitized image and S the sample grid. We assume y is an occurrence of a random field Y, which results from the degradation of a discrete surface d by a white noise W with variance σ^2 and by a gradient of luminosity μ. The surface d is also seen as a realization of a random field D. The degradation model is additive:

$$Y = \mu + D + W. \tag{1}$$

For every occurrence d and every site $s \in S$, we assume $d_s \leq 0$ with $d_s < 0$ when s belongs to a valley and $d_s = 0$ otherwise. From the VBLs of d, we define a binary field ℓ such that $\ell_s = 1$ when s is a VBL site and $\ell_s = 0$ otherwise. The VBLs are defined according to three types of features:

- (f1) curvature feature in d,
- (f2) depth feature,
- (f3) geometrical features.

The field ℓ is also regarded as the occurrence of a random field L. This field is referred as a *virtual hidden* field since it is defined from the hidden field D. At this stage, we notice an analogy between our model and the Geman and Geman's model for the restoration of an image x, that is: $Y = X + W$, in which the virtual hidden field ℓ is issued from the edge elements of x. However, one particular difficulty of our model lies in the unknown surface μ. Since the luminosity gradient is smooth, we choose μ in the vector space of bicubic spline surfaces (Prenter 1971), at a coarse grid resolution c which is deduced from the schedule of our acquisition. So, μ is represented by the model

$$\mu = B' \alpha B, \tag{2}$$

where B is the deterministic matrix of the B-spline functions at the resolution c (Chalmond 1986). This representation has been mainly chosen because of the minimal flexion property of the splines (Eubank 1990). Thanks to this property, the more the resolution c is coarse, the less the spline estimation will be sensitive to the valleys. Finally, the unknown elements in the considered model are composed of the fields D and L and the parameters $\theta = (\alpha, \sigma)$.

Bayes-Markov Approach. A now very well documented approach consists in defining an energy function $U^\theta(d, \ell \mid y)$ which expresses our knowledge given y on the set of configurations (d, ℓ), then in determining a configuration $(\hat{d}, \hat{\ell})$ which minimizes a functional of U^θ. Starting with the innovative paper (Geman and Geman 1984), the Bayes-Markov approach has solved numerous applications of this kind (Azencott 1987, 1992; Chalmond 1988, 1989; Clark 1990; Geman 1990, 1992; among many others). Indeed, U^θ is the energy associated to the posterior Gibbs distribution on (d, ℓ) given y: $P_\theta(d, \ell \mid y) \propto \exp -U^\theta(d, \ell \mid y)$, which is rewritten using the Bayes formula

$$\begin{aligned} P_\theta(d, \ell \mid y) &\propto P_\theta(y \mid d, \ell) P_\theta(d, \ell), \\ &\propto P_\theta(y \mid d) P_\theta(d, \ell), \end{aligned} \tag{3}$$

where $P_\theta(d, \ell)$ is a prior Gibbs distribution. Before presenting in detail the model expression, let us emphasize how much it is crucial to consider the couple

(d, ℓ). This is due to the interactions between d and ℓ modelized in the joint distribution $P_\theta(d, \ell)$, which constrains both estimations of d and ℓ. This is one of the main points of this approach. For instance, in the problem of image restoration cited above and based on the model $Y = X + W$, smoothing y and the detection of the edge elements in x are simultaneously performed in interaction. In our case, it is also better to define an interaction model to perform the restoration of d and the detection of the VBLs ℓ, rather than to estimate first d and then ℓ. Notice also that in the restoration problem $Y = X + W$, the introduction of L improves the smoothing, although the estimation of ℓ is a side issue. In our case the situation is reversed, we mainly want to obtain an accurate estimation of ℓ.

Let us give the expression of $P_\theta(y \mid d)$ which is related to the noise. W being a white noise, it follows that conditionally to d, the random field Y is an independent Gaussian field. $P_\theta(y_s \mid d_s)$ is defined by the Gaussian distribution $\mathcal{N}(\mu_s + d_s, \sigma^2)$, and thus

$$P_\theta(y \mid d) = (2\pi)^{-n/2} \exp -U_1^\theta(y \mid d)$$

$$U_1^\theta(y \mid d) = \sum_{s \in \mathcal{S}} \left[\frac{(y_s - \mu_s - d_s)^2}{\sigma^2} + \log \sigma \right].$$

These expressions show that $P_\theta(y \mid d, \ell)$ is reduced to $P_\theta(y \mid d)$, as it is set in (3). The energy $U_1^\theta(y \mid d)$ expresses the distance between the configuration d and the data y.

The probability $P_\theta(d, \ell)$ is a Gibbs distribution whose energy is $[U_2^\theta + U_3]$:

$$P_\theta(d, \ell) \propto \exp -\left[U_2^\theta(\ell \mid d) + U_3(d)\right]. \quad (4)$$

The energy U_2^θ expresses the features f1, f2 and f3, and the energy U_3 takes into account the low occurrence rate of valleys in the images.

To carry out this Markov approach, we first have to construct the prior energy $[U_2^\theta + U_3]$ (sections 3 and 4), and then to minimize the posterior energy (section 5):

$$U^\theta(d, \ell \mid y) = U_1^\theta(y \mid d) + U_2^\theta(\ell \mid d) + U_3(d). \quad (5)$$

Because of the unknown parameter θ, this optimization will raise some particular difficulties.

3 Construction of U_2

The field L, given d, is regarded as a Markov random field with Gibbs distribution $P_\theta(\ell \mid d) \propto \exp -U_2^\theta(\ell \mid d)$. Ideally, ℓ is organized as a finite set of one pixel wide and connected chains of sites s with label $\ell_s = 1$ and presenting the features f1, f2, f3 which express our prior knowledge. The energy $U_2^\theta(\ell \mid d)$ corresponds to the interaction between the VBL and these features. It takes high values when VBL configurations do not satisfy them, and low values otherwise. This energy is the sum of two terms

$$U_2^\theta(\ell \mid d) = U_2^{\text{ext}}(\ell \mid d) + U_2^{\text{in}}(\ell). \quad (6)$$

The energy U_2^{ext} depends on θ but U_2^{in} does not. Here, the index θ is dropped in order to simplify the expressions. U_2^{ext} incorporates the features f1 and f2: it favors VBLs whose each discrete point is near an "indicator line" of curvature in d (see below) and is situated deep enough under the surface μ. U_2^{in} incorporates the feature f3: it favors smooth, long and one pixel wide VBLs.

3.1 *Energy U_2^{ext}*

Indicator Lines of Curvature. At each site s, we want to extract from d a characteristic of curvature in order to define the feature f1. First, we recall some classical definitions of differential geometry. Let $\tilde{d}$ be the continuous surface from which the discrete field d has been sampled, and M be a point on this surface. We denote $\rho_1 \leq \rho_2$ the principal curvatures of $\tilde{d}$ at M, and τ_1, τ_2 the principal axes respectively associated to ρ_1 and ρ_2 (DoCarmo 1976). In image analysis, principal curvatures are widely used as basic characteristics (see for instance (Parent and Zucker 1988)). If the surface $\tilde{d}$ contains no flat areas, we define its indicator lines as the set of points M for which $\rho_2 > 0$ and τ_2 is an horizontal vector. Figure 1 presents examples of such indicator lines. In Fig. 1a, it is a line of parabolic points. In Fig. 1b, it is the union of two lines of hyperbolic points.

In the discrete case, a point on d with coordinate $s \in \mathcal{S}$ is assumed to be close to an indicator line of $\tilde{d}$ if it is a local minimum of d in a $(2l + 1) \times 1$ window centered at s, whose direction is the orthogonal projection on the sample grid of the principal vector τ_2 associated to the largest principal curvature ρ_2 of d at s. τ_2 and ρ_2 are computed by discretization. We set $T_s = 1$ for such a site and $T_s = 0$ otherwise. When $T_s = 1$, s is referred as a *site of feature* f1.

In practice, in order to regularize the discrete first and second derivatives of d, the principal curvatures are computed from a smoothed version of d obtained by convolution with a 5×5 Gaussian filter whose parameter is chosen equal to 1. Note that $T = T(d)$ is a

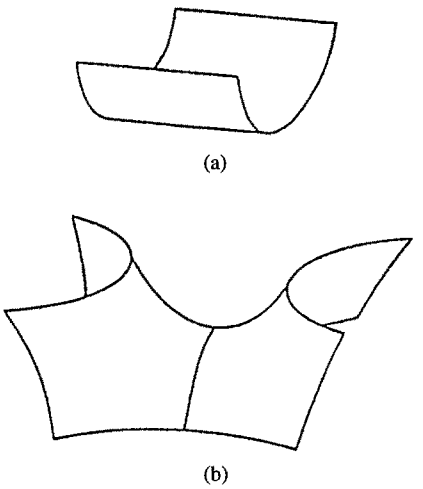


Fig. 1. Two examples of indicator lines of curvature: (a) $\rho_1 = 0$, $\rho_2 > 0$; (b) $\rho_1 < 0$, $\rho_2 > 0$.

deterministic function of d. This operation is very efficient for smooth surfaces. The whole set of points of feature f1 is a set of one or two pixels wide and irregular chains almost covering the valley bottoms. The field T can be regarded as a perturbated VBL field which has to be regularized and from which short lines has to be removed.

Construction of U_2^{ext}. The expression of U_2^{ext} is

$$U_2^{\text{ext}}(\ell \mid d) = -\sum_s 1_{\ell_s=1} \log P(\ell_s = 1 \mid d_s) - \sum_s 1_{\ell_s=0} \log P(\ell_s = 0 \mid d_s). \quad (7)$$

Before presenting in detail the expression of $P(\ell_s \mid d_s)$, let us give the basic ideas which guide the forthcoming development. The energy U_2^{ext} corresponds to a local and independent analysis of the data d at each site. It is built from $P(\ell_s \mid d_s)$ estimating the presence/absence probabilities of a valley at each site. This energy favors valley detection at sites s which have been previously detected as sites of feature f1 and where the SNR is greater than a given empirical threshold k (feature f2). The SNR threshold k guarantees that the detected site belongs to a true valley. So, weak valleys are identified with low noise frequencies and are ignored.

$P(\ell_s = 1 \mid d_s)$ is defined as the product of two probabilities, say P_1 and P_2, as follows. The event $\ell_s = 1$ is true only if s belongs to the "bottom of a deep valley," what we denote by $s \in \mathcal{V}$. Let us examine $P(\ell_s = 1 \mid d_s)$ with regard to the event $s \in \mathcal{V}$. We write

$$P(\ell_s = 1 \mid d_s) = P(\ell_s = 1, s \in \mathcal{V} \mid d_s) = P_1(\ell_s = 1 \mid s \in \mathcal{V}, d_s) P_2(s \in \mathcal{V} \mid d_s).$$

The expression of P_1 can be simplified. As a matter of fact, when one observes $s \in \mathcal{V}$, the event $\ell_s = 1$ only depends on the value of $T_s(d)$. Let Ψ denote this function. So we write

$$P_1(\ell_s = 1 \mid s \in \mathcal{V}, d_s) = P_1(\ell_s = 1 \mid T_s(d)), = \Psi(T_s(d)). \quad (8)$$

This probability is empirically chosen. Since a site s such that $T_s(d) = 1$ is regarded as a potential point of VBL, $\Psi(1)$ is chosen close to 1, and $\Psi(0)$ close to 0. Choosing $\Psi(1)$ strictly lower than 1 allows to regularize the VBLs: the value of this probability defines how distant can be the fields T and ℓ.

Let us present the probability P_2. Deciding that $s \in \mathcal{V}$, given y_s, relies on the classical statistical hypothesis test upon the mathematical expectation of a random variable (Brownlee 1965):

$$\mathcal{H}_0 : E(Y_s \mid d_s) = \mu_s, \quad \text{versus} \quad \mathcal{H}_1 : E(Y_s \mid d_s) < \mu_s.$$

Let us recall $(Y_s \mid d_s)$ is a Gaussian variable $\mathcal{N}(\mu_s + d_s, \sigma^2)$. Thus, the hypothesis $\mathcal{H}_0$ is equivalent to $s \notin \mathcal{V}$ since we have $d_s = 0$. The classical decision rule accepts $\mathcal{H}_0$ if $[(y_s - \mu_s)/\sigma < -k]$ and $\mathcal{H}_1$ otherwise. k is the ϵ-percentile of the Gaussian distribution $\mathcal{N}(0, 1)$ for a given small probability ϵ, i.e. $\Phi(-k) = \epsilon$, where Φ refers to the cumulative distribution function. By definition, the power of a statistical test is the probability of deciding $\mathcal{H}_1$ when $\mathcal{H}_1$ is true and it is given in our case by $\Phi(-k - d_s/\sigma)$. We define P_2 by

$$P_2(s \in \mathcal{V} \mid d_s) = \Phi(-k - d_s/\sigma). \quad (9)$$

Note that k represents the confidence value of the statistical test. Its calibration is easy because it defines the minimal SNR required for valley detection. For instance, in the Gaussian law table, one reads $\Phi(-1.65) = 0.05$. Finally, with (8) and (9), we get

$$P(\ell_s = 1 \mid d_s) = \Psi(T_s(d))\, \Phi(-k - d_s/\sigma). \quad (10)$$

In our model, this term of U_2^{ext} governs the interaction between ℓ and d, a point which we have emphasized in section 2. Clearly, the knowledge of d contributes to the determination of ℓ, and conversely the knowledge of ℓ influences the determination of d. For instance, when $\ell_s = 1$ then we are led to prefer a configuration d maximizing (10) (remember that $\Psi(T_s(d)) = P(\ell_s = 1 \mid T_s(d))$).

3.2 *Energy* U_2^{in}

The role of the second energy term U_2^{in} is to organize the field ℓ in one pixel wide and connected chains of sites s with label $\ell_s = 1$ in order to satisfy the feature f3. As very small valleys can be regarded as acquisition artefacts, this energy tends to suppress short chains. Its expression is

$$U_2^{\text{in}}(\ell) = \sum_s \ell_s H_{V_s}(\ell),$$

$$H_{V_s}(\ell) = \beta_1 1_{n_s=0} + \beta_2 1_{n_s=1} + 1_{n_s=2}(\beta_{3,1} 1_{a_s \geq \pi/2} - \beta_{3,2} 1_{a_s < \pi/2}) + 1_{n_s \geq 3} h(n_s),$$

with the following notations:

- V_s is the 3×3 neighborhood of the site s (s excluded),
- $n_s = \text{card}\,\{t \in V_s \mid \ell_t = 1\}$,
- a_s, only defined when $n_s = 2$, is the absolute value of the angle between the two vectors $s - s'$ and $s'' - s$ where the sites s' and s'' belongs to V_s and are such that $\ell_{s'} = 1$ and $\ell_{s''} = 1$,
- all β are unknown positive parameters.
- $h(n_s)$ is a fixed positive function strongly penalizing all local configurations where $n_s \geq 3$. In that case, ℓ can not be identified with a discrete curve anymore.

The parameter β_1 penalizes isolated detections, and β_2 controls the length of the VBLs. Finally, $\beta_{3,2}$ favors regular chains whose local tangent variation a_s is lower than $\pi/2$, whereas $\beta_{3,1}$ controls the remaining local configurations. The energy U_2^{in} is minimized by fields ℓ containing long and regular chains.

The parameters β and the values of $h(n_s)$ are "calibrated" by the "qualitative box" method (Azencott 1987, 1992) as follows: we define arbitrary minimal or maximal probabilities of observing some typical configuration ℓ. Then the Gibbs distribution $P(\ell) \propto \exp -U_2^{\text{in}}(\ell)$ provides linear inequalities for the vector β and guides its determination. For instance, suppose that ℓ is the null field except at one site s where $\ell_s = 1$. The probability of observing ℓ is proportional to $P(\ell) \propto \exp -\beta_1$. We want the null field $\tilde{\ell}$ to be more likely than ℓ and thus we write $P(\tilde{\ell})/P(\ell) > r$, for a suitably high number $r > 1$. Since $U_2^{\text{in}}(\tilde{\ell}) = 0$ we get:

$$\beta_1 > \log(r).$$

Proceeding in this way for all parameters, we finally select a vector of parameters β satisfying seven linear inequalities.

4 Energy U_3

The introduction of the energy $U_3(d)$ can be justified from several points of view. First, the role of U_3 is to penalize the fields d with too many valleys. Its expression is

$$U_3(d) = \sum_s 1_{d_s<0} \kappa^2, \tag{11}$$

where the parameter κ, like k above, is an Gaussian ϵ-percentile. In the next section, it will be more detailed.

Let us now make a short note that a non-probabilistic reader can omit. In the forthcoming development, we explain how we have determined U_3 and why this energy is mathematically grounded. At first glance, the calculation of the local probabilities $P(\ell_s \mid d_s)$ from the joint distribution $P_\theta(d, \ell)$ given in (4) is not necessarily coherent with the previous definition (10). So, we should prove that there exists an energy U_3 such that a joint probability $P_\theta(d, \ell)$ verifies:

$$P_\theta(d, \ell) \propto \exp -[U_2(\ell \mid d) + U_3(d)],$$

with

$$P(\ell_s = 1 \mid d_s) = \Psi(T_s(d))\Phi(-k - d_s/\sigma) \quad \forall s. \tag{12}$$

After a rather technical development, it is shown in (Coldefy 1993) that the energy expression

$$U_3(d) = -\sum_s 1_{d_s=0} \log(p) - \sum_s 1_{d_s<0} \log(1-p). \tag{13}$$

which is associated to the distribution

$$P(D_s = 0) = p,$$
$$P(D_s < 0) = 1 - p,$$

satisfies (12). With $\kappa^2 = \log[p/(1-p)]$, the expression (13) is equivalent to (11). Note that (11) is not the unique expression which satisfies (12), but it yields very simple analytical forms for the estimation purpose, as we shall see below.

5 Optimization

We *simultaneously* have to reconstruct the fields (d, ℓ) and to estimate $\theta = (\alpha, \sigma)$. The estimation is based

on (1) or similarly on the posterior energy (5) that we recall:

$$U^\theta(d, \ell \mid y) = U_1^\theta(y \mid d) + U_2^\theta(\ell \mid d) + U_3(d).$$

If θ was known, we should compute a local minimum of U^θ with respect to d and ℓ. In that case the classical methods, either stochastic (Geman and Geman 1984) or deterministic (Besag 1989) use the local conditional probabilities associated to $P_\theta(d, \ell \mid y)$ during an iterative process refreshing the value of every pixel. Unfortunately, this process is very expensive in computation time since we have to compute T for every considered field d.

Furthermore θ is unknown. So, a feasible approach consists in seeking a local minimum by using an iterative process which alternates between the estimations of (θ, d) and ℓ. This process belongs to the class of the numerical relaxation algorithms (Cea and Glowinski 1973), and are sometimes referred as "coordinate descent" algorithm. It is a two-stage minimizing procedure which begins to search a first estimate of (d, ℓ). Then this estimate is iteratively improved as follows: at each iteration, with ℓ equal to the estimate obtained at the previous iteration, compute $[\arg\min_{\theta,d} U^\theta(d, \ell)]$; then with (θ, d) equal to this last estimate, compute $[\arg\min_\ell U^\theta(d, \ell)]$. This type of algorithm has been widely used ((Ord 1975; Chalmond 1986), among many others), and has been precisely studied in some specified contexts (see for instance (Carrol 1988)). Its convergence is generally very fast. This algorithm being quite technical in our application, we put stress on the presentation of the first estimation.

First Estimation. This method is a two-step procedure:

1. first we compute $(\hat\theta, \hat d)$ minimizing the energy $(U_1^\theta + U_3)$ subject to the constraint that the discrete field d verifies $d_s \leq 0$ for all s;
2. secondly, the Maximum Posterior Marginal (MPM) estimate of ℓ is obtained from $U_2^{\hat\theta}(\ell \mid \hat d)$.

The step 1 is described in the Appendix. In the step 2, $\hat\ell$ is the MPM estimate calculated from $U_2(\ell \mid \hat d)$. We recall its definition (Chalmond 1982; Marroquin 1987). Let ℓ^* be the "true" field of the VBLs. The field $\hat\ell$ minimizes the mean cost $E_{\hat d}(C(\ell))$ where $C(\ell) = \sum_s 1_{\ell_s \neq \ell_s^*}$, and hence is given by:

$$\forall s \in \mathcal{S}, \qquad \hat\ell_s = \arg\max_{x=0,1} P_{\hat\theta}(L_s = x \mid \hat d).$$

The probabilities $P_{\hat\theta}(L_s = x \mid \hat d)$ are estimated by a Monte-Carlo algorithm using a Gibbs sampler (Geman and Geman 1984).

General Algorithm. From the first estimation, the general relaxation algorithm can continue, by computing alternatively the estimation of (θ, d) and the estimation of ℓ. But instead of minimizing $(U_1^\theta + U_3)$ over (θ, d) as before during the step 1, now we minimize the global energy $U^\theta(d, \ell \mid y)$. So we have to modify the expression of the system (17), since the differentiation of this energy in (θ, d) leads to a new expression of $\hat d_s$.

Comments on U_3. The expression of $\hat d$ in (17) allows us to give another justification of the choice of U_3. In this one, the estimation of d is performed with respect to the weighted noise variance $\kappa\sigma^2$ (As in the Section 3.1, this term could be derived from a statistical hypothesis test for which $\Phi(-\kappa/2) = \epsilon$). Here we want emphasize that ignoring U_3, is equivalent to choose $\kappa = 0$ and then to estimate valleys that are not significant.

6 A Simpler Method

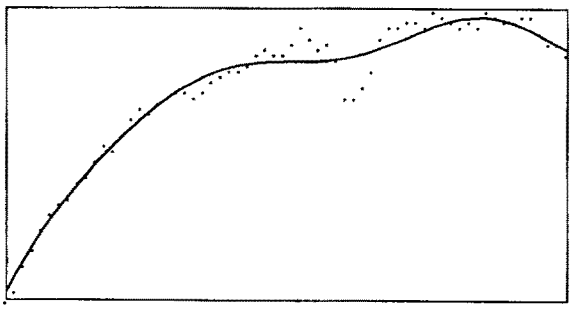
A natural idea would have been to completely separate the estimations of (θ, d) and ℓ. This is the first approach we have considered in (Chalmond and Coldefy 1991) and we now describe briefly. First we minimizes $U_1^\theta(y \mid d)$ over (θ, d) and then $U_2^{\hat\theta}(\ell \mid \hat d)$ over ℓ.

For the first optimization, we adopt a multiresolution representation. μ is defined at a coarse resolution c according to (2): $\mu = B_c' \alpha_c B_c$ and d at a fine resolution f according to: $d = B_f' \alpha_f B_f$, $(c = 5f$ in our application). The surface d is supposed to be orthogonal to the surface μ. The estimations are

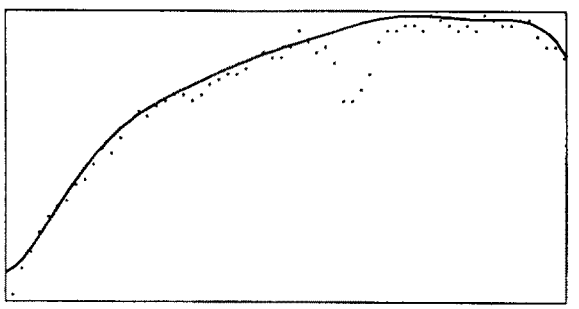
$$\hat\mu = \mathcal{P}_c(y),$$
$$\hat d = \mathcal{P}_f(y - \hat\mu),$$
$$\hat\sigma^2 = \sum_s (y_s - \hat\mu_s - \hat d_s)^2 / \#(\mathcal{S}).$$

The second optimization is based on the MPM principle as it is described in the previous section.

This simple approach, we refer as "rigid", apparently yields good results, but has some drawbacks. Due to the valleys the estimation of μ is biased, and thus the valleys are under estimated. The sophisticated method



(a)



(b)

Fig. 2. A profile of an image with its estimated trend $\hat{\mu}$: (a) rigid method; (b) adaptative method.

presented in the previous sections, we refer as "adaptative", alleviates this drawback. This is illustrated in the Fig. 2.

7 Joint Markov Model for a Pair of Matched Images

We are now dealing with a pair of matched images in order to perform a more accurate detection of the defects. For instance, acquisition artefacts which are randomly and independently distributed on each film have to be removed from the detection. On the opposite, a valley with low depth on one of the images but corresponding to a deep valley on the other one have to be identified as a defect valley.

We assume that the two digitized images y^1 and y^2 are matched: for every site $s \in \mathcal{S}$, we get the couple (y_s^1, y_s^2). The model (1) is replaced by

$$Y^1 = \mu^1 + D^1 + W^1,$$
$$Y^2 = \mu^2 + D^2 + W^2.$$

The surfaces d^1 and d^2 can be seen as two perturbated occurrences of a same field d. Our main unknown is the hidden random field ℓ representing the VBL field of d. Let $\hat{d}^1, \hat{\sigma}_1$ be the estimation of d^1, σ_1 respectively, and $\hat{d}^2, \hat{\sigma}_2$ be the estimation of d^2, σ_2 respectively, obtained from the previous method.

In fact, we forget d^1 and d^2 and we work only with their estimation $\hat{d}^1$ and $\hat{d}^2$. Then we have to define an energy taking account of their weak similarity. We suppose that L is a Markov random field conditionally to $\hat{d}^1$ and $\hat{d}^2$ associated to the fusion energy:

$$H^{\hat{\theta}}(\ell \mid \hat{d}^1, \hat{d}^2) = H_2^{\text{ext}}(\ell \mid \hat{d}^1, \hat{d}^2) + U_2^{\text{in}}(\ell), \quad (14)$$

where $\hat{\theta} = (\hat{\sigma}_1, \hat{\sigma}_2)$. The first term quantifies the quality of the VBLs which are present on both images. The second term is the same energy described above quantifying shape criteria for the VBLs. Before introducing the expression of H_2^{ext}, we need to define some new fields.

New Data Fields. For every site s, we define the mean field $\delta_s^3 = (\hat{d}_s^1 + \hat{d}_s^2)/2$. We can not assume that a valley observed on δ^3 always corresponds to a valley on $\hat{d}^1$ and $\hat{d}^2$. Indeed, it may be due to an artefact present only on one of the images. Thus we define the fields δ^1 and δ^2 by:

$$\delta_s^1 = \hat{d}_s^1 \quad \text{if} \quad T_s(\delta^3) = 0,$$
$$= \hat{d}_j^1 \quad \text{if} \quad T_s(\delta^3) = 1,$$

where j is the nearest site from s, if it exists, in a prescribed centered window such that $T_j(\delta^1) = 1$, and $j = s$ otherwise. Clearly, δ_s^2 is defined in a similar way. These fields allow us to detect the presence/absence of artefacts on each image. Furthermore, we take into account the fact that because of the noise a same valley may differ in form and in intensity from an image to the other. The size of the window corresponds to the maximum of the disparity that we accept between the films. So, for each site s such that $T_s(\delta^3) = 1$, we will associate the values of $\hat{d}^1$ and $\hat{d}^2$ corresponding to the nearest VBL candidates (if it exists) previously detected by $T(\delta^1)$ and $T(\delta^2)$. We then preserve the depth of both valleys.

Energy H_2^{ext}. Its expression is:

$$H_2^{\text{ext}}(\ell \mid \hat{d}^1, \hat{d}^2) = -\sum_s 1_{\ell_s=1} \log P(\ell_s = 1 \mid \delta_s)$$
$$-\sum_s 1_{\ell_s=0} \log P(\ell_s = 0 \mid \delta_s), \quad (15)$$

where $\delta_s = (\delta_s^1, \delta_s^2, \delta_s^3)$ and $P(\ell_s \mid \delta_s)$ are the probabilities defined by:

$$P(\ell_s = 1 \mid \delta_s) = \Psi\big(T_s(\delta^3)\big)\, \Phi\big(-k_1 - \delta_s^1/\hat{\sigma}_1\big)\, \Phi\big(-k_1 - \delta_s^2/\hat{\sigma}_2\big)\, \Phi\big[-k_2 - \big(\delta_s^1/\hat{\sigma}_1 + \delta_s^2/\hat{\sigma}_2\big)/\sqrt{2}\big], \tag{16}$$

in which k_1 and k_2 are two ϵ-percentiles (cf. section 3).

This energy favors detection at sites s which are previously detected by $T(\delta^3)$ and whose SNR is great enough on both images. Theoretically, we should suppress the last term in (16) but in fact a valley may have very different SNR on both images. So we prefer to detect a valley at a very low SNR on one of the images if it corresponds to a deep valley on the other one. So we choose $k_1 < k_2$, k_1 being high enough to avoid detection of artefacts.

Finally, the true field of the common VBLs is estimated by the MPM estimate for the energy H.

8 Experimental Results

Although both radiographies are obtained according to the same process (it is the same view of the pipe) digitized images within a pair differ by a translation shift and a small rotation. This is due to the high resolution of the discretization. So seeking the VBLs leads also to match the two images. Unfortunately, the presence of noise prevents matching by means of classical techniques (maximization of the correlation for instance). Then for this application, we developed a specific three steps method:

- First, we detect defects on each image separately according to the model U given in (5). We obtain two binary fields of VBLs, let say $\hat{\ell}^1$ and $\hat{\ell}^2$, which are transformed into grey level images g^1 and g^2 by convolution with a 5×5 filter.
- Secondly the shift and the rotation between both images are estimated by a first maximization of a functional of g^1 and g^2 quantifying the similarity at low resolution level and by a second maximization of a similar functional at high resolution level. Then the user has to validate this operation.
- Finally according to the Markovian model H given in (15), we detect common VBLs on both images shifted and rotated.

Let us now present our results in non-destructive control. In order to reduce the computation time, original images are reduced from 512×512 to 128×128 after

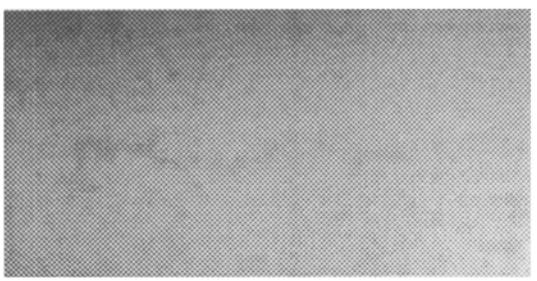


Fig. 3a. γ-image y^1.

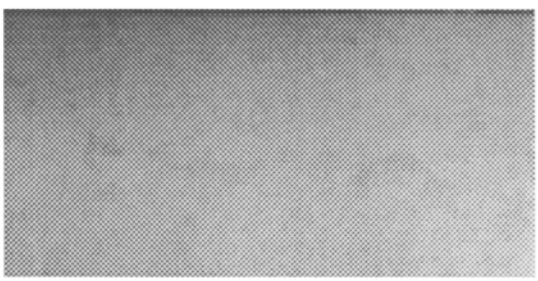


Fig. 3b. γ-image y^2.

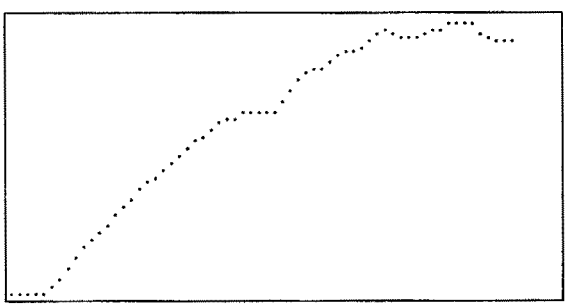


Fig. 3c. A column profile of y^1.

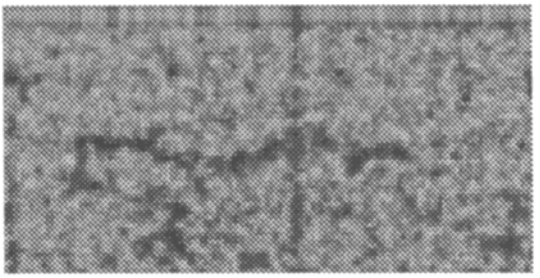


Fig. 3d. First image after removing the luminosity trend: $y^1 - \hat{\mu}^1$.

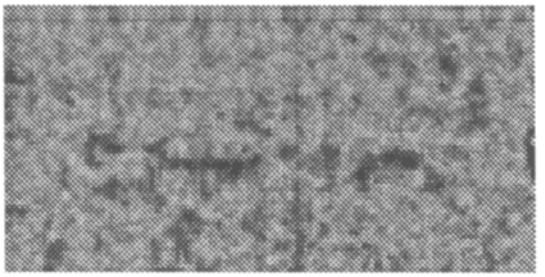


Fig. 3e. Second image after removing the luminosity trend: $y^2 - \hat{\mu}^2$.

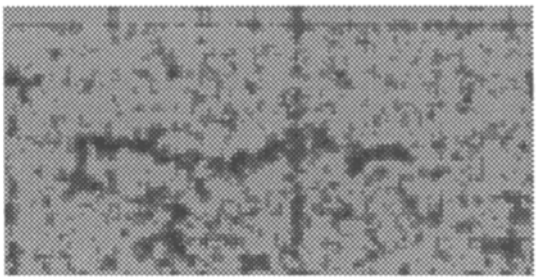


Fig. 3f. Estimated valley surface: $\hat{d}^1$.

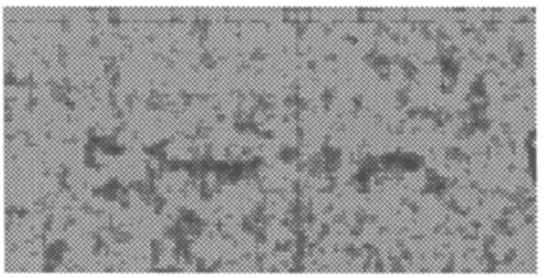


Fig. 3g. Estimated valley surface: $\hat{d}^2$.

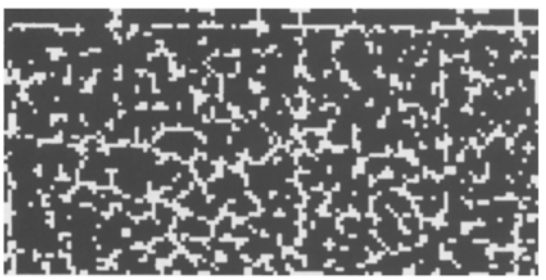


Fig. 3h. Indicator lines of curvature: $T(\hat{d}^1)$.

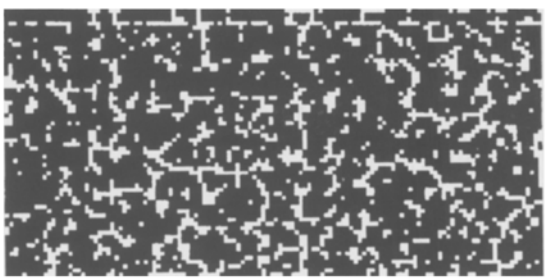


Fig. 3i. Indicator lines of curvature: $T(\hat{d}^2)$.

filtering. Note on Figs. 3a and 3b, that the SNR of the defects is very low. This is due to the luminosity gradient (Fig. 3c) and to the noise. Figures 3d and 3e show both images after removal the luminosity gradients. Note that the defect areas are perceptibly different within the pair. We present the estimation $\hat{d}$ of d for both images on Figs. 3f and 3g (the vertical line is not a defect but an artefact present on both images). The noise explains the erratic behavior of the detected bottom lines $T(\hat{d})$ (Figs. 3h and 3i).

Figures 3j and 3k show the first Bayesian-Markov detection of the VBL fields ℓ on each image separately. Final detection based on the pair of images is presented on Fig. 3l. Only valleys present on both images are now detected. Whereas valleys detection on $\hat{d}^1$ and $\hat{d}^2$ which do not match are removed, defect valleys with very low SNR on one of the images are still detected. Total computation takes 3 minutes on a Sun Spark2 station.

The practical interest of these results is that it gives to the experts a quantified decision which does not change arbitrarily when conditions change. Furthermore, as our method mimics the vision task done by the experts who slide the two films one over the other, it provides a computed artificial task well accepted by them.

Fig. 3j. Detection of VBLs from the first image: $\hat{\ell}^1$.

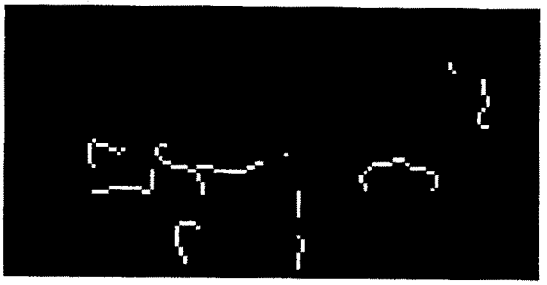


Fig. 3k. Detection of VBLs from the second image: $\hat{\ell}^2$.

Fig. 3l. Final detection of VBLs from the couple: $\hat{\ell}$.

Appendix

Here we describe the step 1 of the first estimation. Our method uses two relaxation algorithms, the second one being called by the first one (Coldefy 1993). Derivating $(U_1^\theta + U_3)$ with respect to d and θ and setting these derivatives to zero, the estimates $\hat{\mu}$, $\hat{d}$ and $\hat{\sigma}$ are the solution of the system

$$\hat{\mu} = \mathcal{P}(y - \hat{d}),$$

$$\hat{\sigma} = \sqrt{\frac{1}{\sum_s I[\hat{d}_s = 0]} \sum_s (y_s - \hat{\mu}_s - \hat{d}_s)^2},$$

$$\forall s, \hat{d}_s = y_s - \hat{\mu}_s \quad \text{if} \quad y_s < \hat{\mu}_s - 2\kappa\hat{\sigma},$$
$$= 0 \quad \text{otherwise}, \tag{17}$$

where $\mathcal{P}$ denotes the projection on the spline space (2), and I is the indicator function $I(A) \equiv 1_A$. An approximated solution is obtained by using a relaxation algorithm for minimizing $(U_1^\theta + U_3)$ alternatively over μ and (d, σ), as follows.

If $\hat{d}^{n-1}$ denotes the estimation of d at the iteration $n-1$ of this algorithm, then those of μ at the iteration n is

$$\hat{\mu}^n = \mathcal{P}(y - \hat{d}^{n-1}).$$

The next estimation of d and σ is itself obtained by applying a second relaxation algorithm to the function $[U_1^{(\hat{\mu}^n, \sigma)}(y \mid d) + U_3(d)]$. If $\hat{d}^{n,p-1}$ and $\hat{\sigma}^{n,p-1}$ denote the estimations of d and σ at the iteration $p-1$ of this second algorithm, then the estimations of d and σ at the iteration p are

$$\hat{\sigma}^{n,p} = \sqrt{\frac{1}{\sum_s I[\hat{d}_s^{n,p-1} = 0]} \sum_s (y_s - \hat{\mu}_s^n - \hat{d}_s^{n,p-1})^2},$$

$$\forall s, \hat{d}_s^{n,p} = y_s - \hat{\mu}_s^n \quad \text{if} \quad y_s < \hat{\mu}_s - 2\kappa\hat{\sigma}^{n,p},$$
$$0 \quad \text{otherwise}.$$

This second procedure stops when $\hat{\sigma}^{n,p}$ does not change anymore, and then we set $\hat{\sigma}^n \doteq \hat{\sigma}^{n,p}$ and $\hat{d}_s^n \doteq \hat{d}_s^{n,p}$. The convergence is very fast, only three iterations are approximately necessary. The whole setup of the step 1 is initialized with $\hat{\mu}^0 = \mathcal{P}(y)$ and $d^0 = 0$, $\hat{\sigma}^0 = \sqrt{\frac{1}{\#S} \sum_s (y_s - \hat{\mu}_s^0)^2}$, and converges within fifteen iterations.

Acknowledgments

The authors acknowledge Electricité De France (DER-SDM) for the financial support provided to this study and for the expert know-how afforded by B.

Lavayssière. This work was also supported by SUDIMAGE Research group which provided a stimulating environment in image analysis. We are grateful to the reviewers for helpful comments on an earlier version of this material.

References

Azencott, R. 1987. Image analysis and Markov field. *Proc. Intern. Conf. on Ind. Appl. Math. SIAM* (Invited Paper).

Azencott, R., Chalmond, B., and Coldefy, F. 1992. Association of adaptative smoothing and Markovian models for detection of valley bottoms on strongly noisy images. *Proc. 11th Intern. Conf. Patt. Recog.*, Den Haag.

Azencott, R., Chalmond, B., and Julien, Ph. 1992. Bayesian 3-D path search and its application to focusing seismic data. In *Lectures Notes in Statistics*, Stochastics Models, Statistical Methods and Algorithms in Image Analysis, Springer-Verlag, 74:46–74.

Azencott, R., Graffigne, C., and Labourdette, C. 1992. Edge detection and segmentation of textured plane images. *Ibid*: Springer-Verlag, 75–88.

Besag, J. 1988. On the statistical analysis of dirty pictures. *J. Royal Statis. Soc.*, B-148.

Brownlee, K.A. 1965. *Statistical theory and methodology in science and engineering*, Wiley.

DoCarmo, M.P. 1976. *Differential Geometry of Curves and Surfaces*. Prentice Hall: Englewood Cliffs, NJ.

Caroll, R.J., Wu, C.F.J., and Ruppert, D. 1988. The effect of estimating weights in weighted least-squares. *J. of the Amer. Statist. Ass.*, 83:1045–1054.

Chalmond, B. 1982. A Bayesian detection-estimation for a shift of the mean in an autoregressive process. *Proc. Comput. Statist. II*, Wien, Physica Verlag.

Chalmond, B. 1986. Regression with spatial autocorrelated residuals and spatial trends search. *Statist. Anal. Donn.*, 11:1–25 (in French).

Chalmond, B. 1988. Image restoration using an estimated Markov model. *Signal Processing*, 15:115–129.

Chalmond, B. 1989. An iterative Gibbsian technique for reconstruction of m-ary images. *Patt. Recog.*, 22:747–761.

Chalmond, B. and Coldefy, F. 1991. Markov model for detecting defects on radiographies of pipe in a nuclear power station. *EDF-DER Internal report HP-21/91-43* (in French).

Chalmond B., Coldefy, F., and Lavayssière, B. 1993. 3D curve reconstruction from degraded projections. In P.J. Laurent, A. Le Méhauté, and L.L. Schumaker, (eds.), *Wavelets, Images, Curve Fitting*, AKPeters Boston (to appear).

Cea, J. and Glowinski, R. 1973. On optimization methods by relaxation. *Rev. Autom. Inform. Recher. Operat.*, R-3:5–32 (in French).

Clark, R. and Yuille, A.L. 1990. *Data fusion for sensory information systems*, Kluwer.

Coldefy, F. 1993. Markov random fields and variational methods applied to low level vision tasks. Ph.D., Paris-Sud University, Orsay (in French).

David, C. and Zucker, S.W. 1990. Potentials, valleys, and dynamic global coverings. *Intern. J. Comput. Vis.*, 5:219–238.

Eubank, P. 1990. *Spline smoothing and non-parametric regression*, Decker.

Gauch, J.M. and Pizer, S.M. 1993. Multiresolution analysis of ridges and valleys in grey-scale images. *IEEE Trans. Patt. Anal. Mach. Intell.*, pp. 635–646.

Geman, S. and Geman, D. 1984. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans. Patt. Anal. Mach. Intell.*, 6:721–741.

Geman, S., Geman, D., Graffigne, C., and Dong, P. 1990. Boundary detection by constrained optimization. *IEEE Trans. Patt. Anal. Mach. Intell.*, 12:609–628.

Geman, D. 1992. Random fields and inverse problems in imaging. *Lecture Notes in Mathematics, Ecole d'Eté de Saint Flour*, 1427:111–193, Springer Verlag.

Marroquin, J.L., Mitter, S., and Poggio, T. 1987. Probabilistic solution of ill-posed problem in computer vision. *J. of the Amer. Statist. Ass.*, 82:76–89.

Ord, K. 1975. Estimation methods for models of spatial interaction. *J. of the Amer. Statist. Ass.*, 70:120–126.

Parent, P. and Zucker, S.W. 1989. Trace inference, curvature consistency and curve detection. *IEEE Trans. Patt. Anal. Mach. Intell.*, 11:823–839.

Prenter, P. 1971. *Spline and variational methods*, Wiley.

Tan, H.L., Gelfand, S.B., and Delp, E.J. 1991. A cost minimization approach to edge detection using simulated annealing. *IEEE Trans. Trans. Patt. Anal. Mach. Intell.*, 14:3–18.

Yuille, A.L., Geiger, D., and Bülthoff, H. 1990. Stereo integration, mean field theory and psychophysics. *Comput. Vis.—ECCV90, Lecture Notes in Computer Sciences*, Vol. 427, Springer-Verlag.