OVERCOMING ACCURACY DETERIORATION IN PIXEL RELAXATION LABELING

J. A. Richards,* D. A. Landgrebe and P. H. Swain†

*School of Electrical Engineering, University of New South Wales, Kensington, New South Wales, 2033 Australia; †School of Electrical Engineering, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, U.S.A. 47906

Abstract

In many applications of probabilistic label relaxation procedures, labeling error reduces with successive iterations as required, only to undergo a minimum and then rise again. If the process is not stopped, this can lead to an error in some cases worse than that initially. This behavavior is particularly significant in remote sensing pixel labeling applications since the iteration of minimum error cannot be ascertained because true labeling is not known. By comparison pixel relaxation labeling exercises in picture processing can often be effectively terminated by visual inspection of the image and its comparison with the desired labeling. This is the situation, for example, in noise removal. In this paper the common relaxation labeling algorithm is analysed whereby the technique is shown to degenerate to a mechanism of weighted averaging in the vicinity of fixed points. When uncontrolled it is demonstrated that this averaging can lead to the deterioration of labeling accuracy observed in practice. However it is also shown that the parameters in relaxation algorithms can be appropriately chosen to control the averaging and thus circumvent the accuracy deterioration problem. Examples are presented to support the analytical results derived. Furthermore, it is suggested that the parameters in the algorithm can be chosen <u>a priori</u>, based upon foreknowledge of image geometry.

Introduction

The results of simple exercises, such as the labeling of sides of a triangle,[1,2] have shown probabilistic relaxation procedures to be attractive techniques for reducing ambiguity and thus labeling errors in image data. In more complex labeling tasks however such as line and edge enhancement[3,4,5] and pixel labeling[6,7], the results obtained to date detract somewhat from the appeal of relaxation since labeling accuracy has been observed to improve during the early iterations of the process only to be followed by a subsequent degradation. In pixel labeling, for example, the labeling error exhibits a turning point at a specific iteration and the final error, in some situations, can be worse than that initially; similarly in line enhancement applications, line broadening is observed to occur late in the process, degrading an otherwise acceptable labeling. From a practical viewpoint, this suggests that the relaxation process in these sorts of applications should be stopped at some particular point to avoid incipient deterioration of the results. Alternatively, it may be better to try to understand the degradation mechanism so that the deterioration of labeling accuracy can be minimized or even avoided, irrespective of the iteration count. Eklundh and Rosenfeld[8] and Peleg[9] have addressed the task of determining suitable stopping rules. However since the reason for the turning point in the error curve has remained undetermined, there is no reason to suppose that stopping rules will circumvent accuracy deterioration. This paper is directed, therefore, towards understanding the mechanism that causes error to increase again after having achieved a minimum. It is shown that this is a process of local averaging once relaxation has approached a fixed point. Consequently, if the algorithm parameters are suitably chosen the error versus number of iterations curve can be made to decrease monitonically to a fixed error.

The Relaxation Algorithm

Consider the probabilistic relaxation algorithm of Rosenfeld, Hummel and Zucker:[1]

$$p_i^{k+1}(\lambda) = p_i^k(\lambda)Q_i^k(\lambda)/\sum_\lambda p_i^k(\lambda)Q_i^k(\lambda) \qquad (1)$$

where $p_i^k(\lambda)$ is the k^{th} estimate of the probability that λ is the proper label for the i^{th} pixel, and $Q_i^k(\lambda)$ is the k^{th} estimate of the neighborhood function, given by

$$Q_i^k(\lambda) = 1 + \sum_{j\epsilon J} d_j \sum_{\lambda'} r_{ij}(\lambda|\lambda')p_j^k(\lambda') \qquad (2)$$

In this expression $r_{ij}(\lambda|\lambda')$ are the compatibility coefficients, the d_j are a set of neighbor weights that can be used to give different neighbors differing degrees of influence in the neighborhood function, and J defines the neighborhood about the particular pixel being considered. The pixel under consideration can be regarded as a member of that neighborhood or else can be excluded; these variations are referred to here as <u>inclusive</u> and <u>exclusive</u> neighborhoods, respectively.

CH1499-3/80/0000-0061$00.75©1980 IEEE

Reprinted from 5th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITON, December 1980

Local Averaging in the Vicinity of Fixed Points and Its Effect on Geometric Features

Suppose a particular relaxation exercise has progressed to a point where the label estimates have all approached 0 or 1. (The stage where the label estimates are at 0 or 1 is called a fixed point in the process. Fixed points with $p_i^k(\lambda)$ other than 0 or 1 can occur; however, they are infrequent in pixel labeling and will not be considered here.) Within homogeneous regions, i.e., where all pixels in a neighborhood have the same predominant label, the mutual support offered among neighbors will not allow the label estimate on any particular pixel to alter by any significant amount. Infact, those estimates will simply move closer to their fixed points. However, the situation at boundaries such as corners of one region within another, can be quite different, as the following discussion reveals.

Consider a λ_1 pixel on the boundary between λ_1 and λ_2 regions. Evidently $p_i^k(\lambda_1)$ is the largest estimate for that pixel and it is reasonable to assume for such a λ_1, λ_2 neighborhood that $p_i^k(\lambda_1) > p_i^k(\lambda_2) >> p_i^k(\lambda_n)$, $\forall_n \neq 1, 2$. Now consider whether the label estimate $p_i^k(\lambda_1)$ will be strengthened or weakened as relaxation proceeds. To do this, it is sufficient to consider the relative strengths of the neighborhood functions as defined in (2). In particular, if

$$Q_i^k(\lambda_1) > Q_i^k(\lambda_2)$$

the λ_1 label will be strengthened at the next iteration; otherwise it will weaken. This will continue with subsequent iterations (since the label estimates at neighbors will not change by any significant amount). Should $Q_i^k(\lambda_2) > Q_i^k(\lambda_1)$, the repeated application of relaxation will ultimately lead to λ_2 being the favored label at the pixel, i.e., the λ_1 label will be removed by further iterations. Consequently, even though labeling error could have been reduced in establishing the λ_1 label on that pixel, it will now (gradually) increase owing to the loss of that label. To avoid this, it is necessary, therefore, to ensure (from (2)) that

$$1 + \sum_{j\varepsilon J} d_j \sum_{\lambda'} r_{ij}(\lambda_1|\lambda')p_j^k(\lambda') >$$

$$1 + \sum_{j\varepsilon J} d_j \sum_{\lambda'} r_{ij}(\lambda_2|\lambda')p_j^k(\lambda')$$

i.e., $$\sum_{j\varepsilon J} d_j \sum_{\lambda'} \{r_{ij}(\lambda_1|\lambda') - r_{ij}(\lambda_2|\lambda')\} p_j^k(\lambda') > 0 \qquad (3)$$

Note that the additive "1" in (2) has been of no significance in determining (3), so that (3) is a result general to all present relaxation algorithms which employ arithmetic averaging over the neighborhood, including particularly that where the $r_{ij}(\lambda|\lambda')$ are chosen on the basis of correlations or mutual information and also that where the $r_{ij}(\lambda|\lambda')$ are mapped to conditional probabilities in which case the "1" does not appear in (2).

The probability that the pixel's label could alter to that of a third class λ_3 has been ignored owing to the earlier assumptions regarding the relative strengths of the label estimates on that pixel.

Since it has been assumed that all the probability estimates are close to 0 or 1, (3) can be modified to

$$\sum_{j\varepsilon J} d_j \{r_{ij}(\lambda_1|\lambda_j) - r_{ij}(\lambda_2|\lambda_j)\} > 0 \qquad (4)$$

where λ_j is the preferred label on the j^{th} neighbor.

Now consider the neighborhood definition explicitly. Let J' be the exclusive neighborhood so that J:{J',i} where i is the pixel whose label is "currently" under consideration. Then (4) can be recast to give

$$d_i > \frac{\sum_{j\varepsilon J'} d_j\{r_{ij}(\lambda_1|\lambda_j) - r_{ij}(\lambda_2|\lambda_j)\}}{r_{ii}(\lambda_2|\lambda_1) - r_{ii}(\lambda_1|\lambda_1)} \qquad (5)$$

as the condition λ_1 be retained as the label for the i^{th} pixel.

To simplify further discussion, now consider some special cases of (5). First suppose the compatibilities $r_{ij}(\lambda|\lambda')$ have been chosen as conditional probabilities, and secondly consider only a two-label problem so that

$$r_{ij}(\lambda_2|\lambda_j) = p_{ij}(\lambda_2|\lambda_j) = 1 - p_{ij}(\lambda_1|\lambda_j).$$

Moreover it is logical that $p_{ii}(\lambda_1|\lambda_1) = 1$ (although $r_{ii}(\lambda_1|\lambda_1)$ based upon other compatibility definitions need not be unity) giving as the condition for avoiding loss of a λ_1 label on the border between λ_1 and λ_2 regions that

$$d_i > \sum_{j\varepsilon J'} d_j\{1 - 2p_{ij}(\lambda_1|\lambda_j)\} \qquad (6)$$

Now consider the particular choice of neighborhood shown in Fig. 1, and let the pixel under consideration be a corner pixel, as depicted. Suppose $d_j = d \; \forall j$, and further assume the compatibility coefficients $p_{ij}(\lambda_1|\lambda_j)$ are the same for each neighbor j of the corner pixel. In addition suppose the d_j have been chosen such that $\sum_j d_j = 1$. Such a choice is strictly only required when the

$r_{ij}(\lambda|\lambda')$ are chosen as correlations. However, it is a useful choice in general and here leads to $4d + d_i = 1$ so that we have

$$d_i > \eta(1+\eta)^{-1}, \; \eta = p_{ij}(\lambda_2|\lambda_2) - p_{ij}(\lambda_1|\lambda_1) \qquad (7)$$

as the required condition that λ_1 corner labels not be lost. This condition also applies to the preservation of single-pixel-wide λ_1 lines that pass through a λ_2 neighborhood. For the simple neighborhood chosen, the only other geometries that are subject to label conversion (deterioration) by the mechanism described are the ends of lines a single pixel wide, and single isolated pixels. From (6) it can be shown that the condition for the preservation of labels at the ends of lines of λ_1 within λ_2 regions is

$$d_i > \frac{3p_{ij}(\lambda_2|\lambda_2) - p_{ij}(\lambda_1|\lambda_1) - 1}{3p_{ij}(\lambda_2|\lambda_2) - p_{ij}(\lambda_1|\lambda_1) + 1} \; . \qquad (8)$$

Likewise, to preserve individual λ_1 labeled pixels in λ_2 regions, it is necessary that

$$d_i > \frac{2p_{ij}(\lambda_2|\lambda_2) - 1}{2p_{ij}(\lambda_2|\lambda_2)} \; . \qquad (9)$$

The predictions of (7-9) were checked using the data chosen in Fig. 2. This is assumed to be a portion of an image for which the compatibilities are $p_{ij}(W|W) = 0.700$, $p_{ij}(\underline{b}|\underline{b}) = 0.800$, where $\underline{b}$ implies blank. Using (7-9), the following conditions can be determined.

To avoid loss of:

1.	a W corner in a b region	$d_i > 0.091$
2.	a b corner in a W region	$d_i > -0.111$
3.	a W line end in a b region	$d_i > 0.259$
4.	a b line end in a W region	$d_i > 0.130$
5.	a W pixel in a b region	$d_i > 0.375$
6.	a b pixel in a W region	$d_i > 0.286$

Consequently we would expect that if

$d_i \simeq 0$,	only b corners retained
$d_i \simeq 0.100$	both b and W corners retained
$d_i \simeq 0.150$	the above plus b lines retained
$d_i \simeq 0.270$	the above plus W lines retained
$d_i \simeq 0.400$	all corners, lines and isolated pixels retained.

As seen in Fig. 2, these predictions are accurate. The image was initialized very close to a fixed point by choosing the initial label estimates as $p_i^o(\underline{b} \text{ or } W) = 0.99$, and thus could be regarded as an image which has approached that condition by some preceding iterations of relaxation; moreover, it is useful to suppose the initial labeling represents the true labeling since then the label conversions observed in Fig. 2 would represent the introduction of labeling errors.

To illustrate the applicability of the results above to real data situations an example using Landsat imagery was chosen. A two category classification (wheat and non-wheat) was performed for a 117 x 196 pixel multitemporal image of a region in Kansas. After classification the remaining error was found, from available ground truth data, to be 9.2%. Using the neighborhood configuration of Fig. 1 a number of relaxation trials using various values of d_i, were performed. In each case the image was initialised close to a fixed point to enable the preceding material to be assessed. Residual error versus number of iterations for these tests are shown in Fig. 3. As observed, for d_i less than 0.150 the error reaches a minimum after about 20 iterations and then steadily increases again. This would have been particularly severe had an exclusive neighborhood definition been used. For d_i in excess of 0.150 the error curve decreases monitonically but leads to a pessimistic result. If each neighbor in an inclusive neighborhood had been weighted equally as has often been the case in practice, this situation would occur. Values of d_i near 0.150 are seen to be optimum for this image with its compatibilities. Equations (7) through (9), along with the compatibilities determined from the ground truth data, indicate that $d_i = 0.150$ will cause individual wheat and non-wheat pixels to be removed during relaxation along with corners on wheat fields, whereas ends of lines a single pixel wide will be retained. Inspection of the ground truth map for the data shows this to be a reasonable action. Indeed, inspection of the ground truth and use of (7) through (9) would have led to an *a priori* assessment of an appropriate value for d_i in the vicinity of 0.150 for this data. In a real situation, where ground truth evidently is not available, values for d_i can still be determined based upon knowledge of likely geometries in the image data being considered.

Labeling Improvement During Relaxation

The intention of applying relaxation to an image is to improve upon a labeling which has been generated beforehand by some "imperfect" process. In endeavoring to examine the improvement, it is useful to view the situation in the following manner. The relaxation algorithm does not know, of course, which are the correct and which are the incorrect labels. It only "knows" which labels are consistent and which are inconsistent with their neighbors. Consequently, an image with initial labeling errors will be treated by the relaxation algorithm as though it were correctly labeled and the "improvement" which it creates is a conversion of locally inconsistent labels. This conversion will take place by mechanisms such as those described in the previous section and, in particular, for pixels that are close to fixed points, equations such as (7) through (9) can be used to describe labeling improvement in addition to likely degradation. Indeed, in the special case when an image is *intentionally* initialized close to a fixed point, those expressions can be used very accurately to describe the labeling improvement phase as well as any deterioration in the labeling that might occur. In such a situation, the predictions of (7) through (9) (for a two-label example) allow the value of d_i to be chosen relative to the compatibilities and other neighbor-weighting coefficients to ensure that some labels are intentionally converted (i.e., those in error), while others are retained. Such a situation is evident in the example of Fig. 3 where isolated pixels had their labels converted since they were considered to be largely erroneous, as were corners. Clearly the requirements for improvement and for avoiding degradation will often conflict in real image segments and, in order to obtain clean-up during relaxation, some correct labels may have to be sacrificed.

Relevance of Accurate Compatibilities

In view of the comments of the previous two sections, it is clear that control of a relaxation process lies significantly in equations of the type (7) through (9) for a two-label problem and similar (albeit more numerous) manifestations of (5) for a multi-label exercise. Consequently, in the removal of initial labeling errors and in avoiding label degradation, the actual values of the compatibility coefficients ($r_{ij}(\lambda|\lambda')$ or $p_{ij}(\lambda|\lambda')$) appear not to be important in pixel labeling so much as their values relative to each other and to d_i as described in (7) through (9). As a demonstration of this consider the example of Fig. 4. From the true labeling figure the actual compatibilities are found as $p_{ij}(W|W) = 0.500$ and $p_{ij}(\underline{b}|\underline{b}) = 0.875$. Suppose it is desired to remove individual W pixels and line ends but that W corners need to be preserved. Then (7) through (9), along with the values for the compatibilities show that d_i must be made greater than 0.273. On the other hand if the compatibilities are arbitrarily set at $p_{ij}(W|W) = 0.600$ and $p_{ij}(\underline{b}|\underline{b}) = 0.700$ then $d_i > 0.091$ is required. The sequence of iterations displayed in Fig. 4 is obtained for both sets of parameters, as noted, showing the required label conversions notwithstanding the "inaccuracy" of one set of compatibilities.

Discussion and Conclusions

The examples presented have shown that it is possible with image data to choose compatibilities and specific values of the neighbor weights d_j such that the relaxation process will converge to a near optimum error which will not subsequently increase owing to label conversion (degradation) mechanisms. However, the discussions above and the predictions of (5) through (9), of course, only hold exactly for an image that has approached a 0,1 fixed point and thus tacitly assumes that the local averaging that gives rise to the conversion of border labels takes place when the label estimates are all near 0 or 1. While this is indeed the case, averaging also takes place earlier when the label probability estimates are not quite so extreme. By initializing the label probabilities further from a fixed point, the predictions from equations such as (7) through (9) will be modified. Empirical tests carried out by the authors indicate that the prediction of (7), at least, is a lower bound.

Should the initial label estimates within a class be all different (as would happen, for example, if they are determined on the basis of Mahalanobis distance considerations in a classification[7]), some incorrectly (and weakly) labeled pixels will be removed preferentially during the early improvement phase in the relaxation. However, all label estimates will then move toward 0 or 1 and the remarks of Section 3, and above, regarding deterioration still apply in principle.

Should an exclusive neighborhood definition be used in a relaxation exercise, then $d_i = 0$ in (5) through (9). Thus λ_1 label deterioration of the types considered will occur unless the right-hand sides of those equations are less than zero. A little thought reveals that these equations can never be satisfied for all complementary pairs of neighbor geometry (i.e., λ_1 corners in λ_2 regions and λ_2 corners in λ_1 regions) so that label degradation leading to an increase in labeling error would always be expected to occur with conventional probabilistic relaxation algorithms applied to real imagery when used with exclusive neighborhood definitions.

From the results presented in the previous sections, it is evident that the compatibilities should not need to be accurately characteristic of a particular image. Rather, as noted, it is biases in the compatibilities, along with the value of d_i (relative to the weights on the other neighbors - here all taken to be the same) that substantially determine how relaxation will behave on particular image data, as demonstrated in Fig. 4. A little thought also reveals that for image data (of the Landsat type especially) the true compatibilities cannot be particularly significant since these are statistically averaged measures computed over the whole or even a part of an image where in fact some regions of an image may bear no geometric or statistical resemblance to other areas of that same image.

References

1. A. Rosenfeld, R. Hummel and S. Zucker, "Scene Labeling by Relaxation Algorithms", IEEE Trans. Sys. Man, Cyber, Vol. SMC-6, June 1976, pp. 420-433.

2. S. Zucker, E. Krishnamurthy and R. Haar, "Relaxation Processes for Scene Labeling: Convergence, Speed and Stability", IEEE Trans. Sys. Man, Cyber, Vol. SMC-8, Jan. 1978, pp. 41-48.

3. S. Zucker, R. Hummel and A. Rosenfeld, "An Application of Relaxation Labeling to Line and Curve Enhancement", IEEE Trans. Comp., C-26, April 1977, pp. 394-403, plus correction, Sept. 1977, pp. 900-929.

4. B. Schachter, A. Lev, S. Zucker and A. Rosenfeld, "An Application of Relaxation to Edge Reinforcement", IEEE Trans. Sys. Man, Cyber, SMC-7, Nov. 1977, pp. 813-816.

5. S. Peleg and A. Rosenfeld, "Determining Compatibility Coefficients for Curve Enhancement Relaxation Processes", IEEE Trans. Sys. Man, Cyber, SMC-8, July 1978, pp. 548-555.

6. A. Lev, S. Zucker and A. Rosenfeld, "Iterative Enhancement of Noisy Images", IEEE Trans. Sys. Man, Cyber, SMC-7, June 1977, pp. 435-442.

7. J. Eklundh, H. Yamamoto and S. Rosenfeld, "Relaxation Methods in Multispectral Pixel Classification", Technical Report 662, Computer Science Center, U. of Maryland, July 1978.

8. J. Eklundh and A. Rosenfeld, "Convergence Properties of Relaxation", Technical Report 701, Computer Science Center, University of Maryland, College Park, Oct. 1978.

9. S. Peleg, "Monitoring Relaxation Algorithms Using Labeling Evaluation", Technical Report 842, Computer Science Center, University of Maryland, College Park, Dec. 1979.

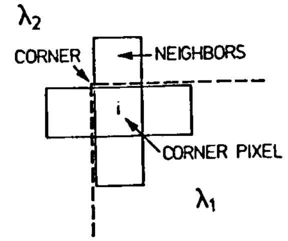


Fig. 1. Neighborhood definition used herein.

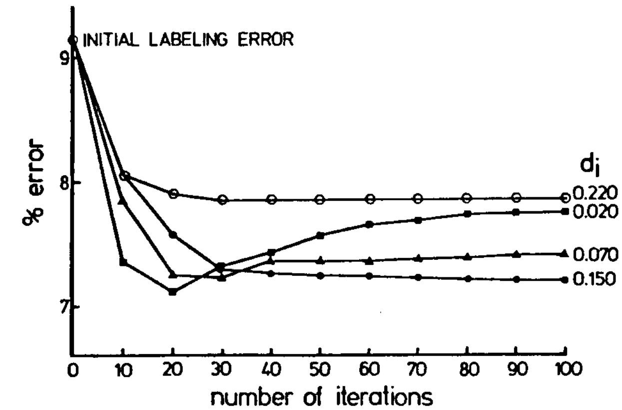


Fig. 3. Labeling error versus number of iterations when relaxation is applied to a 117 x 196 pixel image using several values of d_i. The original image was initialised very close to a fixed point.

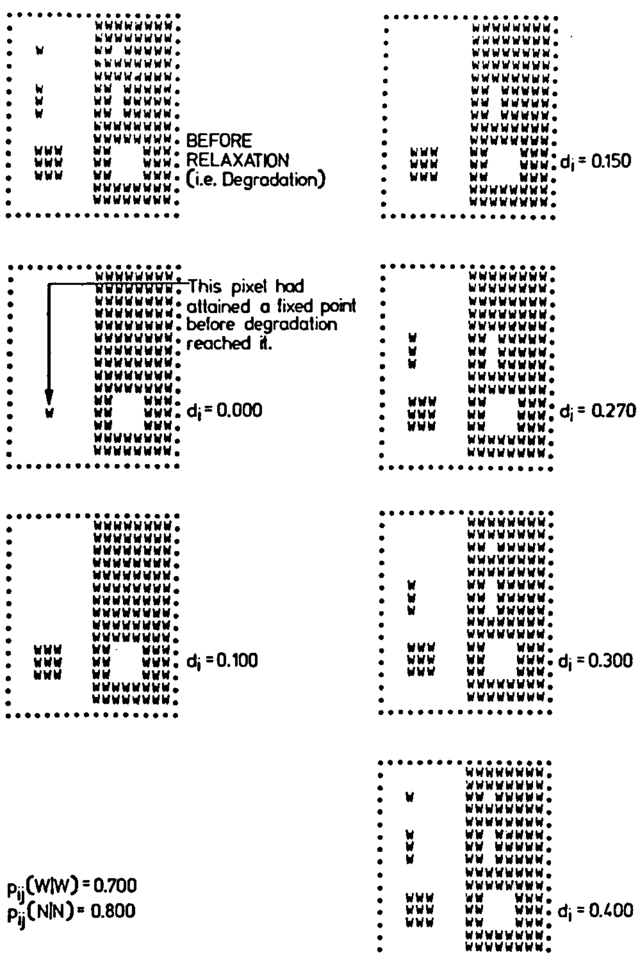
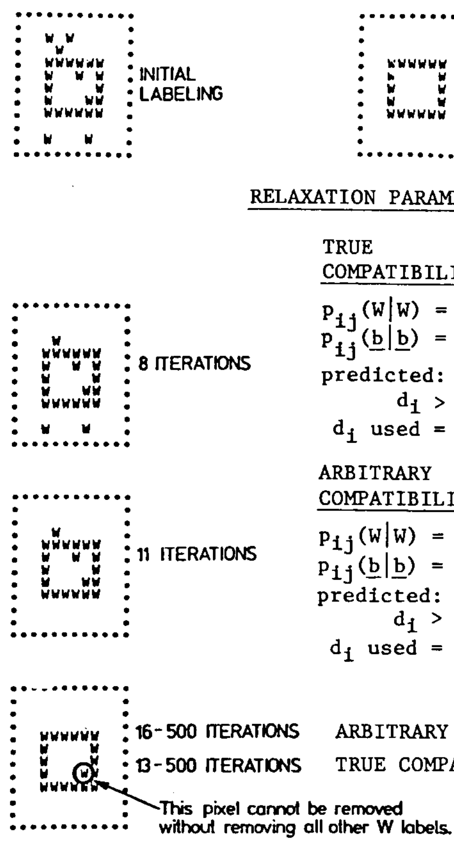


Fig. 2. Verification of predictions made using equations (7) through (9).



RELAXATION PARAMETERS USED

TRUE COMPATIBILITIES

$p_{ij}(W|W) = 0.500$
$p_{ij}(\underline{b}|\underline{b}) = 0.875$
predicted:
$d_i > 0.273$
d_i used $= 0.280$

ARBITRARY COMPATIBILITIES

$p_{ij}(W|W) = 0.600$
$p_{ij}(\underline{b}|\underline{b}) = 0.700$
predicted:
$d_i > 0.091$
d_i used $= 0.096$

16-500 ITERATIONS — ARBITRARY COMPATIBILITIES
13-500 ITERATIONS — TRUE COMPATIBILITIES

Fig. 4. Illustration of the effectiveness of arbitrary as against the true compatibilities when using relaxation labeling. Note that the final labeling was achieved in 13 iterations with the true compatibilities and 16 iterations with the chosen values. All other intermediate labelings occurred after the same number of iterations.