Probabilistic and Dictionary-Based Relaxation Techniques Applied to a Statistical Method of Edge Detection

André Hiroshi Hayashi Alves[1]
Nelson Delfino d'Ávila Mascarenhas[2]

[1]INPE - Instituto Nacional de Pesquisas Espaciais
Caixa Postal 515
12201 São José dos Campos, SP, Brasil
hiroshi@dpi.inpe.br

[2]INPE - Instituto Nacional de Pesquisas Espaciais
Caixa Postal 515
12201 São José dos Campos, SP, Brasil
nelson@dpi.inpe.br

Abstract. Two relaxation schemes, a probabilistic and a dictionary-based one, applied to edge detection, are described. The problem of edge detection is defined using a statistical approach. The solution, in terms of statistical decision theory, leads to a test among hypotheses of configurations of sets of four pixels (quadriplets). The relaxation schemes are also developed using the quadriplets as labelling objects. The initial probabilities for the label set of each object are sinthesized from the values obtained in the statistical tests. The interaction neighbourhood adopted for the two methods is the 4-neighbourhood. The iterative label probability updating is performed using a classical heuristic procedure in the two schemes. Tests using noisy synthetic and real images are presented. An experimental analysis of convergence to a consistent and non-ambiguous labelling and speed of convergence is performed for the two schemes and the results are compared. A change in the dictionary according to a modification in the definiton of consistency is proposed and the resulting scheme is tested and compared with the two other ones.

1. Introduction

The classification of the pixels in a scene as objects of the type edge or non-edge is useful in a variety of applications. It is a fundamental step for the image understanding task, where it is necessary to identify the objects that are contained in the image.

Several edge detection techniques have been proposed in the literature. For a survey of the main methods, the reader may consult [Lim, 1990] or [Gonzalez and Woods, 1992].

Two major problems faced by edge detectors are: a) the inevitable presence of noise in the image and b) the need to incorporate context in order to obtain meaningful results.

In this paper an attempt to overcome these two problems is made. First, by formulating the local edge detection problem as a statistical hypothesis testing problem and modifying a method that has been proposed by [Mascarenhas and Prado, 1980]; second, by using two relaxation techniques, one with a probabilistic formulation [Kittler, 1987] and the other based on the use of a dictionary [Hancock and Kittler, 1990], to incorporate spatial context in the local technique.

2. Review of the Local Edge Detection Method

The method that was used as a basis for the local edge detector was derived through a bayesian formulation of a statistical hypothesis testing problem [Mascarenhas and Prado, 1980], by taking explicitly into consideration the randomness of signal and noise.

A two-dimensional autoregressive model was adopted for the signal, according to the following expression:

$$s(k+1,l+1) = \rho_1.s(k+1,l) + \rho_2.s(k,l+1) - \\ - \rho_1.\rho_2.s(k,l) + \sqrt{(1-\rho_1^2)(1-\rho_2^2)}\ .U(k,l) \qquad (1)$$

In this model, the random variables are supposed to have null means; ρ_1 (ρ_2) is the correlation coefficient between non-noisy pixels s(k,l) on the horizontal (vertical) direction; and {U(k,l)} is a set of noncorrelated gaussian random variables, with the same variance as $\{s(k+1,l+1)\}$. The same model was used for the additive, independent noise.

This model leads to a separable structure of the covariance matrices for the processes, in terms of Kronecker products of the covariance matrices on the horizontal and vertical directions.

The edge detection problem is proposed in the following terms: having observed a set of four noisy pixels $v(k,l)$, $v(k,l+1)$, $v(k+1,l)$, $v(k+1,l+1)$, where $v(k,l) = s(k,l) + n(k,l)$ (i. e., noisy signal = signal + noise), we want to take a decision about the signal without noise.

The edge characterization is based on a set of hypotheses for decision about a non-noisy quadriplet (set of four pixels as in the previous paragraph), each one associated with a particular edge type, besides the non-edge decision.

The original work of [Mascarenhas and Prado, 1980] adopted a set of seven hypotheses, associated to edges that characterized the direction but not the sense of the contour. In order to obtain a refinement on the labelling, aiming at well defined contours, it is necessary to modify the original formulation, taking into account the sense of the edges. This leads to a set of thirteen hypotheses. The hypotheses are composite (each one of them involves a region in the space of the signal), and overlapping, since a given set of non-noisy pixels can satisfy more than one hypothesis constraint.

More specifically, Figure 1 below describes the set of four pixels (quadriplets) used as a basis for the model definition. Figure 2 displays the edge of the first type and its definition. Figure 3 shows the remaining eleven types of edges, as well as the non-edge configuration. Their definitions are analogous.

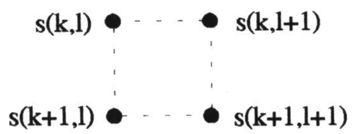
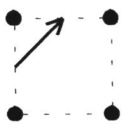


Figure 1: Quadriplet used as basis for the model definition.



$s(k,l)-[s(k,l+1)+s(k+1,l)+s(k+1,l+1)]\,/3 \geq \Delta$

Figure 2: Edge of the first type and its definition.

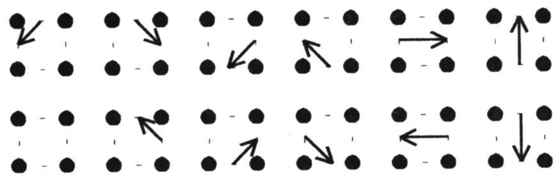
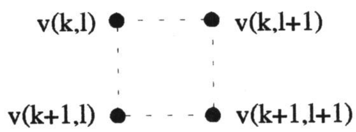


Figure 3: The other eleven types of edge and the non-edge configuration.

The non-negative parameter Δ allows to take into consideration psychophysical considerations, adjusting the result to a visual judgement.

The formulated local edge detection problem consists in taking a decision about the non-noisy signal $s(k,l)$, according to one out of the thirteen hypotheses, from the observation of a set of four pixels $v(k,l)$, as shown in figure 4.



$v(k,l) = s(k,l) + n(k,l)$

Figure 4: Observed quadriplet for local edge detection.

According to the bayesian approach, the optimal local decision rule is the one that minimizes the average risk.

The average risk for a decision γ_i is given by:

$$R(\gamma_i) = \int_V dv \int_S ds\, C(s,\gamma_i) f(v|s) \sigma(s) \qquad (2)$$

In this equation, $\sigma(s)$ defines the probability density function of the nonnoisy signal; $f(v|s)$ gives the conditional probability density function of the noisy signal given the nonnoisy signal. $C(s,\gamma_i)$ is the cost function associated with a signal value s and a decision γ_i. Observe that this optimal solution depends on the choice of the cost function $C(s,\gamma_i)$, as it is usual in Bayesian formulations.

This average risk is minimized by selecting the decision γ_i that corresponds to the minimal inner integral, given by:

$$A_i(v) = \int_s C(s,\gamma_i).\sigma(s).f(v|s)\, ds \qquad (3)$$

In order to circumvent computational difficulties, the original work of [Mascarenhas and Prado, 1980] involved a redefinition of the problem, with preliminary decisions between two non-overlapping hypotheses, by means of likelihood ratios of the type:

$$\frac{\int_{|\,|\geq\Delta} \sigma(s)f(v|s)ds}{\int_{|\,|<\Delta} \sigma(s)f(v|s)ds} \underset{0}{\overset{1}{\gtrless}} \frac{C_{01}-C_{00}}{C_{10}-C_{11}} \qquad (4)$$

The same technique was also applied in this work, but, instead of only two, we now have three hypotheses, because the sense of the edge is taken into account. Therefore, the modified testes involve three quantities:

$$\begin{aligned} A_{-1}(v) &= C_{-1-1}.I_{-1}(v)+C_{0-1}.I_0(v)+C_{1-1}I_1(v) \\ A_0(v) &= C_{-10}.I_{-1}(v)+C_{00}.I_0(v)+C_{10}I_1(v) \\ A_1(v) &= C_{-11}.I_{-1}(v)+C_{01}.I_0(v)+C_{11}I_1(v) \end{aligned} \quad (5)$$

where:

$$\begin{aligned} I_{-1}(v) &= \int_{(\,)\le-\Delta} \sigma(s).f(v|s).ds \\ I_0(v) &= \int_{|\,|<\Delta} \sigma(s).f(v|s).ds \\ I_1(v) &= \int_{(\,)\ge\Delta} \sigma(s).f(v|s).ds \end{aligned} \quad (6)$$

- C_{ij} represents the cost of taking a decision j when the correct hypothesis is i.
- A_{-1},A_0 and A_1 are the risks associated with each of the hypotheses in the individual tests.
- $\int_{(\,)\le-\Delta}$,$\int_{|\,|<\Delta}$ and $\int_{(\,)\ge\Delta}$ denote the integrals over the domain of definition of hypotheses -1, 0 and 1, respectively, in the individual tests.

Two further simplifications are necessary in order to obtain a computationally feasible problem: 1) a reduction in the number of observed variables v(k,l) from four to two; 2) an approximation on the integrals computation by performing two dimensional integrals, instead of four dimensional integrals. Under the gaussian assumption, these integrals are performed by computing error functions and Gauss-Hermite quadrature.

In order to apply the relaxation schemes, it is necessary to derive initial probabilities that are associated with each hypothesis. These probabilities are obtained from the risks given by equations (5) and (6) through the following heuristic procedure. First, the risk for each hypothesis is inverted:

$$B_i = \frac{1}{A_i},\ i = -1,0,1 \quad (7)$$

Then, in order to normalize the results of the six tests (each one involving three hypotheses), the following transformations were performed:

$$\begin{aligned} BB_{-1} &= \frac{B_{-1}}{B_0} \\ BB_0 &= \frac{B_0}{B_0} = 1 \\ BB_1 &= \frac{B_1}{B_0} \end{aligned} \quad (8)$$

Finally, the initial probabilities for the iterative relaxation procedures were computed by dividing the previous values by the sum of the thirteen values, corresponding to the thirteen hypotheses.

3. Relaxation Labelling Techniques

Once the local edge detection is completed, one can start to use the spatial contextual information provided by the adjacent quadriplets. Two methods to incorporate this information have been used: probabilistic and dictionary based relaxation labelling techniques.

3.1. Probabilistic Relaxation Labelling Techniques

The sets of objects that we will be dealing with will be the quadriplets previously defined. The labels associated with each quadriplet will be those defined in the previous section.

The compatibility between objects will be dictated by two rules. The first one determines that the contour derived from the local edges should be continuous (it deals only with direction of edges). The second one imposes that the sense of the edges should not be incompatible.

In our case these rules will be applied to objects in a 4-neighbourhood (center, north, south, east and west). Furthermore, the probabilistic formulation avoids combinatorial problems by dealing with only pairs of objects.

The iterative procedure that has been proposed by [Kittler, 1987] is formulated by the following equations:

$$P^{n+1}(\theta_0=w_{0i}) = \frac{P^n(\theta_0=w_{0i}).q^n(\theta_0=w_{0i})}{\sum_{r=1}^{13} P^n(\theta_0=w_{0r}).q^n(\theta_0=w_{0r})} \quad (9)$$

$$q^n(\theta_0=w_{0i}) = \prod_{l=1}^{4} \sum_{\theta_l\in\Omega} \frac{P(\theta_0=w_{0i}|\theta_l)}{P(\theta_0=w_{0i})}.P^n(\theta_l) \quad (10)$$

The iterative process starts with the probabilities given by the labelling performed by the local edge detection process. The a priori probabilities $P(\theta_0 = w_{0i})$, i = 1,...,13 for the center quadriplet are assumed to be equal in such a way that no configuration is favoured. The support functions $q^n(\theta_0 = w_{0i})$, i = 1,...,13 convey the contextual information for each labelling of the center quadriplet and it is expressed in terms of the conditional probabilities $P(\theta_0 = w_{0i}|\theta_l)$ involving only pairs of quadriplets. An example of a matrix C of conditional probabilities $P(\theta_0 = w_{0i}|\theta_j) = c_{ij}$ is given below for the left neighbour of the central quadriplet.

$$C_{left}=\begin{bmatrix}
\frac{1}{7} & 0 & 0 & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & 0 & 0 & 0 & 0 & \frac{1}{7} & \frac{1}{7} \\
\frac{1}{7} & 0 & 0 & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & 0 & 0 & 0 & 0 & \frac{1}{7} & \frac{1}{7} \\
\frac{1}{7} & 0 & 0 & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & 0 & 0 & 0 & 0 & \frac{1}{7} & \frac{1}{7} \\
0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 \\
0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 \\
0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\
\frac{1}{7} & 0 & 0 & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & 0 & 0 & 0 & 0 & \frac{1}{7} & \frac{1}{7} \\
\frac{1}{7} & 0 & 0 & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & 0 & 0 & 0 & 0 & \frac{1}{7} & \frac{1}{7} \\
0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 \\
\frac{1}{7} & 0 & 0 & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & 0 & 0 & 0 & 0 & \frac{1}{7} & \frac{1}{7} \\
\frac{1}{7} & 0 & 0 & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & 0 & 0 & 0 & 0 & \frac{1}{7} & \frac{1}{7}
\end{bmatrix} \quad (11)$$

3.2. Dictionary-Based Relaxation Techniques

The highly structured characteristics of the contextual edge detection problem impose severe restrictions on the type of compatibility between neighbour labellings allowed. This is a strong point in favour of dictionary-based relaxation techniques [Hancock and Kittler, 1990]. For example, by adopting a 4-neighbourhood, among the 13^5 possibilities, only a relatively small number (4743) of configurations are allowed, from the point of view of consistent contour labelling.

As compared with probabilistic relaxation, this method has the advantage of simultaneously considering the 5 quadriplets of a 4-neighbourhood and the central quadriplet for specifying the allowed configurations, instead of only pairs of quadriplets with probabilistic relaxation. However, an increase on the computational effort may be required at each iteration, due to the larger number of configurations (4743 versus 676 in our case).

The updating scheme, that starts with the local edge detection labelling probabilities, is given by equation (9), as in the previous case. The support function is expressed as

$$q^n(\theta_0=w_{0i})=\sum_{k=1}^{Z_0(\theta_0=w_{0i})}\left\{\prod_{l\in I_0}\frac{P^n(\theta_l)}{P(\theta_l)}\right\}.P(\Lambda_0^k(\theta_0=w_{0i})) \quad (12)$$

In this equation, $P(\Lambda^k_0(\theta_0=w_{0i}))$ represents the probability associated with the k^{th} entry of the dictionary for central labelling $\theta_0=w_{0i}$. It is reasonable to assume that this probability should be given by the inverse of $Z_0(\theta_0=w_{0i})$ (the number of entries for $\theta_0=w_{0i}$). It was also assumed that the a priori probabilities $P(\theta_l)$ for each quadriplet labelling on the 4-neighbourhood denoted by I_0 should be equal . $P^n(\theta_l)$, l in I_0, denotes the previous labelling probabilities of the configurations on the 4-neghbourhood that occur in the current entry of the dictionary.

4. Experimental Results

The computational implementation was performed using C language on a network of Sun workstations, running under the SunOS version 4.1.3 operating system and the OpenWindows version 3.0 graphical interface. In order to visualize the images the SAO application program from the Smithsonian Astrophysical Observatory was used. Image formatting was done through the Khoros program, from the University of New Mexico.

The matrix Co of the costs defined in (5) was selected in the form:

$$Co=\begin{bmatrix} C_{-1\,-1} & C_{0\,-1} & C_{1\,-1} \\ C_{-1\,0} & C_{0\,0} & C_{1\,0} \\ C_{-1\,1} & C_{0\,1} & C_{1\,1} \end{bmatrix}=\begin{bmatrix} 0 & 1 & 1 \\ K & 0 & K \\ 1 & 1 & 0 \end{bmatrix} \quad (13)$$

with K>1. The reason for this is the fact that the non-edge decision has to be discouraged, since, with the reformulation of the statistical detection problem in terms of non-overlapping hypotheses, the non-edge hypothesis was favoured [Alves, 1993].

The computation of the integrals $I_{-1}(v)$, $I_0(v)$ and $I_1(v)$ was performed through a previous construction of a table look up to reduce the effort during execution of the edge detection program.

The mapping from the quadriplets classification to the pixels is done by superposition, as explained by the following. Each pixel is labelled as edge or non-edge. In a quadriplet, the pixels that are labelled as edges are the ones to the left of the edge of the quadriplet, according to its sense. Since each pixel can belong to up to four quadriplets simultaneously, the pixel will have four classifications. It was assumed that the classification will be done by favouring the edge label, that is, if one or more of the classifications indicate edge, this hypothesis will be adopted.

Several tests using synthetic and real images were performed [Alves,1993]. In this work we will present two examples, involving a synthetic and a real image.

Figures 5 and 6 show the original and noisy synthetic images respectively (the variance of the additive gaussian noise is four). Figure 7 displays the result of the local edge detector. One can observe gaps on the contour of the circle as well as several isolated points and segments. In figures 8 and 9 we find the results of the application of the probabilistic and the dictionary-based relaxation, respectively, after 13 iterations. A slightly better result was obtained with the last method, at the price of an increased computational cost. Table 1 below summarizes the data for these tests.

Image size	100x100
Signal to noise ratio	5.12
Hardware	SparcStation IPX
Network status	Single user
Vert. and hor. correl. coef. (signal)	0.96
Vert. and hor. correl. coef. (noise)	0
Δ	7
K	2.0×10^6
CPU time - table construction	383.7 s
CPU time - classification	3.1 s
Probabilistic relaxation 1st iteration	28.6 s
Dictionary-based relax. 1st iteration	590.7 s

Table 1: Data for synthetic image experiment.

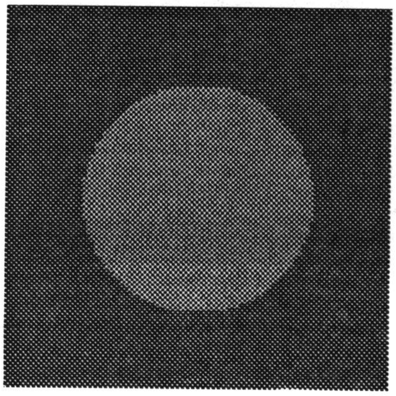


Figure 5: Original synthetic image.

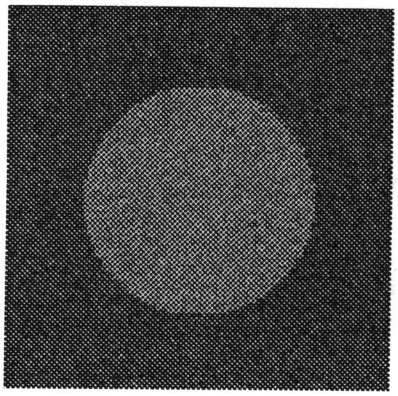


Figure 6: Noisy sinthetic image.

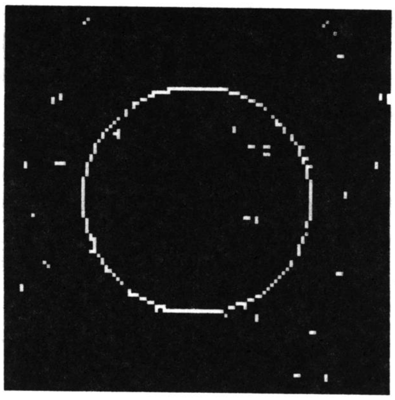


Figure 7: Local edge detection - synthetic image.

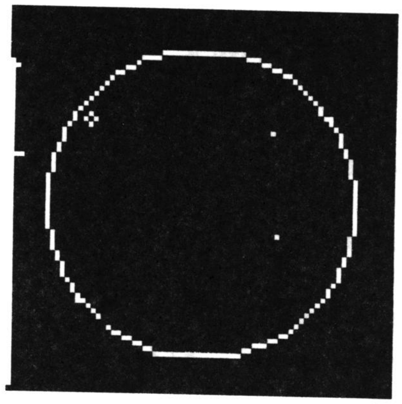


Figure 8: Probabilistic relaxation over synthetic image after 13 iterations.

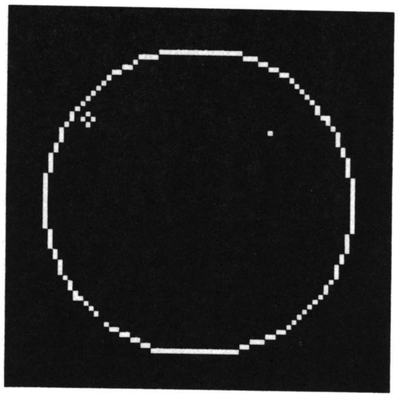


Figure 9: Dictionary-based relaxation over synthetic image after 13 iterations.

Figure 10 displays the original real image. It was acquired by an airborne synthetic aperture radar (SAR) over Tapajós area, Brazil, on April 1992 (SAREX Mission). The theoretical and practical number of looks are seven and four, respectively, and, under these conditions, the resulting speckle noise is approximately gaussian, although signal-dependent. Figures 11, 12 and 13 show the results of the local edge detection, the probabilistic and the dictionary-based relaxation, respectively, after 15 iterations. Table 2 below summarizes the data for this experiment.

Image size	250x250
Signal to noise ratio	5.95
Hardware	Sparc 10
Network status	Single user
Vert. and hor. correl. coef. (signal)	0.96
Vert. and hor. correl. coef. (noise)	0
Δ	7
K - diagonal tests - vert. and hor. tests	$1.0x10^6$ $5.0x10^5$
CPU time - table construction	121.6 s
CPU time - classification	6.7 s
Probabilistic relaxation 1st iteration	67.3 s
Dictionary-based relax. 1st iteration	1536.4 s

Table 2: Data for real image experiment

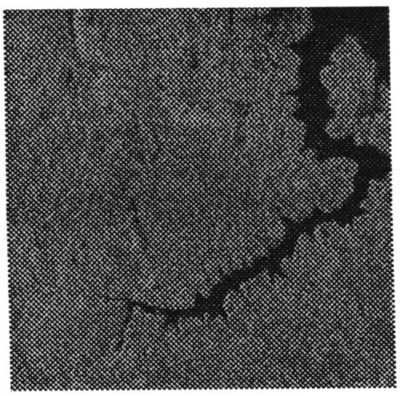


Figure 10: Original real image.

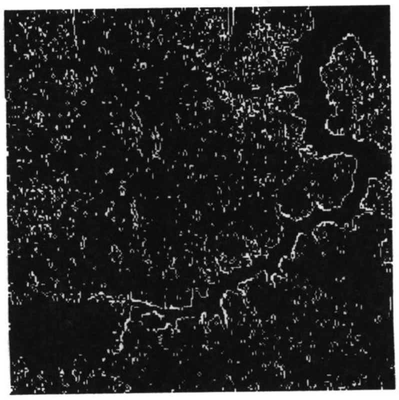


Figure 11: Local edge detection of real image.

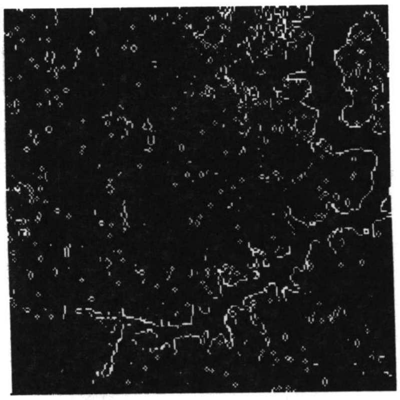


Figure 12: Probabilistic relaxation over real image after 15 iterations.

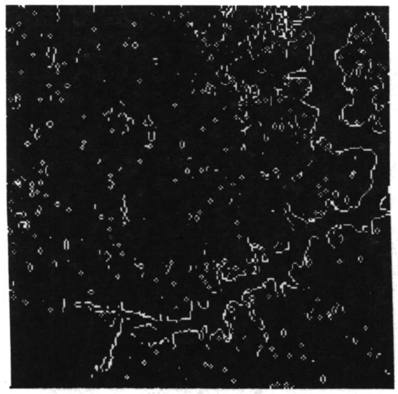


Figure 13: Dictionary-based relaxation over real image after 15 iterations.

The visual results with both relaxation schemes are similar but an elimination of isolated pixels and two pixel size segments can be obtained through a modification of the dictionary. The result is shown in figure 14 (CPU time for 1st iteration = 1409.3 s).

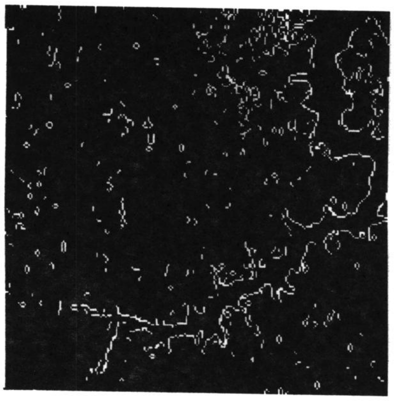


Figure 14: Modified dictionary-based relaxation over real image after 15 iterations.

5. Concluding Remarks

Although both iterative relaxation procedures are not guaranteed to converge, no deterioration on the labelling was observed with the increase in the number of iterations. On the other hand, both procedures displayed a tendency to achieve local consistency, but sometimes at the price of global inconsistency. This was observed, for example, in the case of gaps on the circumference, where the contextual information stabilized the decision at the non-edge label, before the global information of edge conveyed from both sides could be reached at that point.

As far as a comparison between both relaxation methods, one could state that the convergence speed per iteration is roughly the same, but the CPU time per iteration is much longer for the dictionary-based relaxation for the size of the dictionary. Yet, this method gives slightly better results under low signal-to-noise ratio and is able to be easily modified to emphasize or remove the support to certain spatial edge configurations.

6. Acknowledgement

The authors wish to acknowledge the support of FAPESP through the thematic Project Process 91/3532-2.

References

ALVES, A.H.H. Aplicação de técnicas de relaxação probabilística e baseada em dicionário em um algoritmo de deteção de bordas com formulação estatística, tese de mestrado, ITA, 1993.

GONZALEZ, R. C.; WOODS, R. E. *Digital image processing*, New York, Addison-Wesley, 1992.

HANCOCK, E. R.; KITTLER, J. Edge labeling using dictionary-based relaxation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2):165-181, Feb. 1990.

KITTLER, J. Relaxation labelling, in: *Pattern recognition theory and Applications*, Berlin, Springer-Verlag, 1987. pp. 99-108. (NATO ASI Series, Vol. F30)

LIM, J. S. *Two-dimensional signal and image processing*, New Jersey, Prentice Hall, 1990.

MASCARENHAS, N. D. A.; PRADO, L. O. C. A bayesian approach to edge detection in images, *IEEE Transactions on Automatic Control*, AC-25(1):36-43, Feb. 1980.