COMPUTER GRAPHICS AND IMAGE PROCESSING 13, 242–256 (1980)

Compatibilities and the Fixed Points of Arithmetic Relaxation Processes*

ROBERT M. HARALICK,† JOHN L. MOHAMMED, AND STEVEN W. ZUCKER

Computer Vision and Graphics Laboratory, Department of Electrical Engineering, McGill University, Montreal, H3A 2A7 Canada

Received September 5, 1979

There are two different sources of information in relaxation labeling processes: the initial certainty factors of the labels and the compatibility coefficients. In this paper we examine two ways in which the compatibility coefficients influence the fixed point achieved: (1) we demonstrate how the coefficients can bias the process toward the instantiation of a subset of the labels, and (2) we show how the coefficients precisely define the set of possible fixed points. We also indicate how eigenanalysis of the derivative of a relaxation labeling process at a fixed point can be used to study the stability of the fixed point. Finally, we present an empirical comparison of two statistical interpretations of the compatibility coefficients.

1. INTRODUCTION

Relaxation labeling processes are a class of parallel, iterative algorithms for reducing the sets of labels attached to nodes in a graph. They have most commonly been applied in computer vision problems, with the nodes indicating entities of some kind, and the labels indicating assertions about those entities. For example, the entities might be picture points, and the labels assertions about the presence of edges [3] or interpretive classifications [2]. (For a recent review of vision applications, see [9].) More recently, they have been applied in other domains such as handwriting analysis [5] and traffic light control [1].

The essential idea behind relaxation is the iterative use of context to effect the ambiguity reduction. In the continuous case, this is accomplished by updating label certainty factors on the basis of compatibility relationships between n-tuples of labels on m-tuples of neighboring nodes. (In this paper we consider pairs of labels on pairs of nodes; for a discussion of the full generality in the discrete case, see [4].) Thus, there are two ways in which information can enter a relaxation process: through the initial certainty factors and through the compatibility relationships. In this paper we are primarily concerned with the constraints that compatibilities exert on the final distribution of label certainty factors. Our discussion will end with a precise characterization of the fixed points (i.e., final certainty factor distributions) for a class of relaxation processes in terms of these compatibilities. First, however, after introducing relaxation more formally, we briefly discuss the importance of proper compatibility functions together with an associated design criterion.

2. RLPs WITH BALANCED COMPATIBILITY FUNCTIONS

The relaxation labeling process (RLP) that we consider here was suggested by Rosenfeld, Hummel, and Zucker in [10]. To define it, we use the random variable

*This research was supported by NSERC Grant No. A4470.

†Permanently with the Department of Electrical Engineering, Virginia Polytechnic Institute & State University, Blackburg, Virginia.

0146-664X/80/070242-15$02.00/0

notation often employed in statistics. Let $p^t\ (i = \lambda)$ be the certainty with which node i has label λ at iteration t. If there are l nodes with the label set Λ at each node, then the relaxation process defines how to compute $\{p^{t+1}\ (i = \lambda),\ i = 1,2,\ldots,l;\ \lambda = 1,2,\ldots,\ |\Lambda|\}$ from $\{p^t\ (i = \lambda),\ i = 1,2,\ldots,\ l;\ \lambda = 1,2,\ldots,\ |\Lambda|\}$.[1] The process is based on predefined and fixed node–label pair compatibility coefficients, $r\ (i = \lambda, j = \lambda')$, which indicate the compatibility between label λ on node i and λ' on j. The compatibilities are bounded within $[-1, +1]$. Furthermore, if nodes i and j do not constrain each other in any way, then $r\ (i = \lambda, j = \lambda') = 0$ for each label pair (λ, λ').

The supporting evidence, or contextual influence, given to label λ on node i by node j is given by

$$\sum_{\lambda' \in \Lambda} r(i = \lambda, j = \lambda') p^t(j = \lambda').$$

If we further let C_{ij} be the influence that node j can have on node i, where

$$0 \leq C_{ij} \leq 1 \qquad \text{and} \qquad \sum_j C_{ij} = 1,$$

then the supporting evidence given to λ on i by all nodes is given by

$$q^t(i = \lambda) = \sum_j C_{ij} \sum_{\lambda'} r(i = \lambda, j = \lambda') p^t(j = \lambda'). \tag{2.1}$$

The new label probabilities can be defined in terms of supporting evidence and the old label probabilities:

$$p^{t+1}(i = \lambda) = \frac{p^t(i = \lambda)\left[1 + q^t(i = \lambda)\right]}{\sum_\eta p^t(i = \eta)\left[1 + q^t(i = \eta)\right]}. \tag{2.2}$$

Since the updating rule defined by Eqs. (2.1) and (2.2) uses arithmetic averages, we refer to it as the arithmetic rule; for a discussion of families of related rules, see [13].

The two sources of information in the arithmetic rule are the initial certainty factors, $p^0(\cdot)$, and the compatibility matrices. Clearly, when the initial certainty factors are such that they contain no information, then the iterations (2.2) will be determined solely by biases introduced through the compatibility functions. If these compatibilities are also intended to introduce no explicit biases, then we have the:

Necessary design condition. An RLP with (i) unbiased compatibility functions and (ii) no information introduced through the initial probabilities should be at a fixed point. □

More precisely, the requirement of no information in the initial certainty factors translates into the uniform distribution:

$$p^0(i = \lambda) = 1/|\Lambda| \qquad \text{for all } i \text{ and } \lambda, \tag{2.3}$$

[1] $|A|$ designates the number of elements in the set A.

when there are $|\Lambda|$ possible labels. For (2.3) to be a fixed point, we must have

$$f_i \equiv \sum_{\mu} q(i = \mu) = q(i = \lambda) \qquad \text{for all } i \text{ and } \lambda. \tag{2.4}$$

Notice that the left-hand side of (2.4) is only a function of i, which means the supporting evidence for each label of node i must be the same. By substituting the expression defining q in Eq. (2.1) for the right-hand side of (2.4), we obtain the constraint

$$f_i = \sum_j C_{ij} \sum_{\lambda'} r(i = \lambda, j = \lambda') \tag{2.5}$$

for all i and λ. This constraint over the node interaction and the compatibility weights is sufficient to meet our necessary design criterion.

A simpler form of the constraint (2.5) arises for homogeneous RLPs.

DEFINITION. A relaxation labeling process is said to be *homogeneous* when each node has identical label sets, initial certainty factors, identical neighborhood relations, and identical compatibility functions. □

In this case, since the relationship to each neighbor is identical, the constraint (2.5) reduces to

$$h_i = \sum_{\lambda'} r(i = \lambda, j = \lambda'). \tag{2.6}$$

For the remainder of this paper, it is more suitable algebraically to consider compatibilities in the range [0,1] rather than [−1,1]. To emphasize this difference, we denote such compatibilities by $p\ (i = \lambda | j = \lambda')$, indicating again the compatibility between λ on i and λ' on j. If the $r\ (i = \lambda, j = \lambda')$ have already been specified, then the $p\ (i = \lambda | j = \lambda')$ can be obtained by the mapping:

$$p(i = \lambda | j = \lambda') = C_{ij} \frac{1 + r(i = \lambda, j = \lambda')}{\alpha} \tag{2.7}$$

for a suitably defined constant α. In this case we obtain,

$$\begin{aligned} 1 + q^t(i = \lambda) &= 1 + \sum_j C_{ij} \sum_{\lambda'} r(i = \lambda, j = \lambda') p^t(j = \lambda') \\ &= \alpha \sum_j \sum_{\lambda'} p(i = \lambda | j = \lambda') p^t(j = \lambda'), \end{aligned}$$

which, when substituted into Eq. (2.2), gives

$$p^{t+1}(i = \lambda) = \frac{p^t(i = \lambda) \sum_j \sum_{\lambda'} p(i = \lambda | j = \lambda') p^t(j = \lambda')}{\sum_{\eta} p^t(i = \eta) \sum_j \sum_{\lambda'} p(i = \eta | j = \lambda') p^t(j = \lambda')}. \tag{2.8}$$

It is this form of the arithmetic rule that we use to characterize the fixed points.

However, before proceeding, we would like to remark that compatibilities of this form suggest interpreting them as conditional (or subjective) Probabilities [7,14]. For this to be the case, we must also have

$$\sum_{\lambda} p(i=\lambda \mid j=\lambda') = 1 \qquad \text{for every } i,j,\lambda',$$

which implies the further constraint (with Eq. (2.7)):

$$\sum_{\lambda} C_{ij}[1 + r(i=\lambda, j=\lambda')] = \alpha \qquad \text{for every } i,j,\lambda'.$$

This is a significant constraint since the left-hand side is, in general, a function of i, j, and λ', while the right-hand side is a constant.

It should be stressed that, although statistical interpretations of the compatibility functions are possible, they are not necessary. However, if one wishes to adopt such a viewpoint, then it can be used to derive updating rules of similar (but not identical) form. For a discussion of one such derivation, see Peleg [7].

3. THE TRIANGLE LABELING EXAMPLE

In order to illustrate the ideas of the previous section, as well as to provide an example around which several new ideas can be motivated, consider the triangle labeling problem for a line drawing (see also [10, 12, 13]). In this example, which is a simplified Waltz-like line labeling problem [11], a line drawing is given of some real world scene. The problem is to interpret each line segment in the drawing with one of four interpretations: occluding edge, forward object above (λ_1); occluding edge, forward object below (λ_2); convex fold (λ_3); and concave fold (λ_4) (see Fig. 1.) In one part of the line drawing there is a triangle. The triangle labeling problem is to label or interpret the sides of the triangle. Interpretation of the triangle sides is possible because of constraints. Any pair of sides of a triangle can have only a pair of interpretations from a subset of all possible interpretation pairs. That subset is $\{\lambda_1\lambda_1, \lambda_2\lambda_2, \lambda_3\lambda_1, \lambda_4\lambda_2, \lambda_1\lambda_3, \lambda_2\lambda_4\}$.

The conditional probabilities of a triangle side having one interpretation given that it is adjacent to a side with a second interpretation can be obtained by assuming that all pairs are equally likely and counting the elementary events. We use this model for the compatibilities in this paper to obtain empirical results. Thus

$$p(i=\lambda \mid j=\lambda') = \begin{array}{c} \\ i=\lambda_1 \\ i=\lambda_2 \\ i=\lambda_3 \\ i=\lambda_4 \end{array} \begin{array}{c} \begin{array}{cccc} j=\lambda_1 & j=\lambda_2 & j=\lambda_3 & j=\lambda_4 \end{array} \\ \begin{bmatrix} 0.5 & 0 & 1 & 0 \\ 0 & 0.5 & 0 & 1 \\ 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \end{bmatrix} \end{array} . \tag{3.1}$$

Since the triangle labeling RLP is homogeneous, no information in the initial certainty factors corresponds to the uniform distribution $\{p^0(i=\lambda) = 0.25$, for all i and $\lambda\}$. If we let this RLP run, we see that it converges to a fixed point with

$$p^*(i=\lambda_1) = p^*(i=\lambda_2) = 0.5$$
$$p^*(i=\lambda_3) = p^*(i=\lambda_4) = 0 \qquad \text{for all } i.$$

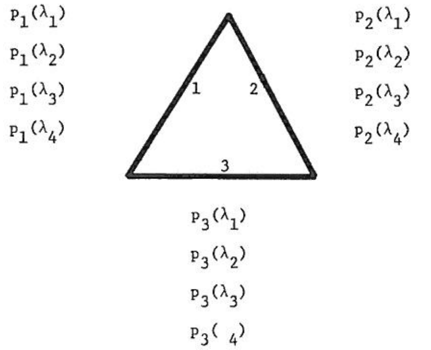
LABEL SET =

λ_1 : occluding edge : forward object below

λ_2 : occluding edge : forward object above

λ_3 : convex edge

λ_4 : concave edge



	$p(\lambda_1)$	$p(\lambda_2)$	$p(\lambda_3)$	$p(\lambda_4)$
SIDE 1				
SIDE 2				
SIDE 3				

FIG. 1. Labeling a triangle.

We refer to this limit as the "no-information fixed point" and denote it by p^*. Thus, the compatibility matrix contains a bias for λ_1 and λ_2, a fact that the designer of such processes certainly ought to know about. Note furthermore that this bias reflects the structure of the space of possible label pairs.

The above computation of the no-information fixed point raises two immediate questions. First, how stable is it? In other words, how much variation in the initial certainty factors is necessary to cause this particular RLP to converge to a different fixed point? Second, and more generally, since this fixed point is essentially caused by the compatibility values, is it possible to characterize this fixed point, and others, in terms of such compatibilities? The answers to these two questions will occupy us for the next two sections.

4. STABILITY OF THE FIXED POINT

To study the stability of the no-information fixed point of the triangle RLP, we now turn to an examination of the local neighborhood around this fixed point. To proceed, consider an equivalent model of the relaxation operator (2.8) in vector terms:

$$P^{k+1} = FP^k, \qquad k = 1, 2, \ldots, \tag{4.1}$$

where p is a ν-dimensional vector and F is the nonlinear mapping that implements the update (2.8) for each element of p. The dimension $\nu = l \times |\Lambda|$. For the triangle

RLP, $\nu = 3 \times 4 = 12$. The structure around a fixed point of (4.1) is revealed by the Fréchet derivative of the process:

$$F' = \frac{\partial p_\alpha^{k+1}}{\partial p_\beta^k} \qquad \alpha = 1, 2, \ldots, \nu; \quad \beta = 1, 2, \ldots, \nu.$$

This derivative can be used to determine whether (4.1) is a contraction mapping in the neighborhood of a given fixed point. Or, to put it another way, the Fréchet derivative evaluated at a fixed point can provide an indication of whether that fixed point can be approached from any direction, or whether it must be approached only from certain directions.

For the triangle RLP, the Fréchet derivative has the form

$$F' = \begin{bmatrix} \dfrac{\partial p^{t+1}(n1 = \lambda_1)}{\partial p^t(n1 = \lambda_1)} & \cdots & \dfrac{\partial p^{t+1}(n1 = \lambda_1)}{\partial p^t(n3 = \lambda_4)} \\ \dfrac{\partial p^{t+1}(n1 = \lambda_2)}{\partial p^t(n1 = \lambda_1)} & & \vdots \\ \vdots & & \\ \dfrac{\partial p^{t+1}(n3 = \lambda_4)}{\partial p^t(n1 = \lambda_1)} & & \dfrac{\partial p^{t+1}(n3 = \lambda_4)}{\partial p^t(n3 = \lambda_4)} \end{bmatrix}$$

where $n1$, $n2$, and $n3$ indicate nodes 1, 2, and 3 of the RLP, respectively.

If we evaluate F' at a particular fixed point P^f, we can obtain an understanding of the convergence properties of the process in an open neighborhood around P^f. The fixed point that we are immediately interested in is the no-information fixed point, P^*, computed in the previous section. Figure 2 lists F' at P^*.

FIXED POINT:

0.5	0.5	0.0	0.0
0.5	0.5	0.0	0.0
0.5	0.5	0.0	0.0

FRECHET DERIVATIVE:

0.5	-0.5	-0.5	-0.5	0.5	-0.5	1.0	-1.0	0.5	-0.5	1.0	-1.0
-0.5	0.5	-0.5	-0.5	-0.5	0.5	-1.0	1.0	-0.5	0.5	-1.0	1.0
0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.5	-0.5	1.0	-1.0	0.5	-0.5	-0.5	-0.5	0.5	-0.5	1.0	-1.0
-0.5	0.5	-1.0	1.0	-0.5	0.5	-0.5	-0.5	-0.5	0.5	-1.0	1.0
0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
0.5	-0.5	1.0	-1.0	0.5	-0.5	1.0	-1.0	0.5	-0.5	-0.5	-0.5
-0.5	0.5	-1.0	1.0	-0.5	0.5	-1.0	1.0	-0.5	0.5	-0.5	-0.5
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

FIG. 2. The Fréchet derivative of the triangle RLP evaluated at the "no-information" fixed point.

```
EIGENVALUES:  [0, 3, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1]

EIGENVECTORS:

1.  [-0.913, -0.183, 0.0, 0.0, 0.183, -0.183, 0.0, 0.0, 0.183, -0.183, 0.0, 0.0]
2.  [-0.408, 0.408,  0.0, 0.0, -0.408, 0.408, 0,0, 0.0, -0.408, 0.408, 0.0, 0.0]
3.  [0.255,  -0.843, 0.0, 0.0, -0.275, 0.275, 0.0, 0.0, -0.275, 0.275, 0.0, 0.0]
4.  [0.0,    -0.0,   0.0, 0.0, 0.167,  -0.167,0.0, 0.0,  0.5,   0.833, 0.0, 0.0]
5.  [0.0,     0.0,   0.0, 0.0, 0.833,  0.167, 0.0, 0.0, -0.5,   0.167, 0.0, 0.0]
6.  [0.0,    -0.0,   0.0, 0.0, 0.167,  0.833, 0.0, 0.0,  0.5,  -0.167, 0.0, 0.0]
7.  [-1.5,    0.5,   1.0, 0.0, -0.0,   0.0,   0.0, 0.0, -0.0,   0.0,   0.0, 0.0]
8.  [0.5,    -1.5,   0.0, 1.0,  0.0,  -0.0,   0.0, 0.0, 0.0,   -0.0,   0.0, 0.0]
9.  [0.0,     0.0,   0.0, 0.0, -1.5,   0.5,   1.0, 0.0, -0.0,  0.0,    0.0, 0.0]
10. [-0.0,   -0.0,   0.0, 0.0,  0.5,  -1.5,   0.0, 1.0,  0.0,  -0.0,   0.0, 0.0]
11. [ 0.0,    0.0,   0.0, 0.0, -0.0,   0.0,   0.0, 0.0, -1.5, 0.5,     1.0, 0.0]
12. [-0.0,   -0.0,   0.0, 0.0,  0.0,  -0.0,   0.0, 0.0,  0.5, -1.5,    0.0, 1.0]
```

FIG. 3. Eigenvalues and eigenvectors of the Fréchet derivative matrix in Fig. 1. Each entry is listed as an ordered pair with the real part followed by the imaginary part. The components are in the order: λ_1, λ_2, λ_3, λ_4, first for side 1, then side 2, and finally side 3.

ITERATION NO.		(a) λ_1	(a) λ_2	(a) λ_3	(a) λ_4	(b)	(b)	(b)	(b)
0	side 1	.3	.3	.2	.2	.1	.1	.3	.3
	side 2	.3	.3	.2	.2	.1	.1	.3	.3
	side 3	.3	.3	.2	.2	.1	.1	.3	.3
1		.39	.39	.11	.11	.25	.25	.25	.25
		.39	.39	.11	.11	.25	.25	.25	.25
		.39	.39	.11	.11	.25	.25	.25	.25
2		.42	.42	.08	.08	.38	.38	.13	.13
		.42	.42	.08	.08	.38	.38	.13	.13
		.42	.42	.08	.08	.38	.38	.13	.13
5		.46	.46	.04	.04	.45	.45	.05	.05
		.46	.46	.04	.04	.45	.45	.05	.05
		.46	.46	.04	.04	.45	.45	.05	.05
LIMIT		.5	.5	0	0	.5	.5	0	0
		.5	.5	0	0	.5	.5	0	0
		.5	.5	0	0	.5	.5	0	0

FIG. 4. Two initial distributions for the triangle RLP with equal biases for λ_1 and λ_2 at every node.

ITERATION NO.				
0	.4	.2	.2	.2
	.4	.2	.2	.2
	.4	.2	.2	.2
1	.57	.21	.14	.07
	.57	.21	.14	.07
	.57	.21	.14	.07
2	.74	.12	.12	.02
	.74	.12	.12	.02
	.74	.12	.12	.02
5	.92	0	.08	0
	.92	0	.08	0
	.92	0	.08	0
LIMIT	1	0	0	0
	1	0	0	0
	1	0	0	0

FIG. 5. Iterations toward a unique labeling.

To interpret this information, we make use of a theorem in Ostrowski [6]. This theorem states that if the spectral radius (the highest eigenvalue) $\rho(F' (P^{\mathrm{f}})) > 1$, then P^{f} is a point of repulsion for any sequence of iterates that enters a certain open neighborhood around it. That is, the sequence P^K, $K = 1, 2 \ldots$ will diverge from P^{f}. On the other hand, if $\rho(F,(P^{\mathrm{f}})) < 1$, the sequence will be attracted to P^{f}.

The eigenvalues of $F'(P^*)$ are given in Fig. 3, from which we see that $\rho(F'(P^*)) = 3 > 1$. Thus P_i^* is a point of repulsion. However, as we saw in Section 3, it is

ITERATION NO.				
0	.1	.1	.5	.3
	.1	.1	.5	.3
	.1	.1	.5	.3
1	.42	.27	.19	.12
	.42	.27	.19	.12
	.42	.27	.19	.12
2	.58	.23	.14	.05
	.58	.23	.14	.05
	.58	.23	.14	.05
5	.91	0	.09	0
	.91	0	.09	0
	.91	0	.09	0
LIMIT	1	0	0	0
	1	0	0	0
	1	0	0	0

FIG. 6. Cancellation effect of conflicting biases.

possible to obtain this fixed point. This seeming contradiction is clarified by looking at the eigenvectors of $F'(P^*)$ (see Fig. 3). These eigenvectors show that it is not the individual label probabilities that are important, but rather it is the relative differences between the labels (λ_1 and λ_2) and (λ_3 and λ_4) at each node that determine whether P^* will be achieved. As long as their likelihoods remain perfectly in balance, neither of the two occlusion labels will be selected over the other (see Fig. 4). However, as soon as one label is favored either directly (Fig. 5) or indirectly (Fig. 6), a unique interpretation obtains. The extreme sensitivity to even minor deviations from these balanced configurations is indicated by the magnitude of $\rho(F'(P^*))$, and brings to mind a decision process like the one studied in [13].

This discussion has shown that it is possible to analyze, at least in special circumstances, the structure imposed on an RLP by the compatibilities separately from the influence of the initial certainty factors. In the next section we provide a different, and more general, characterization of this structure.

5. A CHARACTERIZATION OF THE FIXED POINTS

Although both the compatibilities and the initial certainties affect the final fixed point of an RLP, in this section we show that the matrix of compatibilities restricts the set of possibilities in a nontrivial manner. In particular, we derive an algebraic expression over the compatibilities that the fixed point must satisfy, given that we know which label certainties go to zero. To facilitate understanding, we do the analysis first for homogeneous RLPs, and then for arbitrary ones.

The homogeneous case. By definition, in the special case of a homogeneous RLP, $p(i = \lambda \mid j = \lambda')$ is the same for every i and j; hence, we can write the compatibilities as $p(\lambda \mid \lambda')$. Also, since a homogeneous RLP is started with the same initial certainty factors at each node, it is easily seen that the fixed points achieved by such a process would also have the same distribution of certainty factors at each node.

Let $\{p(\lambda) \mid \lambda \in \Lambda\}$ be one of these fixed points. Further, let $\Lambda^+ = \{\lambda \mid p(\lambda) > 0, \lambda \in \Lambda\}$ be the set of labels having nonzero certainty factors at this fixed point. Every fixed point of (2.8) must satisfy (see also [12])

$$q(i = \lambda) = \sum_{\lambda' \in \Lambda} p(i = \lambda')q(i = \lambda') = \text{constant}, \qquad \forall \lambda \in \Lambda^+, \tag{5.1}$$

where

$$q(i = \lambda) = \sum_j \sum_{\lambda' \in \Lambda^+} p(i = \lambda \mid j = \lambda')p(j = \lambda') = \text{constant}, \qquad \forall \lambda \in \Lambda^+.$$

The homogeneity assumption simplifies this latter expression to

$$\sum_{\lambda' \in \Lambda^+} p(\lambda \mid \lambda')p(\lambda') = \text{constant} = \alpha, \qquad \forall \lambda \in \Lambda^+,$$

which can be rewritten in matrix form as

$$R_H P_H = \alpha I,$$

where R_H is the $|\Lambda^+| \times |\Lambda^+|$ matrix of compatibility functions, $p(\lambda|\lambda')$, for λ, $\lambda' \in \Lambda^+$; P_H is the $|\Lambda^+| \times 1$ vector of $p(\lambda)$, $\lambda \in \Lambda^+$; and I is a $|\Lambda^+| \times 1$ vector of all 1's.

If R_H is of full rank, then

$$P_H = \alpha R_H^{-1} I.$$

(If R_H is not of full rank, pseudo-inverses may be appropriate.) Moreover, since

$$\sum_\lambda p(\lambda) = 1 \tag{5.2}$$

at every node, we have

$$\alpha = \left(I' R_H^{-1} I\right)^{-1}$$

or

$$P_H = R_H^{-1} I / \left(I' R_H^{-1} I\right).$$

For comparison with the general case presented later in the paper, note that this can also be written

$$P_H = R_H^{-1}\left(I' R_H^{-} I\right)^{-1} I. \tag{5.3}$$

Thus, every fixed point P_H of a homogeneous RLP must satisfy Eq. (5.3), which is an explicit function of the compatibility values.

The above analysis can be extended to the more general situation in which the RLP is not homogeneous and the distribution of certainty factors is different at each node. Let neigh(i) be the set of nodes neighboring i, and Λ_i be the set of labels for node i. Now, Eq. (5.1) becomes

$$\sum_{j \in \text{neigh}(i)} \sum_{\lambda \in \Lambda_n} p(i = \lambda | j = \lambda') p(j = \lambda') = \alpha_i \tag{5.4}$$

for all nodes i and for all λ such that $p(i = \lambda) > 0$.

Note that α_i is constant for all $\lambda \in \Lambda_i^+$, but may be different at each node.

To write Eq. (5.4) in matrix form, we must first introduce the following notation. Let

$N_i = |\Lambda_i^+|$ be the number of labels with nonzero certainty factors at each node.

Then

$N = \sum_i N_i$ is the total number of labels in the RLP with nonzero certainty.

Also let P_g be the $N \times 1$ vector of $p(i = \lambda)$, $i = 1, 2, \ldots, l$, $\lambda \in \Lambda_i^+$; and let R_g be the $N \times N$ matrix of compatibilities, the entries of which are $p(i = \lambda \mid j = \lambda')$, where $p(i = \lambda)$ and $p(j = \lambda')$ are both nonzero. If we let the nodes be represented by the symbols $1, 2, \ldots, l$, and the labels $\lambda \in \Lambda_i^+$ be represented by $\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{in_i}$, where $n_i = |\Lambda_i^+|$ for each node i, then the form of R_g is given by

$$R_g = \begin{bmatrix} P(1 = \lambda_{11} | 1 = \lambda_{11}) \ldots P(1 = \lambda_{11} | 1 = \lambda_{1n_1}) \ldots \\ \ldots P(1 = \lambda_{11} | l = \lambda_{l1}) \ldots P(1 = \lambda_{11} | l = \lambda_{ln_l}) \\ \vdots \\ P(1 = \lambda_{1n_1} | 1 = \lambda_{11}) \ldots P(1 = \lambda_{1n_1} | 1 = \lambda_{1n_1}) \ldots \\ \ldots P(1 = \lambda_{1n_1} | l = \lambda_{l1}) \ldots P(1 = \lambda_{1n_1} | l = \lambda_{ln_l}) \\ \vdots \\ P(l = \lambda_{ln_l} | 1 = \lambda_{11}) \ldots P(l = \lambda_{ln_l} | 1 = \lambda_{1n_1}) \ldots \\ \ldots P(l = \lambda_{ln_l} | l = \lambda_{l1}) \ldots P(l = \lambda_{ln_l} | l = \lambda_{ln_l}) \end{bmatrix}$$

Let A_g be the $l \times 1$ vector of α_i. (Recall that there are l nodes.) We shall construct a $l \times N$ matrix $\mathcal{H}$ as follows: Let $h_1, h_2, \ldots, h_l$ be $N \times 1$ vectors of 1's and 0's defined by

$$h_1^t = (\underbrace{1, \ldots, 1}_{N_1}, 0, \ldots, 0)$$
$$h_2^t = (\underbrace{0, \ldots, 0}_{N_1}, \underbrace{1, \ldots, 1}_{N_2}, 0, \ldots, 0)$$
$$\vdots$$
$$h_{|l|}^t = (0, \ldots, 0, \underbrace{1, \ldots, 1}_{N_{|(l)|}}).$$

Then $\mathcal{H}$ is the matrix

$$\mathcal{H} = [h_1, h_2, \ldots, h_l].$$

With this notation, we can write (5.4) as

$$R_g P_g = \mathcal{H} A_g.$$

Again, if R_g is of full rank,

$$P_g = R_g^{-1} \mathcal{H} A_g. \tag{5.5}$$

At this point we reuse the constraint (5.2) that the certainty factors sum to 1 at every node to solve for the α_i. Note that this constraint can be written as

$$h_i' P_g = 1$$

at every node i. Applying this constraint to Eq. (5.5) yields

$$\mathcal{H}' P_g = \left(\mathcal{H}' R_g^{-1} \mathcal{H}\right) A_g = I,$$

which implies

$$A_g = \left(\mathcal{H}' R_g^{-1} \mathcal{H}\right)^{-1} I.$$

This result, together with Eq. (5.5), gives us the equation for determining the nonzero confidence levels:

$$P_g = R_g^{-1} \mathcal{H}\left(\mathcal{H}' R_g^{-1} \mathcal{H}\right)^{-1} I. \tag{5.6}$$

Hence, given that we know which labels have nonzero confidence levels at the fixed point, we can use Eq. (5.3) or Eq. (5.6) to determine exactly what these confidence levels are.

Furthermore, it should be noted that the above analyses imply an algorithm for determining the set of all possible fixed points of an RLP. Consider a partition of the set Λ of all possible labels on all nodes into two classes:

Λ^+ labels assumed to have nonzero confidence at the fixed point;

Λ^0 labels assumed to have zero confidence at the fixed point.

If we consider only partitions for which every node has at least one nonzero label, then the number of such partitions is

$$\pi_i \left(2^{|\Lambda|} - 1\right).$$

Some of these partitions may require that one or more labels which are totally incompatible with their neighbors belong to Λ^+. For such a partition there can be no fixed point which has nonzero confidence for all the labels in Λ^+ and zero confidence for all labels in Λ^0, since the incompatible labels would be eliminated in the ensuing iterations. In such cases the solution of Eq. (5.6), or Eq. (5.3) as appropriate, would require that α_i be zero for some i.

Thus, to determine all possible fixed points, one would apply Eq. (5.6) or Eq. (5.3) to each of the partitions and accept the solutions as fixed points only when all the α_i are greater than zero.

Finally, the analyses that led to Eqs. (5.3) and (5.6) assume that the matrix $R_{\mathcal{H}}$ and R_g are nonsingular. In the event that this assumption is not valid for a particular partition, then the fixed point corresponding to the partition may not be unique. That is, the equations may be underdetermined, admitting a family of solutions.

6. FURTHER REMARKS ON THE INTERPRETATION OF COMPATIBILITIES

The final test of any set of compatibilities will be how they perform in practice. Such empirical data can often lead to practical insights concerning the relative differences between two models for compatibilities, and we will now switch to this methodology for comparing the conditional probability interpretation in Section 3 with the "mutual information" interpretation proposed by Peleg [7]. This latter interpretation is being used in RLPs for, e.g., classifying multispectral images [2]. Moreover, this comparison is interesting because mutual information is a function of the conditional probabilities: the mutual information between two events A and B is defined to be

$$MI(A,B) = \log \frac{P(A|B)}{P(A)}.$$

The mutual information compatibilities for the triangle example computed by the method in [7] are

$$R_{MI} = \begin{bmatrix} 0.081 & -1 & 0.22 & -1 \\ -1 & 0.081 & -1 & 0.22 \\ 0.22 & -1 & -1 & -1 \\ -1 & 0.22 & -1 & -1 \end{bmatrix}$$

written in the same form as (3.1).

	(a)				(b)			
Initial Certainty factors	.25	.25	.25	.25	1	0	0	0
	.25	.25	.25	.25	1	0	0	0
	.25	.25	.25	.25	.25	.25	.25	.25
Limit: Mutual Information process	.5	.5	0	0	1	0	0	0
	.5	.5	0	0	1	0	0	0
	.5	.5	0	0	0	0	1	0
Limit: Conditional Probability Process	.5	.5	0	0	1	0	0	0
	.5	.5	0	0	1	0	0	0
	.5	.5	0	0	.5	0	.5	0

	(c)				(d)				(e)			
Initial certainty factors	.5	0	.5	0	.6	0	.4	0	.6	0	.4	0
	.7	.3	0	0	1	0	0	0	1	0	0	0
	.25	.25	.25	.25	.25	.25	.25	.25	.3	.25	.2	.25
Limit:mutual information process	0	0	1	0	1	0	0	0	0	0	1	0
	1	0	0	0	1	0	0	0	1	0	0	0
	1	0	0	0	0	0	1	0	1	0	0	0
Limit: conditional process	1	0	0	0	1	0	0	0	1	0	0	0
	1	0	0	0	1	0	0	0	1	0	0	0
	1	0	0	0	7/8	0	1/8	0	1	0	0	0

FIG. 7. Comparison of the mutual information and conditional probabilities in the triangle RLP.

Figure 7 contains a number of comparison examples for different initial certainty factor distributions. The first example, Fig. 7a, shows that both sets of compatibilities exhibit the same bias for uniform initial certainties. (These final values could also have been computed by the analytical methods of the previous section.) The next two examples, Figs. 7b and c, show that the mutual information coefficients contain a bias toward events that are a priori unlikely, while the conditional probabilities show a bias, in the second case, toward likely a priori labelings. Finally, as the last three examples show (Figs. 7c, d, and e), the mutual information coefficients maintain their predisposition toward unlikely events for different initial configurations.

7. SUMMARY AND CONCLUSIONS

There are two sources of information in a relaxation labeling process: the initial certainty factors and the compatibilities between labels on neighboring nodes. In this paper we have concentrated on the compatibility factors, and have studied them with respect to four issues. First, to characterize biases in the compatibilities, we proposed the idea of computing "no-information" fixed points. This was tied to the formulation of the first design criterion for compatibilities. Second, in studying one such fixed point, we attempted to characterize the local neighborhood around this fixed point with respect to its stability characteristics. Third, we characterized the fixed points of arithmetic of RLPs as an algebraic relation over the compatibilities, thereby expressing the compatibility biases in a more general way. Finally, we presented an empirical comparison of two statistical interpretations of compatibilities and illustrated their respective differences in a subjective manner for a triangle labeling problem. This comparison revealed a third kind of bias that is possibly introduced through the compatibilities—a bias for labelings that are either likely or unlikely a priori. It is our position that the designers of RLPs will need to understand all of these biases in terms of their specific problems before they can guarantee a successful application.

REFERENCES

1. R. S. Brooks and V. R. Lesser, Distributed problem solving using iterative refinement, Technical Report, Computer and Information Sciences, University of Massachusetts at Amherst, May 1979.
2. J. O. Eklundh, H. Yamamoto, and A. Rosenfeld, A relaxation method for multi-spectral pixel classification, *IEEE Trans. PAMI*, to appear.
3. A. Hanson and E. Riseman, Segmentation of natural scenes, in *Computer Vision Systems*, (A. Hanson and E. Riseman, Eds.), Academic Press, New York, 1978
4. R. M. Haralick and L. G. Shapiro, The consistent labeling problem: Part I, *IEEE Trans. PAMI* **PAMI-1**, 1979, 173–183.
5. K. C. Hayes, Reading handwritten words using hierarchical relaxation, Ph.D. Thesis, Computer Science Department, University of Maryland, College Park, July 1979
6. A. M. Ostrowski, *Solution of Equations and Systems of Equations*, Academic Press, New York, 1966.
7. S. Peleg, A new probabilistic relaxation scheme, TR-711, Computer Science Center, University of Maryland, College Park.
8. S. Peleg and A. Rosenfeld, Determining compatibility coefficients for curve enhancement relaxation processes, *IEEE Trans. Systems, Man, and Cyber.* **SMC-8**, 1978, 548–554.
9. A. Rosenfeld, Iterative methods in image analysis, Proc. IEEE Conf. Pattern Recognition and Image Proc., Troy, N. Y., 1977.
10. A. Rosenfeld, R. A. Hummel, and S. W. Zucker, Scene labeling by relaxation operations, *IEEE Trans. Systems, Man and Cybernetics*, **SMC-6**, 1976, 420–433.
11. D. Waltz, Generating semantic descriptions from drawings of scenes with shadows, AI Memo 271, Artificial Intelligence Lab., MIT, Cambridge, Mass., 1972.

12. S. W. Zucker, E. V. Krishnamurthy, and R. Haar, Relaxation operations for scene labeling: Convergence, speed and stability, *IEEE Trans. Systems, Man, and Cybernetics* **SMC-8**, 1978, 41–48.
13. S. W. Zucker, Y. G. Leclerc, and J. L. Mohammed, Continuous relaxation and local maxima selection: Conditions for equivalence, Proc. Sixth Int. Joint Conf. on Artificial Intelligence, Tokyo, 1979; Technical Report 78-15R, department of Electrical Engineering, McGill University, Montreal, 1978.
14. S. W. Zucker and J. L. Mohammed, Analysis of probabilistic relaxation processes, Proc. IEEE Conf. Pattern Recognition and Image Proc., Chicago, Ill., 1978.