Conf-9111129--2

Received by: NOV 1 8 1991

SAND--91-2451C

DE92 002988

Segmentation of X-ray Images Using Probabilistic Relaxation Labeling

Tan Q. Thai
Sandia National Laboratories
Albuquerque, NM 87185

This paper presents a Probabilistic Relaxation Labeling (PRL) method to segment X-ray baggage images. PRL segmentation is an iterative algorithm that labels pixels in an image by cooperative use of two information sources: the pixel probability and the degree of certainty of its probability supported by the neighboring pixels.

Presented at the First International Symposium on Explosive Detection Technology, November 13-15, 1991

Abstract

Segmentation is a process of separating objects of interest from their background or from other objects in an image. Without a suitable segmentation scheme, it is very difficult to detect contraband in X-ray images. In this paper, a Probabilistic Relaxation Labeling (PRL) segmentation scheme is presented and compared with other segmentation methods. PRL segmentation is an iterative algorithm that labels each pixel in an image by cooperative use of two information sources: the pixel probability and the degree of certainty of its probability supported by the neighboring pixels. The practical implementation and results of the PRL segmentation on X-ray baggage images are also discussed and compared with other segmentation methods.

1. INTRODUCTION

One of the most important tasks of an image analysis system for contraband detection using X-ray images is to separate objects or regions of interest from their background or other objects in the image. Although segmentation of an image can be done with many different methods, there are three general approaches to the problem: 1) pixel-based approach, 2) region-based approach, and 3) model-based approach.

Pixel-based algorithms segment an image using information such as grey level, gradient magnitude or color of each pixel independently from its neighboring pixels. The pixels' information can then be used in a cummulative fashion as in histogram thresholding [3] or thresholding based on the degree of membership of the image pixels using fuzzy set concept [7]. *Region-based* approach takes into consideration the information of the neighboring pixels and their relations with the examined pixel. A pixel is assigned to the same region (class, cluster, etc.) if it has similar properties to its neighbors. One method of region-based approach is region growing [6]. Using this method the image is first divided into atomic regions of constant grey levels, then similar adjacent regions are merged sequentially until the adjacent regions become sufficiently different. Examples of region-based algorithms are the K-means clustering [12], the split-and-merge [6] and morphological segmentation using watershed transform with markers [4]. Unlike the pixel-based and region-based approaches that make no assumption about the image content and its noise, the *model-based* approach attempts to model both the image content and the image noise. Markov random field is often used to model the local properties and dependencies among neighboring pixels of regions in an image and Gaussian noise is assumed as the image noise mode [1,5].

Probabilistic relaxation labeling (PRL) segmentation is an iterative algorithm that labels each pixel in an image by cooperative use of two information sources: the pixel probability and the degree of certainty of its probability supported by the neighboring pixels. PRL algorithm can be considered as a hybrid approach of both region-based and model-based algorithms because of the dependency of each pixel to its neighbors in the labeling process and the assumption that each pixel can be assigned a probability index.

Details of the PRL algorithm are shown in section 2. Practical implementation of the PRL segmentation algorithm is discussed in section 3. Its application to X-ray baggage images and a comparison of the results with other methods are presented in section 4.

2. PROBABILISTIC RELAXATION LABELING SEGMENTATION

The idea of cooperative use of two sources of information in pixel classification though the mechanism of probabilistic relaxation was first developed by Rosenfeld et al. [11]. In essence the basic concept of the method is to iteratively reduce local ambiguities in classifying a pixel using local contextual information of the neighboring pixels. Its goal is to optimize a probabilistic index associated with a pixel; however, the method does not guarantee a unique optimal solution but rather seeks a practical suboptimal solution.

The relaxation labeling process is defined as the "best" assignment of a set of pixels $A = \{a_1, a_2, \ldots, a_N\}$ to a set of labels (or classes) $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_M\}$ where N is the total number of pixels in the image and M is the total number of labels (classes). Intially each pixel a_i is given a probability that it belongs to a label λ_l, $p_i(\lambda_l)$. These probabilities must sastify the following conditions

$$0 \leq p_i(\lambda) \leq 1 \quad \text{and} \quad \sum_{l=1}^{M} p_i(\lambda_l) = 1$$

For each pair of neighboring pixels a_i, a_j and each pair of labels λ_k, λ_l, we assume that there exists a measure of compatibility (compatibility coefficient) that $a_i \in \lambda_k$ and $a_j \in \lambda_l$. This measure of compatibility, denoted as $r_{ij}(\lambda_k, \lambda_l)$ has the following properties:

a) $-1 \leq r_{ij}(\lambda_k, \lambda_l) \leq 1$

b) If the assignment of pixels a_i, a_j to labels λ_k, λ_l is compatible then $r_{ij}(\lambda_k, \lambda_l) > 0$

c) If the assignment of pixels a_i, a_j to labels λ_k, λ_l is not compatible then $r_{ij}(\lambda_k, \lambda_l) < 0$

d) If the assignment of a_i to λ_k and a_j to λ_l is independent of each other then $r_{ij}(\lambda_k, \lambda_l) = 0$

The iterative update process of the image pixels incoorporates the information of both the pixel's initial probability and the influence of neighboring pixels based on the compatibility coefficients. One heuristic update process [10,11] is given as follows

$$p_i^{m+1}(\lambda) = \frac{p_i^m(\lambda)\,(1 + q_i^m(\lambda))}{\sum_{l=1}^{M} p_i^m(\lambda_l)\,(1 + q_i^m(\lambda_l))} \qquad (1)$$

where

$$q_i^m(\lambda) = \sum_{j=1}^{K} w_{ij} \sum_{l=1}^{M} r_{ij}(\lambda, \lambda_l)\, p_i^m(\lambda_l), \qquad (2)$$

m is the iteration number, K is the total number of neighboring pixels, w_{ij} are the weighting coefficients for the contribution of the neighboring pixels in the labeling process to pixel i, and M is the total number of labels (classes).

The update rule is simply a product of both p_i the pixel probability and q_i the degree of certainty of its probability supported by the neighboring pixels. Since the range of r_{ij}'s is [-1, 1], 1 is added to q_i to ensure that $p_i^{m+1}(\lambda)$ is always in the range of [0, 1]. The denominator of the update rule which is a normalizing factor uses to guarantee that the probability of pixel i is summed to 1 for all possible M labels.

Given the update rule, how can we determine the initial probabilities of all image pixels? What is a good number for the neighboring system in the update rule? How do we select the compatibility coefficients r_{ij}'s? Although different answers to these questions can lead to slightly different results in evaluating the relaxation update rule, the overall relaxation update should consistently reduce the ambiguity in the labeling process. Fekete et al [2] suggested two criteria to measure the performance of the probabilistic relaxation process: the rate of change between consecutive updates

$$R(m) = \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{l=1}^{M} \left(p_{il}^{m+1} - p_{il}^{m} \right)^2 \right)^{1/2} \qquad (3)$$

and the entropy of the pixel probabilities evaluated at each iteration

$$H(i, m) = - \sum_{l=1}^{M} p_{il}^m \ln p_{il}^m \qquad (4a)$$

$$H(m) = \sum_{i=1}^{N} H(i,m) \qquad (4b)$$

where N is the number of pixels in the image, M is the total number of labels and m is the iteration index. Instead of using the measurements as a performance index, we use them as convergence criteria for the update process.

2.1 Assignment of intial labeling probability

Most commonly the image's histogram is used to assign the initial labeling probability to each pixel in the image. A histogram represents the relative frequency of occurences of the grey levels in the image.

$$p(x_i) = \frac{n(x_i)}{N}, \qquad i = 0, 1, ..., L\text{-}1$$

where L is the number of grey levels, $n(x_i)$ is the number of occurence of pixels with grey level i in the image, and N is the total number of pixels in the image. This assignment, however, doesn't take into consideration the a-priori knowledge about the classes that we want to segment the image into. In our application of segmenting X-ray baggage images, the a-priori knowledge can be the average grey level of the X-ray images of a certain type of contraband that we want to extract from the input image. Since the a-priori knowledge is only an estimate, it is more appropriate to model the assignment of the initial probability by the S function [13, 7] which is defined as

$$S(x; a, b, c) = \begin{cases} 0, & x \leq a \\ 2\left(\frac{x-a}{c-a}\right)^2, & a \leq x \leq b \\ 1 - 2\left(\frac{x-a}{c-a}\right)^2, & b \leq x \leq c \\ 1, & x \geq c \end{cases} \qquad (5)$$

with $b = \frac{a+c}{2}$ and $a \leq b \leq c$. Figure 1 shows an example of the S function with the same crosspoint b for different values of a's and c's. If the crosspoint b is an estimate of the average grey level of an X-ray contraband image, then the S function defines the membership function corresponding to a fuzzy set "grey level x is similar to the contraband average grey level". The "spread" of the S function which controls by the distance (c - a) is a measure of fuzziness (uncertainty) in associating a grey level x to the contraband average grey level. As (c - a) approaches zero, the S function becomes a simple thresholding function about the crosspoint value b.

2.2 <u>Selection of compatibility coefficients</u>

The compatibility coefficients are the measures of the degree of support (or unsupport) for pairs of neighboring pixels in the process of assigning them to a set of labels. There are many ways to select these coefficients according to different definitions of compatibility functions [9].

A simplest selection of compatibility coefficients is to restrict them to the extreme values -1 and 1

$$r_{ij}(\lambda,\lambda') = +1 \qquad \text{if } \lambda = \lambda'$$

$$r_{ij}(\lambda,\lambda') = -1 \qquad \text{if } \lambda \neq \lambda'$$

Another way to estimate the coefficients is using conditional probabilities. Let $p_i(\lambda)$ be the initial estimate of the probability that a pixel i having label λ, then the probability of all pixels in the image having label λ is given by.

$$p(\lambda) = \frac{1}{N}\sum_{i=1}^{N} p_i(\lambda)$$

The joint probability of a pixel pair having label λ at pixel i and λ' at pixel j relative to i (e.g. pixel j is one pixel on the south side from i) is estimated as follows

$$p_{ij}(\lambda,\lambda') = \frac{1}{N}\sum_{i=1}^{N} p_i(\lambda)\, p_{i+j}(\lambda')$$

Given the above probabilities, the conditional probability is given by

$$p_{ij}(\lambda\,|\lambda') = \frac{p_{ij}(\lambda,\lambda')}{p(\lambda')} = \frac{\sum_{i=1}^{N} p_i(\lambda)\, p_{i+j}(\lambda')}{\sum_{i=1}^{N} p_i(\lambda')} \qquad (6)$$

Note that the estimate of the compatibility coefficients using conditional probabilities are global estimates because it is computed only once for a given neighboring configuration system. Once computed the coefficients are kept constant during the update process. For images with high textural details, local dependencies can be preserved by using a different estimate that applies only to local neighboring pixels. Such estimate can be computed by simply replacing N (total number of pixels in the image) with K (total number of pixels in the neighboring system). One drawback, however, is the significant increase in the overall computation of the algorithm. If conditional probabilities are used for compatibility coefficients, the relaxation update rule can be reformulated using Bayesian probability theory [8] with the assumption that the probability of pixel i given its label is λ is independent of its neighboring pixel's label. The relaxation update scheme is now given by

$$p_i^{m+1}(\lambda) = \frac{p_i^{m}(\lambda)\; q_i^{m}(\lambda)}{\sum_{l=1}^{M} p_i^{m}(\lambda_l)\, q_i^{m}(\lambda_l)} \qquad (7)$$

where q_i is in the same formulation as in Equ. (2). The update rule is basically the same as Equ. (1); the only difference is the elimination of 1 because the compatibility coefficients are [illegible] in the range of [0, 1] instead of [-1, 1] as before.

Experimentally we find that one set of compatibility coefficients does not work well for X-ray baggage images because the estimate is done globally. For example, if a bag is small and the background level is dominant then the estimated compatibility coefficients tend to bias toward the background. To compensate for the bias in the estimate, we propose the use of two update passes using two different set of compatibility coefficients. The first set of coefficients can be viewed as estimates from images with more background than baggages; while the second set is the estimates from images with cluttered baggage details. The biggest advantage of this method is that the compatibility coefficients do not have to be computed on-line for each image but they can be "learned" *off-line* from a number of different images.

<u>3. ALGORITHM IMPLEMENTATION</u>

Because the compatibility coefficients can be calculated off-line, the implementation of the PRL algorithm becomes very simple for the two class segmentation problem. Recal from Equ. (2), the local dependency information in the labeling process for the two class problem is given by a vector. (For notation convenience, from now on $q_i(\alpha)$ is written as $q_{i\alpha}$.)

$$\begin{bmatrix} q_{i\alpha}^{n} \\ \\ q_{i\beta}^{n} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{K} w_{ij}\, q_{j\alpha}^{n} \\ \sum_{j=1}^{K} w_{ij}\, q_{j\beta}^{n} \end{bmatrix} \qquad (8)$$

where

$$\begin{bmatrix} q_{j\alpha}^{n} \\ \\ q_{j\beta}^{n} \end{bmatrix} = \begin{bmatrix} r_{ij}(\alpha\alpha) & r_{ij}(\alpha\beta) \\ r_{ij}(\beta\alpha) & r_{ij}(\beta\beta) \end{bmatrix} \begin{bmatrix} p_{j\alpha}^{n} \\ \\ p_{j\beta}^{n} \end{bmatrix}$$

with α and β are the two classes. Let

$$c1 = r_{ij}(\alpha\alpha)$$
$$c2 = r_{ij}(\alpha\beta)$$
$$c3 = r_{ij}(\beta\alpha)$$
$$c4 = r_{ij}(\beta\beta)$$

then

$$q_{i\alpha} = c1 \sum_{j=1}^{K} w_{ij}\, p_{j\alpha} + c2 \sum_{j=1}^{K} w_{ij}\, p_{j\beta} \qquad (9)$$

since $p_{j\alpha} = 1 - p_{j\beta}$ and $\sum_{j=1}^{K} w_{ij} = 1$, we have

$$q_{i\alpha} = (c1 - c2) \sum_{j=1}^{K} w_{ij}\, p_{j\alpha} + c2 \qquad (10a)$$

Similarly,

$$q_{i\beta} = (c3 - c4) \sum_{j=1}^{K} w_{ij}\, p_{j\beta} + c4 \qquad (10b)$$

Therefore (7) becomes

$$p_{i\alpha}^{m+1} = \frac{p_{i\alpha}^{m}\, q_{i\alpha}^{m}}{p_{i\alpha}^{m}\, q_{i\alpha}^{m} + (1 - p_{i\alpha}^{m})\, q_{i\beta}^{m}} \qquad (11)$$

Theoretically, the range of pixel values in a PRL should be [0.0, 1.0]; however, for display purposes the pixel values must range from [0, 255] for an 8 bit image. The mapping from real values to integer values is simply done as follows

$$p'^{m}_{i\alpha} = \mathrm{ceil}(\, p_{i\alpha}^{m} * 255.0\,)$$

where p' is a displayed pixel value and ceil(x) is a function that returns a smallest integer which is greater than or equal to x. If we neglect the round off errors in the mapping, Equ. (11) can be implemented with a 64K integer look-up-table (LUT). The 64K size of the update rule LUT is used to account for all possible results in the product of p's and q's. Figure 2 shows that the entire PRL algorithm for two class problem can be implemented using two LUTs: a 256 byte LUT for the S function and a 64K LUT for the update rule. In addition, for each iteration of the PRL, one convolution is needed to calculate the summation term. Because both the LUT and convolution (especially with 3x3 and 5x5 masks) can be executed in real-time (30 frames/second) in many inexpensive image processing hardware, each iteration of the PRL can be completed in less than 80ms.

3. EXPERIMENTAL RESULTS

The experimental results are based on X-ray images that are taken from a local airport. The image size is 512 x 512 with 256 grey levels. In this paper, we only show the results from two X-ray baggage images using four different methods: PRL, adaptive histogram thresholding, constrast intensification using fuzzy set concept [7] and Iterated Conditional Modes (ICM) [1, 3]. The adaptive histogram thresholding is a simple, one pass segmentation using an adaptive threshold which is defined as the peak of the image's histogram that falls within the segmenting class' variance and has a certain height. (The peak height is used for size discrimation purpose.) The peak is found using the top-hat transformation [4]

$$\mathrm{Peak} = H - O(\,H\,)_n$$

where H is the image histogram and $O(H)_n$ is the opening of H by a line structure element of n pixels long. The opening operation is defined by an erosion followed by a dilation.

The constrast intensification using fuzzy concept is an iterative algorithm which attemps to assign individual pixel values into different classes based on their initial membership function. The assigment is done recursively with the following operation

$$T'(x) = \begin{cases} 2\,[T(x)]^2, & 0 \le x \le 0.5 \\ 1 - 2\,(1 - T(x))^2, & 0.5 \le x \le 1.0 \end{cases}$$

Details of the algorithm can be found in [7].

The ICM algorithm is a model-based method. In its simplest form, the method iteratively minimizes the following energy function

$$U = \frac{1}{2}\ln(\sigma_\alpha^2) + \frac{(x-\mu_\alpha)^2}{2\sigma_\alpha^2} + \sum_{r=1}^{c} \beta_\sigma \left[J(\alpha \rho_{+r}) + J(\alpha \rho_{-r})\right]$$

where α is the class label, σ_α^2 and μ_α are the variance and mean of class α, respectively, β_r is the clique[1] parameter, and J(a,b) is the spatial interaction among the clique neighbors.

3.1 Test parameters

All parameters of the four tested algorithms are fixed during the entire segmentation process. They are the followings.

- *Adaptive histogram thresholding:*
  - Structure element = {1, 1, 1, 1, 1}
  - Minimum peak's height = 800
  - $\mu_1 = 15$ (average grey level of class 1)
  - $\sigma_1 = 6.5$

- *Contrast intensification with fuzzy set:*
  - $\mu_1 = 15$
  - Fe = 2

- *Iterated conditional modes:*
  - $J(a, b) = 1$ if $a = b$
  - $= 0$ if $a \neq b$
  - $\mu_1 = 15$
  - $\mu_2 = 60$
  - $\sigma_1 = \sigma_2 = 6.5$
  - $\beta_r = 1.5$ for all r
  - c = 4 and the relative neigborhood configuration (r's) is as follows:

-3	-2	+4
-1	0	+1
-4	+2	+3

- *Probabilistic relaxation labeling:*
  - S function parameters: a = 0.0, c = 25.0
  - Compatibility coefficients for pass 1:
    $r_{00} = 0.62$; $r_{01} = 0.38$; $r_{10} = 0.57$; $r_{11} = 0.43$
  - Compatibility coefficients for pass 2:
    $r_{00} = 0.49$; $r_{01} = 0.51$; $r_{10} = 0.45$; $r_{11} = 0.55$
  - 3x3 convolution mask = {1, 1, 1, 1, 2, 1, 1, 1, 1}

[1] A clique is a set of points that are neighbors of each other.

Except the adaptive histogram threshold, the other three methods are iterative. The iteration loop is fixed to 7 for those methods.

3.2 Results and Discussion

Figure 3 and 8 show two original baggage X-ray images. The results of adaptive histogram threshold method are shown in Fig. 4 and 9. Only the final result from the last iteration are displayed for each iterative algorithms. Fig. 5 and 10 are the results of PRL segmentation. These images are taken from the second pass of the algorithm. The results of the contrast intensification using fuzzy set concept are shown in Fig. 6 and 11. Finally, Fig. 7 and 12 are the results from the ICM algorithm.

Assigning a "measure of performance" to the results of X-ray baggage images from different algorithms is a difficult task. The rate of change between consecutive updates and the entropy measurements proposed by Fekete et al. [2] do not guarantee the correctness of the segmentation results. For example, among the iterative algorithms, the rate of change between updates and entropy of the Iterated Conditional Modes method decreases faster than other methods; however, by visual inspection, one can conclude that its results are poorer than those of other methods. Using "true map" (a known segmentation result) then computing a deviation (e.g. mean square error) from the segmented results and the true map is also not a valid measure of performance because different parameters in the algorithms can give different results. Therefore, a fair comparison must not only be done with various possible combinations of the algorithms' parameters but also over a large number of data.

After testing the four different algorithms on a number of X-ray baggage images, the PRL is selected over other methods based on the following considerations:

1) Acceptable and consistent performance.
2) Capable of dealing with multiple classes (i.e. more than 2 classes).
3) Segmentation features can be color, texture, etc. instead of grey level values.
4) Ease of implementation.
5) Computational speed

4. CONCLUSION

The paper presents a method of segmenting X-ray baggage images using probabilistic relaxation labeling (PRL). To compensate for the bias in estimating the compatibility coefficients, we propose the use of two different set of coefficients for two update passes. These coefficients can be estimated off-line from a number of baggage images instead of on-line from the examined image as proposed in various literature. For a two class problem it is shown that the PRL can be implemented with two LUTs and a convolver. In addition, the algorithm can also be extended to segment contrabands based on other features such as colors, textures, lines and edges.

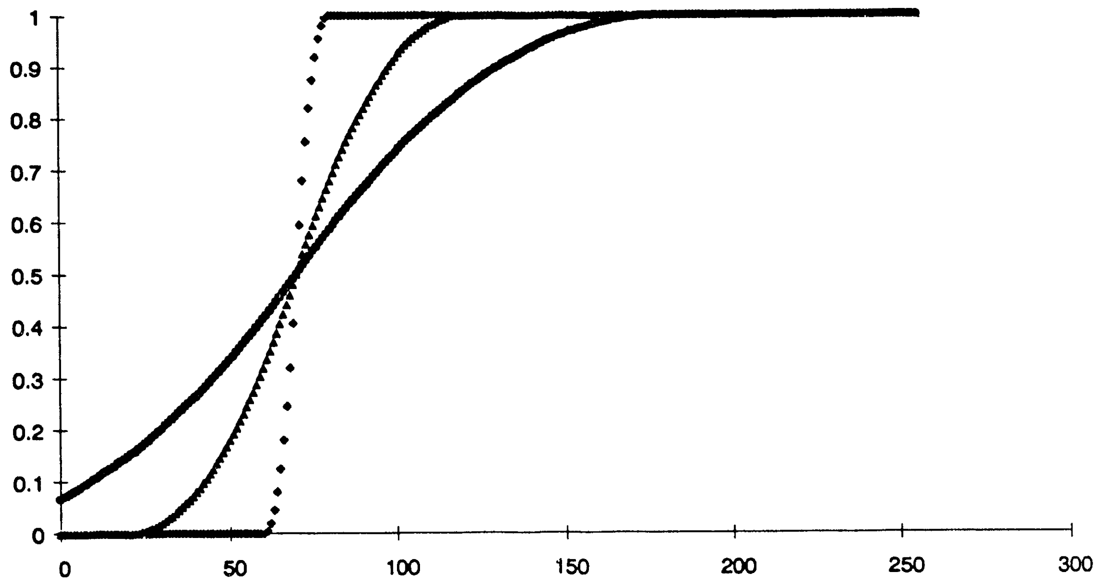

Figure 1: S function with different parameters

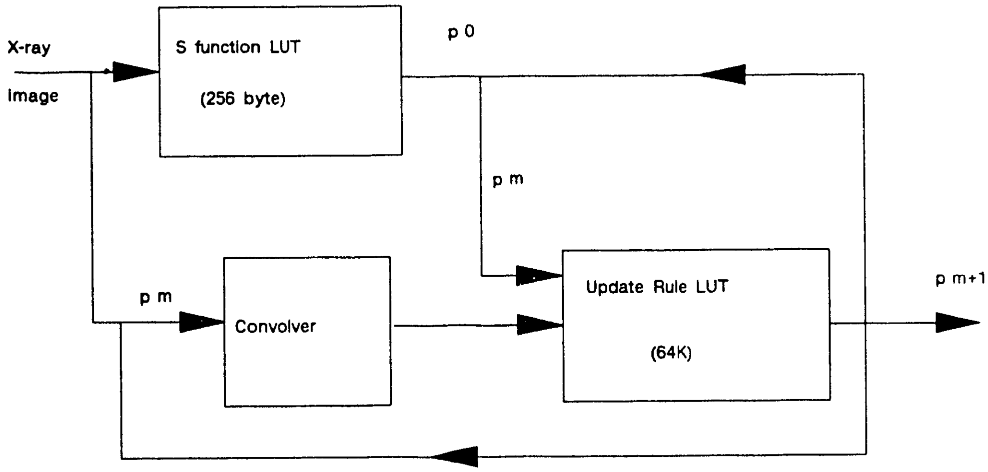

Figure 2: Implementation scheme of PRL method for two class problem.

REFERENCES

1. J. Besag, "On the statistical analysis of dirty pictures", J. Roy. Statist. Soc. B., vol. 48, no. 3, pp. 2181-2197, 1984.
2. G. Fekete, J. Eklundh, and A. Rosenfeld, "Relaxation: evaluation and applications", IEEE Trans. on PAMI, vol. 3, no. 4, pp. 459-469, July 1981.
3. K. V. Mardia and T. J. Hainsworth, "A spatial thresholding method for image segmentation", IEEE Trans. PAMI, vol. 10, pp. 919-927, 1989.
4. F. Meyer and S. Beucher, Morphological Segmentation, Centre de Morphologie Mathematique, School of Mines, Tech. Report, Apr. 1990.
5. T. N. Pappas and N. S. Jayant, "An adaptive clustering algorithm for image segmentation', IEEE Second International Conf. on Computer Vision, pp. 310-315, 1988.
6. T. Pavlidis, Algorithms for Graphics and Image Processing, Computer Science Press, 1982.
7. S. K. Pal and R. A. King, "Image enhancement using smoothing with fuzzy sets", IEEE Trans. Syst., Man, Cybern., vol. SMC-11, no. 7, pp. 494-501, July 1981.
8 S. Peleg, "A new probabilistic relaxation scheme", IEEE Trans. on PAMI, vol. 2, no. 4, July 1980.
9 S. Peleg and A. Rosenfeld, "Determining compatibility coefficients for curve enhancement relaxation processes," IEEE Trans. Syst., Man, Cybern., vol. SMC-8, no. 7, pp. 548-556, July 1978.
10. A. Rosenfeld, "Iterative methods in image analysis," Pattern Recognition, vol. 10, pp.181-187, 1978.
11. A. Rosenfeld, R. A. Hummel, and S. W. Zucker, "Scene labeling by relaxation operations," IEEE Trans. Syst., Man, Cybern., vol. SMC-6, pp.420-433, 1976.
12. J. Tou and R. Gonzales, Pattern Recognition Principles, Addison-Wesley, 1974.
13. L. A. Zadeh, K. S. Fu, K. Tanaka, and M. Shimura, Eds., Fuzzy Set and Their Applications to Cognitive and Decision Processes. Academic, 1975.

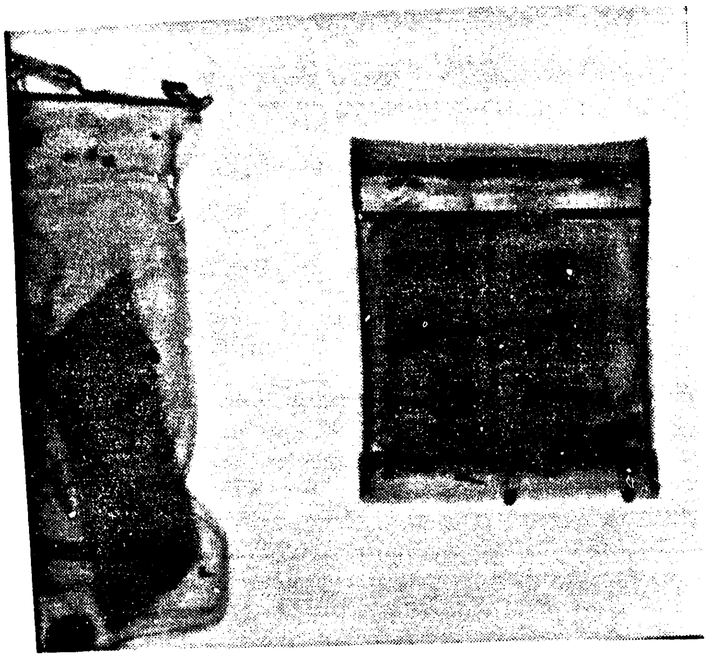

Figure 3: Original X-ray image 1

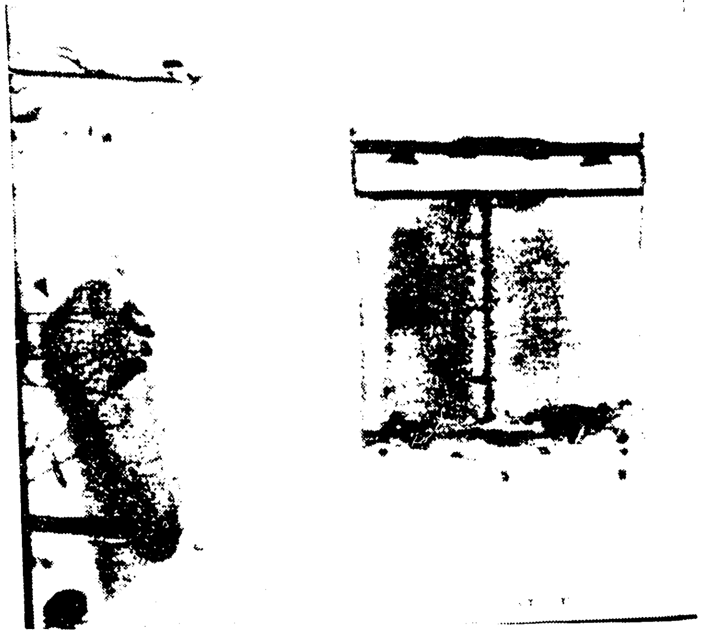

Figure 4: Result from adaptive histogram threshold on image 1

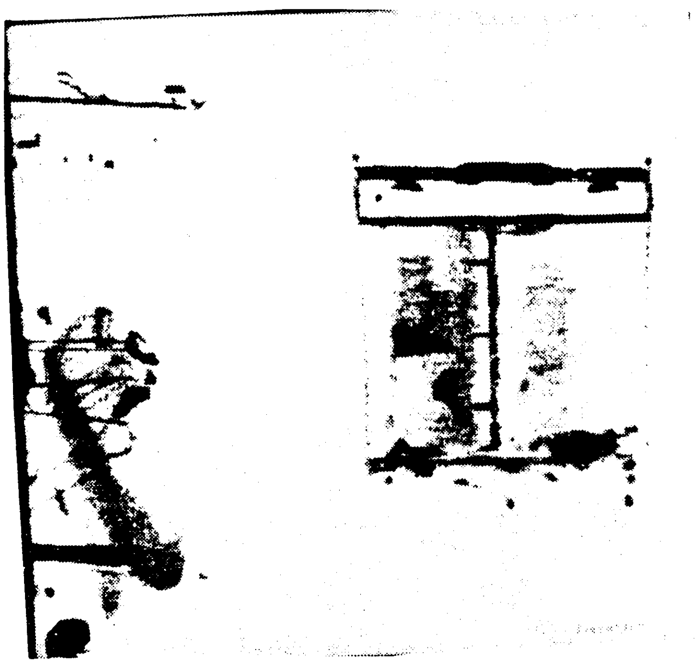

Figure 5: Result from PRL method on image 1

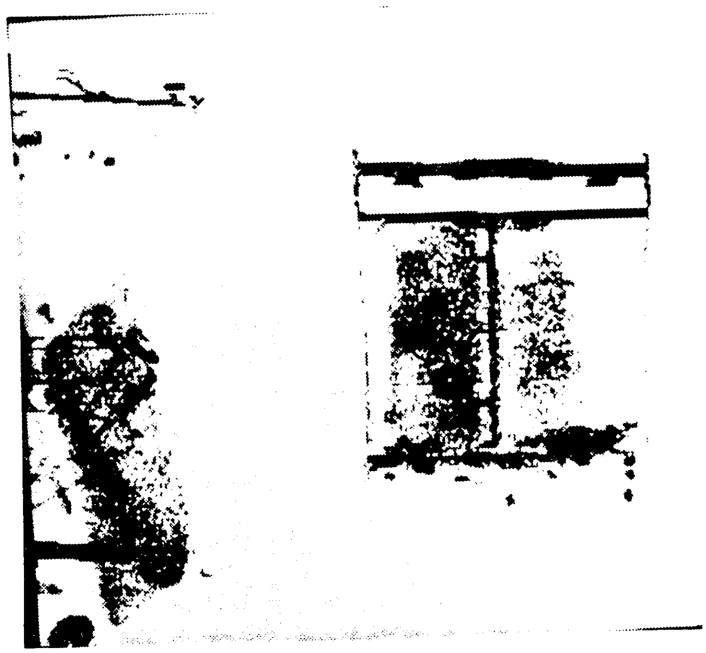

Figure 6: Result from constrast intensification on image 1

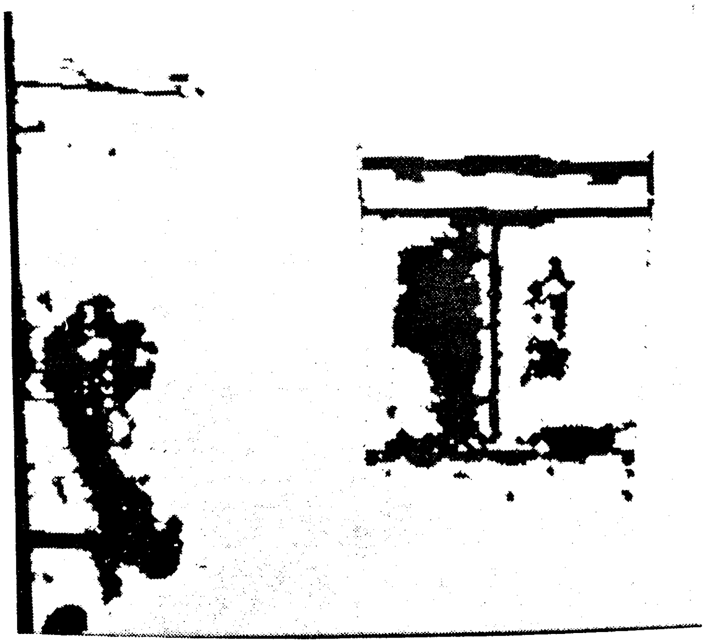

Figure 7: Result from the ICM method on image 1

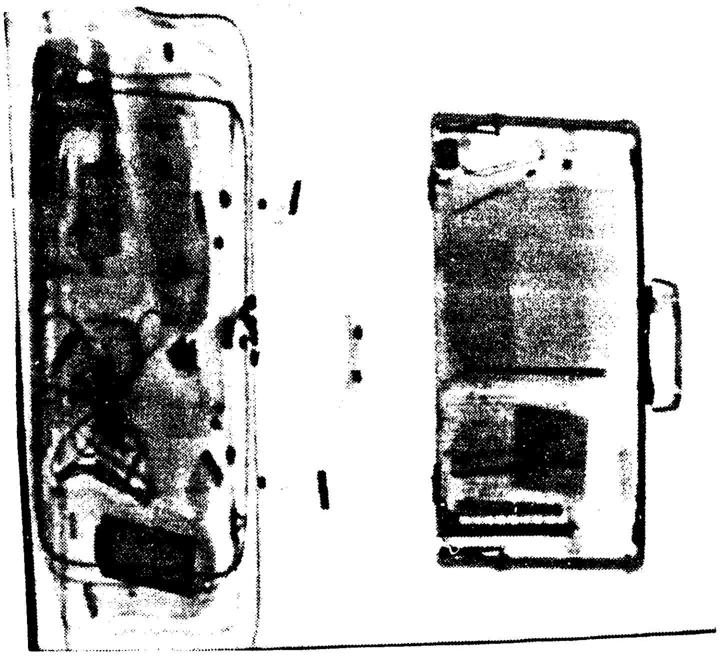

Figure 8: Original X-ray image 2

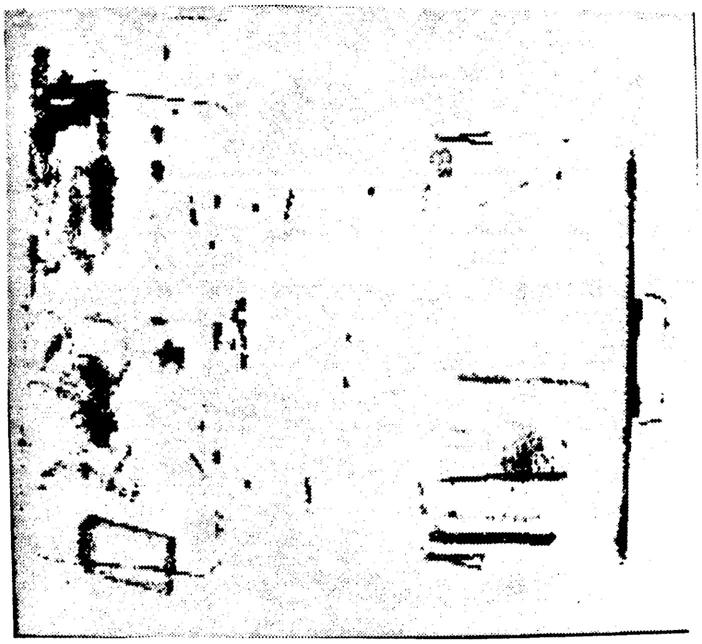

Figure 9: Result from adaptive histogram threshold on image 2

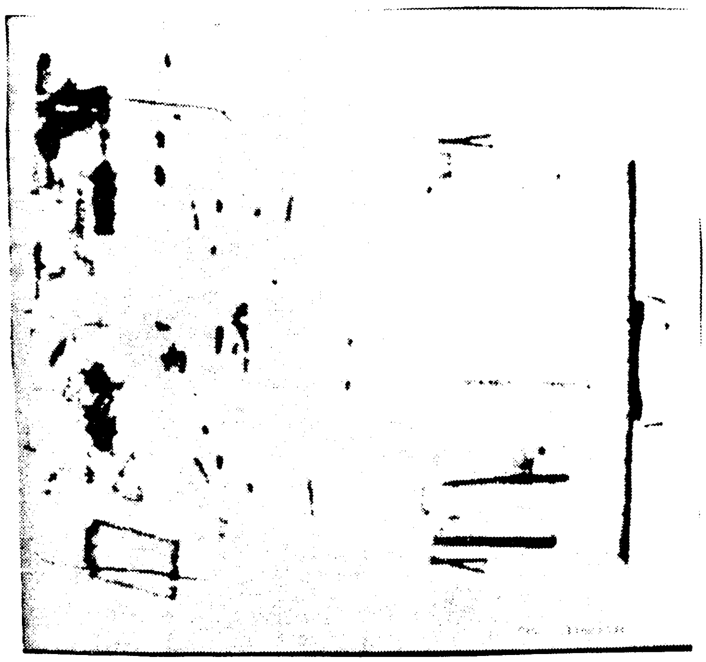

Figure 10: Result from the PRL method on image 2

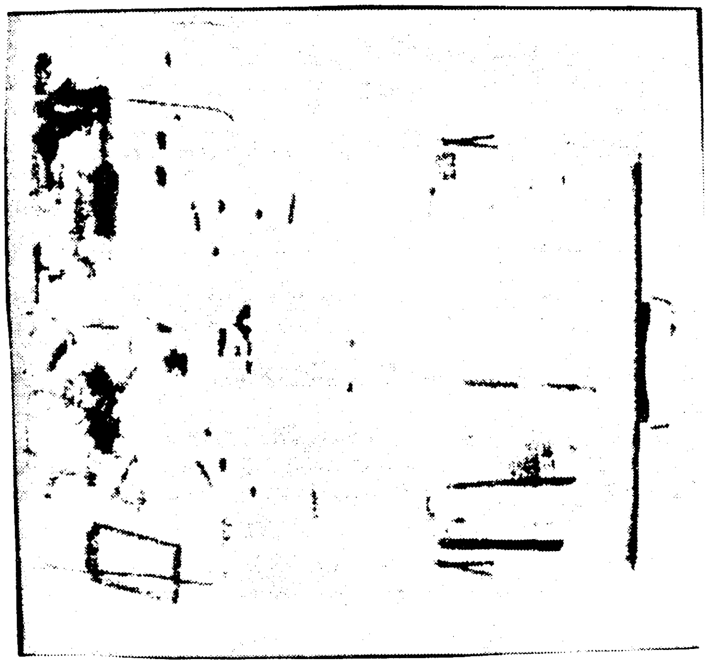

Figure 11: Result from constrast intensification on image 2

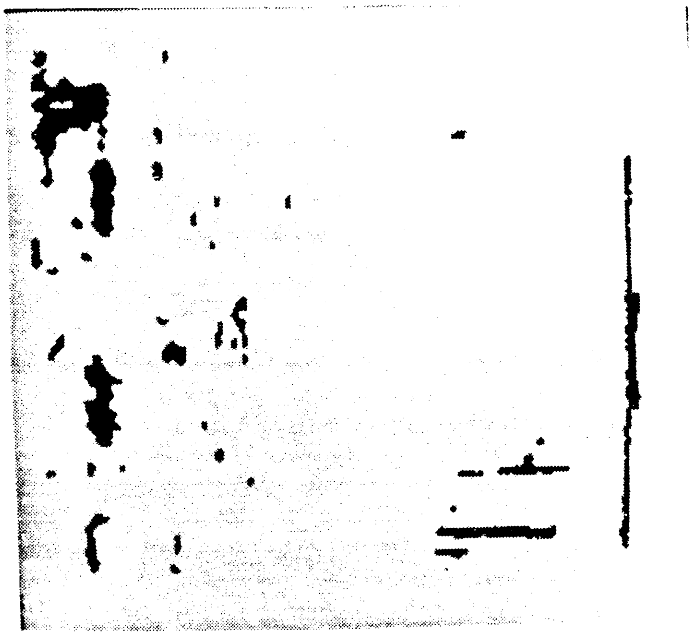

Figure 12: Result from the ICM method on image 2

DATE
FILMED
01/17/92

I