Fuzzy Relaxation Labeling Reconsidered

Hermann Borotschnig, Axel Pinz, Isabelle Bloch

Abstract— **Relaxation labeling is an iterative scheme for parallel information fusion forcing neighborhood constraint satisfaction. In the seminal paper of Rosenfeld et.al.[1] the authors discussed various models for relaxation labeling: discrete, probabilistic, non-linear probabilistic and fuzzy. Since then many new relaxation labeling schemes have been proposed only a few of which are incorporating methodologies from fuzzy set theory. It is shown that this is partly due to the traditional, but questionable classification of relaxation labeling schemes. A new classification is introduced which clarifies the role of fuzzy relaxation labeling. The new classification scheme is complemented by the introduction of a whole range of new fuzzy relaxation operators.**

I. Introduction

Many different problems can be re-stated as the problem of assigning labels to objects given some relational constraints. Such problems can be tackled within the framework of (continuous) relaxation labeling, in which we have:

- a set O of objects i.
- a set Λ_i of possible labels λ for each object i.
- an assignment weight $p_{i\lambda} \in [0,1]$ for each possible labeling (i,λ).
- support values $R_{ij\lambda\mu}$ measuring the support (either positive or negative) for label λ at object i arising from label μ at object j, where it is understood that $i \neq j$. Whenever $R_{ij\lambda\mu}$ is symmetric, i.e. $R_{ij\lambda\mu} = R_{ji\mu\lambda}$, it can be interpreted to measure the compatibility of the labelings (i,λ) and (j,μ).

Continuous relaxation labeling [2] is an iterative scheme that updates the assignment weights $p_{i\lambda}$ according to the information contained in the set of compatibility weights $R_{ij\lambda\mu}$ and the current assignment weights. The aim is to sieve out assignments that are hardly compatible and to promote highly compatible assignments by reducing/increasing the assignment weights properly. The discrete formulation of relaxation labeling can be considered as the limiting case $p_{i\lambda} \in \{0,1\}$, $R_{ij\lambda\mu} \in \{0,1\}$.

Subsequently we denote by $\mathbf{p}$ the vector of all assignment weights $p_{i\lambda}$, by R the set of all support weights $R_{ij\lambda\mu}$ and by $\mathbf{p}_i$ the vector of assignment weights for one particular object i and all its possible labelings, that is $\mathbf{p}_i := (p_{i1},..,p_{im_i})$ with $m_i = card(\Lambda_i)$. In addition to the introduction of support weights $R_{ij\lambda\mu}$ with $j \neq i$ we may also use support weights $R_{ii\lambda\mu}$ that encode how strongly two labelings (i,λ) and (i,μ) of the same object i support each other. This quantity becomes especially interesting as soon as more than one labeling of a single object i may be assigned a high assignment weight.

We will usually assume $R_{ij\lambda\mu} \in [-1,1]$ for convenience. When discussing various approaches we will also mention the range of values allowed for $R_{ij\lambda\mu}$. If $R_{ij\lambda\mu} \in [-1,1]$ then $R_{ij\lambda\mu} = -1$ stands for total conflict while $R_{ij\lambda\mu} = +1$ signifies total support. $R_{ij\lambda\mu} = 0$ indicates indifference or no influence. If $R_{ij\lambda\mu} \in [0,1]$ then usually low values indicate low support (either because of conflict or because of indifference) and high values indicate strong support.

H.Borotschnig and A.Pinz are with the Institute for Computer Graphics and Vision, TU-Graz, Münzgrabenstraße 11, A-8010 Graz, Austria. {borotschnig,pinz}@icg.tu-graz.ac.at

I.Bloch is with the Département Images, ENST, F-75634 Paris Cedex 13, France. bloch@ima.enst.fr

II. Classifying Relaxation Schemes

Relaxation schemes fall into two broad classes: methods which incorporate a normalization constraint on the assignment weights, requiring them to sum to one for each object separately:

$$\sum_{\lambda=1}^{m_i} p_{i\lambda} = 1, \qquad \text{for all } i = 1,..,n \tag{1}$$

with $n = card(O), m_i = card(\Lambda_i)$. And, on the other hand, relaxation models that do not incorporate this constraint.

Following the terminology in [1],[3] the unanimous traditional classification of relaxation schemes equates all fuzzy models with the second class of models where there is no normalization constraint. Simultaneously one can find in the literature the inclination to call approaches of the first class "probabilistic models", again starting with the seminal paper of Rosenfeld [1].

In this paper we propose the following alternative terminology that leads to a reconsideration of the domain of fuzzy relaxation models:

- We call all approaches incorporating constraint (1) *normalized continuous relaxation models* or for short *normalized (relaxation) models.*
- Those that do not incorporate constraint (1) are called *non-normalized continuous relaxation models* or for short *non-normalized (relaxation) models.*

We will call a relaxation model

- a *fuzzy model* if fuzzy information and/or fuzzy fusion operations are used in establishing the relaxation scheme.
- a *probabilistic model* if arguments based on Bayesian reasoning are used to establish the relaxation scheme.
- an *optimization model* if optimization of a functional is used to derive the scheme.

Note that the domain of optimization models might overlap with the set of fuzzy models and probabilistic models.

It is clear that probabilistic models necessarily belong to the class of normalized models. The latter class, however, also contains optimization models, ad-hoc models, and, as will be shown in this paper: normalized fuzzy models.

0-7803-4863-X/98 $10.00©1998 IEEE

Since traditional terminology equates fuzzy models with non-normalized models it may appear natural that all non-normalized models are fuzzy models. However, the implication does not hold the other way round. Fuzzy models are not defined through a normalization constraint but through the use of fuzzy information and/or fuzzy fusion operations.

There is also the issue of interpretation involved: normalized models often assume that each object should get one label but we do not know which one. While this is a common interpretation of probability distributions we are not enforced to restrict normalization only to the regime of Bayesian reasoning. In fuzzy models we usually interpret the assignment weights as degrees to which a particular object assumes a specific label. These degrees of membership to a particular class of labels may be non-normalized or normalized in the sense of equation (1). If they are normalized one may interpret the weights as to what percentage an object belongs to a specific class [1] (e.g. in the mixed pixel/voxel case).

Within the regime of fuzzy information fusion we may also interpret assignment weights as confidence values for specific claims which again can be normalized or non-normalized. The real distinction must be drawn when deciding if a method uses fuzzy information or not.

III. Basics of Relaxation Labeling

Relaxation labeling amounts to updating the vector $\mathbf{p}$ using the following iteration scheme:

$$\mathbf{p}^{k+1} = F(\mathbf{p}^k, R) \qquad k = 0, 1, 2, 3, ... \tag{2}$$

with k denoting the number of the current iteration and $\mathbf{p}^0$ being the initial assignment weights which have been either obtained from some other information source or otherwise set to consistent but ad-hoc initial values. The support coefficients R have to be obtained either from knowledge about the modeled process or through learning.

The geometric structure of the weighted labeling space follows directly from the definitions of normalized and non-normalized relaxation. We define

$$\mathbf{K}^f := \{\ \mathbf{p} \mid 0 \le p_{i\lambda} \le 1, \quad \forall\, i = 1, .., n\}$$

and

$$\mathbf{K}^n := \{\ \mathbf{p} \mid 0 \le p_{i\lambda} \le 1 \quad \wedge \quad \sum_{\lambda=1}^{m_i} p_{i\lambda} = 1, \quad \forall\, i = 1, .., n\}.$$

Let $\mathbf{C}$ be the set of corner points of $\mathbf{K}^n$, i.e. the set of completely unambiguous labelings. We will write subscripts $\mathbf{K}_i^f$, $\mathbf{K}_i^n$, $\mathbf{C}_i$ to indicate the restriction of the above concepts to the labelings $p_{i\lambda}$ of a particular object i. Fig. 1 visualizes the spaces $\mathbf{K}^f$ and $\mathbf{K}^n$ in the simple case of one object $i = 1$ and three labels $\lambda_1..\lambda_3$. $\mathbf{K}^n$ is the simplex with corner point set $\mathbf{C} = \{(0,0,1);(0,1,0);(1,0,0)\}$. In the case of n-objects, m-labels $\mathbf{K}^n$ consists of n different $m-1$ dimensional simplexes. For non-normalized relaxation models (1) is not valid and the space $\mathbf{K}^f$ of possible assignment values consists of the n different m dimensional unit cubes enclosing the above simplexes.

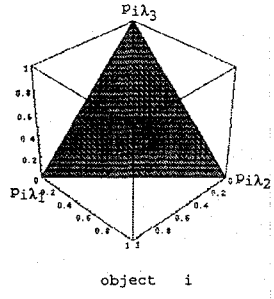

Fig. 1. *Exemplary locus of possible assignment weights* $\mathbf{p}_i = (p_{i\lambda_1}, p_{i\lambda_2}, p_{i\lambda_3})$ *for one object* i *in normalized relaxation.*

There are a few desirable common sense characteristics every relaxation operator F should possess [1]. Figure 2 summarizes some important properties of F in tabular form. In that figure "+" means $p_{i\lambda}$ should increase, "−" means that it should decrease, and "0" that it should remain relatively unchanged. For example, the assignment weight $p_{i\lambda}$ should be increased ("+") if another object j has a high assignment weight for a label μ, i.e. $p_{j\mu}$ is high, and the support $R_{ij\lambda\mu}$ for (i,λ) from (j,μ) is high too. Thus the signs indicate the influence of particular situations on $\Delta p_{i\lambda} := p_{i\lambda}^{k+1} - p_{i\lambda}^k$. The total change $\Delta p_{i\lambda}$ results from fusing the individual influences exerted by each labeling (j,μ).

Support for (i,λ) from (j,μ)
$R_{ij\lambda\mu}$

$p_{j\mu}$ \ $\Delta p_{i\lambda}$	High	Low
High	+	−
Low	0	0

Fig. 2. *Table of qualitative rules for the change in assignment weight* $\Delta p_{i\lambda} = p_{i\lambda}^{k+1} - p_{i\lambda}^k$ *of the labeling* (i,λ) *induced by the labeling* (j,μ).

IV. Normalized Models

In this section a few, representative, normalized relaxation models will be reviewed (see [4] for an extensive survey). The first normalized model which is still widely used has been proposed by Rosenfeld [1], who has termed it "non-linear probabilistic model". In light of the terminology established in section II the model belongs to the class of optimization models or ad-hoc models depending on the properties of R.

Computation consists of two major steps. The first one computes

$$q_{i\lambda} := \sum_{j \neq i}^{n} \sum_{\mu=1}^{m_j} R_{ij\lambda\mu} p_{j\mu} \tag{3}$$

which is interpreted to give the total support for (i,λ) from all other objects j. Then new assignment values

[1] Note that the same interpretation is used in the FCM-algorithms.

are defined to replace the current values according to the formula

$$p_{i\lambda}^{k+1} := \frac{p_{i\lambda}^k(1+q_{i\lambda}^k)}{\sum_{\nu=1}^{m_i} p_{i\nu}^k(1+q_{i\nu}^k)} \tag{4}$$

The denominator in eq. (4) ensures eq. (1) while the numerator updates $p_{i\lambda}$, with $1+q_{i\lambda}$ being the relative change of $p_{i\lambda}$. We let $R_{ij\lambda\nu} \in [-\frac{1}{n}, \frac{1}{n}]$ to keep $|q_{i\lambda}^k| < 1$. [2]

It is easy to show [5] that in the case of symmetric $R_{ij\lambda\mu}$, equation (3) is the gradient of the "gain function"

$$G(\mathbf{p}) = \frac{1}{2}\sum_i\sum_j\sum_\lambda\sum_\mu R_{ij\lambda\mu}p_{i\lambda}p_{j\mu}. \tag{5}$$

With the Baum-Eagon theorem it is therefore possible to show that the updating fixed point iteration in (4) performs a sort of gradient ascent [6]. But, of course, the above iteration equation is only one of many choices to do this. Indeed, most relaxation labeling algorithms with symmetric $R_{ij\lambda\mu}$ that use eq. (3) essentially perform constrained maximization of $G(\mathbf{p})$ based on the gradient $\nabla G(\mathbf{p})$ and therefore belong the the class of optimization models. For these models maximizing $G(\mathbf{p})$ is thus equivalent to increasing the consistency of the labeling $\mathbf{p}$ using the $R_{ij\lambda\mu}$. Faugeras and Berthod [7] and Hummel and Zucker [2] have investigated the general form of relaxation schemes based on step-wise constrained optimization by gradient projection:

$$\mathbf{p}_i^{k+1} := \mathbf{p}_i^k + \alpha\Psi(\mathbf{p}_i^k, \mathbf{q}_i^k) \tag{6}$$

where $\Psi(\mathbf{p}_i^k, \mathbf{q}_i^k)$ is an operation that performs gradient projection and α is a small step-size. The operator $\alpha\Psi$ returns an updating vector which ensures that the updated vector $\mathbf{p}_i^{k+1}$ lies within $\mathbf{K}^n$, i.e. fulfills eq. (1). We will use below also the (conceptually and implementationally much simpler) projection operator Ψ which projects on $\mathbf{K}^f$.

For non-symmetric $R_{ij\lambda\nu}$, the interpretation in terms of maximizing a gain function breaks down. Hummel and Zucker [2] have shown that the above schemes will still lead to consistent labelings (with consistency appropriately defined). These arguments replace the previous heuristics given by Rosenfeld and summarized in figure 2. Another approach that is based on Bayesian reasoning for only one single step of the whole relaxation process has been widely used, notably by Haralick [8] and Christmas and Kittler [9]:

$$p_{i\lambda} := \frac{p_{i\lambda}\left(\prod_{j=1}^n\sum_{\mu=1}^m R_{ij\lambda\mu}p_{j\mu}\right)}{\sum_\lambda p_{i\lambda}\left(\prod_j\sum_\mu R_{ij\lambda\mu}p_{j\mu}\right)} \tag{7}$$

In eq. (7) the $R_{ij\lambda\mu}$ are positive quantities measuring the ratio of the probability of the labeling (j,μ) conditional on the given labeling (i,λ) , to the unconditional probability of that labeling [9].

[2]Of course it is also possible to retain $R_{ij\lambda\nu} \in [-1,1]$ by using appropriate weight factors d_{ij} in eq. (3) as was done in [1].

Relaxation schemes are built to reduce ambiguity and to improve consistency [7]. Doing normalized relaxation, ambiguity reduction amounts to slowly changing $\mathbf{p}_i$ towards one of the unit vectors in $\mathbf{C}_i$, which are the only completely unambiguous labelings in $\mathbf{K}_i^n$. This means that normalized relaxation implicitly assumes that we want to associate with each object only one label. A separate rejection-label has to be introduced to include the possibility that none of the labels is correct.

As with every iterative scheme the question of convergence and/or termination criteria is imminent. For normalized relaxation it has been shown that updating rules (6) and (4) will give convergence under fairly general conditions [7],[2],[6]. For optimization approaches it must be noted, that according to the problem at hand one does either seek global convergence to the maximum of the gain function or only local convergence to the nearest maximum. If the initial setting $\mathbf{p}^0$ is the most reliable input to the relaxation scheme and the aim is to improve $\mathbf{p}^0$ slightly by taking into account relational information $R_{ij\lambda\mu}$ the aim is to find a nearby maximum of $G(\mathbf{p})$. If on the other hand the problem specification relies mostly on relational knowledge expressed through $R_{ij\lambda\mu}$ and $\mathbf{p}^0$ might even be random one is interested in finding the global maximum. In this case deterministic relaxation schemes (as discussed in this paper) might get stuck at local maxima of $G(\mathbf{p})$.

V. Non-normalized Models

Until now, non-normalization of a relaxation scheme has been regarded as the essential definition of a fuzzy model and to the knowledge of the authors no normalized fuzzy models exist in the literature. Thus the few existing models described below are all fuzzy models.

Non-normalized models permit one object to carry multiple labels by allowing more than one assignment weight to become greater than 0.5. Two exemplary applications that may lead to such labeling problems are: The mixed pixel problem in image processing and stereo feature matching or tracking. One pixel may indeed belong to multiple classes and we may like to express this in a way different from having low confidence in both classes. Also stereo features in one frame may correspond to more than one feature in the other frame when lines or regions break up. Non-normalized labelings are well suited whenever it appears more naturally to stick to non exclusive and/or non exhaustive classes.
We begin with the first fuzzy model ever proposed ([1],[3],[10]):

$$p_{i\lambda}^{k+1} := \min_{j=1}^{n}\left[\max_{\mu=1}^{m_j} R_{ij\lambda\mu} \wedge p_{j\mu}^k\right] \tag{8}$$

where $\wedge$ means "and", that is e.g. min or product. It is understood that $R_{ij\lambda\mu} \in [0,1]$. It has been shown [1] that the iteration expressed by eq. (8) converges to the *greatest consistent fuzzy labeling*, with consistency being defined by

$$\max_{\mu=1}^{m_j} R_{ij\lambda\mu} \wedge p_{j\mu} \geq p_{i\lambda}. \tag{9}$$

and *greatest consistent labeling* standing for a labeling whose individual $p_{i\lambda}$ are each greater than the corresponding assignment weights of other consistent labelings.

There exists, however, one difficulty with the above algorithm: low weights are contagious. If for some object j we have $p_{j\mu} \leq 0.5$ for all μ then after application of (8) we have $p_{i\lambda} \leq 0.5$ for all i, λ.

Another suggestion [3] has been to use

$$p_{i\lambda}^{k+1} := \frac{1}{n} \sum_{j=1}^{n} \left[\max_{\mu=1}^{m_j} R_{ij\lambda\mu} \wedge p_{j\mu}^{k} \right]. \qquad (10)$$

This updating scheme is no longer suffering from the problem of contagious low weights. Both increases and decreases of $p_{i\lambda}$ are possible. However, it may appear unfavorable that the current weight $p_{i\lambda}^{k}$ gives only one term in an average when determining its new value.

A few other fuzzy relaxation schemes have been used [11],[12] but since they are quite specific solutions suited mainly for particular applications, we will not discuss them further. Recently a promising new candidate for a completely different approach to fuzzy relaxation based on the *Maximum Weight Bipartite Matching* problem has been presented [13].

VI. Considerations on Fuzzy Relaxation

Continuing the reasoning started in section II, it is to be investigated whether the broader viewpoint adopted there helps in finding new fuzzy models. Two initial observations are in order:

- It has been argued that fuzzy relaxation has to be defined through the use of fuzzy information and/or fuzzy fusion operations. Thus all the normalized models reviewed (except the probabilistic one) may as well be interpreted to be fuzzy models if they deal with fuzzy data.
- It has been observed that it is highly desirable for fuzzy approaches to allow for increases as well as decreases and to emphasize the importance of individual assignment weights.

We may think of two major classes of techniques that can be employed in a fuzzy relaxation framework:

- Analytical methods such as optimization of an appropriate gain function by iteration.
- Algebraic methods such as various fuzzy fusion operations or fuzzy rule bases.

A. Analytical Methods.

We shall restrict discussion to the constrained optimization framework which is definitely the setting to model relaxation in case a gain function can be defined, i.e. whenever $R_{ij\lambda\mu}$ is symmetric. Even with purely symmetric $R_{ij\lambda\mu}$ relaxation by optimization includes many useful application domains. Furthermore, the updating formulae obtained with symmetric $R_{ij\lambda\mu}$ may also give reasonable results in case no gain function can be defined [2].

It is desirable to choose functions $G(\mathbf{p})$ for which the gradient $\nabla G(\mathbf{p})$ can be evaluated locally for each object i by taking into account only some neighborhood of i. This is usually the case in equation (3) because only some neighboring objects j will contribute to $q_{i\lambda}$, i.e. have $R_{ij\lambda\mu} \neq 0$. If the locality property gets violated the gradient projection algorithm remains still valid but can no longer be parallelized. Note also that, if $G(\mathbf{p})$ should be non-differentiable still other optimization techniques can be used such as simulated annealing.

When trying to modify the gain function the meaning of the terms to be put into the gain function has to be clarified. Faugeras and Berthod [7] have argued in favor of two major terms, a consistency measure and a measure for ambiguity: $G(\mathbf{p}) = G_c(\mathbf{p}) - G_a(\mathbf{p})$. Maximizing $G(\mathbf{p})$ therefore amounts to increasing consistency $G_c(\mathbf{p})$ and decreasing ambiguity $G_a(\mathbf{p})$. We adhere to the common practice of identifying the consistency measure $G_c(\mathbf{p})$ with the terms in equation (5) and proceed to discuss various possible definitions of G_a in the following.

A.1 Normalized Models.

The theoretical basis for local and stepwise constrained optimization of normalized models has been investigated by Hummel and Zucker [2]. Their gradient projection algorithm has been summarized in section IV. The only question here is if introducing an extra measure of ambiguity $G_a(\mathbf{p})$ is interesting; and if so, if fuzzy set theory has something to contribute to its definition.

Today many different measures for various types of uncertainty including ambiguity are at our disposal [14]. Expressing the wanted effect of the measure for ambiguity $G_a(\mathbf{p})$ during relaxation in geometrical terms we search for "repulsive potentials" (i.e. penalty terms) in the vicinity of the center points $\mathbf{c}_i^n := \frac{1}{m_i}\mathbf{1}_i$, $\mathbf{1}_i = (1, .., 1)$ of the simplexes in $\mathbf{K}_i^n$. These points correspond to the most ambiguous configurations in $\mathbf{K}_i^n$. Note that an entropy term $G_a(\mathbf{p}) := -\sum_{i\lambda} p_{i\lambda} \ln(p_{i\lambda})$ would amount to such a repulsive potential, as would other entropy-like terms [14]. Another possibility is to introduce a term $G_a(\mathbf{p}) := -\sum_i \phi_i(d_{m_i}^2)$ with $d_{m_i}^2 := \sum_\lambda (p_{i\lambda} - \frac{1}{m_i})^2$ being the squared distance from the center point $\mathbf{c}$. Each ϕ_i has its negative local minimum at the distance 0 and increases towards zero for larger distances.

A.2 Non-normalized Models.

In the case of non-normalized fuzzy models the $\mathbf{p}^k$ are allowed to wander around in $\mathbf{K}^f$ through repeated application of eq. (6) with $\Psi(\mathbf{p}_i^k, \mathbf{q}_i^k)$ projecting onto $\mathbf{K}^f$ and $\mathbf{q} = \nabla G(\mathbf{p})$. $\mathbf{K}^f$ not only contains many more potential labeling situations than $\mathbf{K}^n$ but also for most points in $\mathbf{K}^f$ the distance to the closest unambiguous labeling $\mathbf{p}^c \in \mathbf{C}$ is bigger than the greatest possible distance of the points in $\mathbf{K}^n$ to a point in $\mathbf{C}$. Putting it informally: There are many more strongly ambiguous labelings in $\mathbf{K}^f$ than there are in $\mathbf{K}^n$. Therefore an additional term $G_a(\mathbf{p})$ in the gain function which explicitly tends to reduce ambiguity becomes much more important now.

Adopting once again a geometrical viewpoint we can easily state what $G_a(\mathbf{p})$ should accomplish: For points $\mathbf{p}$ close to one of the unambiguous corner points in $\mathbf{C}$ ambiguity $G_a(\mathbf{p})$ should be low. The further away from the corners $\mathbf{p}$ lies the higher $G_a(\mathbf{p})$ should become. A little experimenting shows that none of the common definitions reviewed in [14] captures these features satisfactorily for the task at hand. Thus we proceed to define a new measure of fuzzy ambiguity. G_a should amount to a landscape with basins in the vicinity of the unambiguous labelings

and elevations elsewhere. In a first approach one might set

$$G_a(\mathbf{p}) := -\sum_i \sum_{\mathbf{p}_i^c \in \mathbf{C}_i} \phi_i(\mathbf{p}_i - \mathbf{p}_i^c) \qquad (11)$$

with $\phi_i(\mathbf{x})$ being high for $\mathbf{x} = \mathbf{0}$ and decreasing towards 0 as $|\mathbf{x}|$ increases. The above definition shows two practical drawbacks: The double sum in the definition of $G_a(\mathbf{p})$ requires a lot of additional computations for each step of relaxation and each object i. Secondly the $\phi_i(\mathbf{x})$ are *not* functions of $|\mathbf{x}|$ but necessarily depend also on the direction of $\mathbf{x}$. This makes the single terms complicated and frequent evaluation of the gradient even more costly.

We therefore propose a different measure of ambiguity which is more quickly evaluated while keeping the most important properties of the above, intuitive definition. To this end we split the term $G_a(\mathbf{p}) := G_a^{(1)}(\mathbf{p}) + G_a^{(2)}(\mathbf{p})$ in two parts with separate characteristics. $G_a^{(1)}(\mathbf{p})$ is a measure based on the signed normal distances of the $\mathbf{p}_i$ to the simplexes $\mathbf{K}_i^n$, while $G_a^{(2)}(\mathbf{p})$ depends on the rescaled distances of the $\mathbf{p}_i$ to the diagonals $\mathbf{d}_i$ in $\mathbf{K}_i^f$: $\mathbf{d}_i(t) := t\,\mathbf{e}_i$ with $t \in [0, \sqrt{m_i}]$ and $\mathbf{e}_i := \frac{1}{\sqrt{m_i}}\mathbf{1}_i$.

$$G_a^{(1)}(\mathbf{p}) := \sum_{i=1}^{n} \psi_i^{(1)}\left((\mathbf{p}_i - \mathbf{c}_i^n)\cdot \mathbf{e}_i^T\right) \qquad (12)$$

$$G_a^{(2)}(\mathbf{p}) := \sum_{i=1}^{n} \psi_i^{(2)}\left(\frac{\mathbf{p}_i\cdot\mathbf{p}_i^T - (\mathbf{p}_i\cdot\mathbf{e}_i^T)^2}{\mathbf{1}_i\cdot\mathbf{p}_i^T - (\mathbf{p}_i\cdot\mathbf{e}_i^T)^2}\right). \qquad (13)$$

As can be verified with a little effort the argument of $\psi_i^{(1)}$ is indeed the signed normal distance of $\mathbf{p}_i$ to the simplex $\mathbf{K}_i^n$ while the argument of $\psi_i^{(2)}$ is the square of the distance of $\mathbf{p}_i$ to the diagonal $\mathbf{d}_i(t)$ rescaled by the numerator in such a manner that all corner points of $\mathbf{K}_i^f$ are at the same "unit" distance.

The functions $\psi_i^{(1)}(x)$ are zero for $x = 0$ and increase as $|x|$ increases such that the final effect of $G_a^{(1)}$ is to obtain a "valley"-potential at the locations of $\mathbf{K}_i^n$. Thus $G_a^{(1)}$ will tend to drive $\mathbf{p}^k$ to the vicinity of $\mathbf{K}^n$. As the labelings of $\mathbf{K}^n$ are on the average less ambiguous than most labelings far from $\mathbf{K}^n$ this will already introduce a bias for unambiguous labelings. Regarding $G_a^{(1)}$ to be an approximation to equation (11) we see that the major errors introduced by reducing eq. (11) to equation (12) are to be expected for $\mathbf{p}_i$ near to the strongly ambiguous line segments $\mathbf{d}_i(t)$.

The purpose of $G_a^{(2)}$ is exactly to reduce errors in that region. The functions $\psi_i^{(2)}(x)$ are high for $x = 0$ and decrease towards 0 as x increases until $\psi_i^{(2)}(1) = 0$. This way $G_a^{(2)}$ penalizes labelings near to $\mathbf{d}(t)$ by assigning a high measure of ambiguity to them. The combined effect of $G_a^{(1)}$ and $G_a^{(2)}$ is to keep ambiguity values low especially near to $\mathbf{K}^n$ and there the lowest values are obtained far from $\mathbf{d}$, i.e. at the corners $\mathbf{C}$. It is therefore similar to the effect of eq. (11) but note that the double sum is no longer present in the new definitions. This has become possible because using eq. (12) and (13) we no longer explicitly enumerate potentials ϕ_i having high values only in the domain $\mathbf{D}$ around $\mathbf{C}$ where G_a should be low. Instead the new equations enumerate potentials $\psi_i^{(1)}$ and $\psi_i^{(2)}$ which assume high values in the complement $\mathbf{K}^f \setminus \mathbf{D}$, a region which can be represented much more compactly than $\mathbf{D}$.

It should also be noted that for many applications it will be possible to define how much ambiguity should be tolerated without penalizing. The width of the functions $\psi_i^{(1)}(x)$ can then be chosen accordingly. For example in the case of stereo matching of lines, the maximum number w of pieces into which a line may have been broken up can be determined by searching for collinear, nearby lines. The ambiguity measuring potential $G_a(\mathbf{p})$ can then be chosen to become very high for labelings with more than w high assignment weights.

For non-normalized fuzzy models the quantities $R_{ii\lambda\mu}$ may also be of special importance in the process of reducing ambiguity. Quite often specific $R_{ii\lambda\mu}$ have to be set to negative values near to -1 and therefore the compatibility term (5) will already help in eliminating ambiguous configurations because it will tend to keep either $p_{i\lambda}$ or $p_{i\mu}$ or both low. Again an example for the application to structural matching of stereo images of lines can be given. To set $R_{ii\lambda\mu}$ to a low value indicates that one does not believe that line i in the left image has broken up into the two lines λ and μ in the right image. This may be the case because λ and μ are not collinear, too far apart, etc. Note that it is not even possible to model this case with normalized relaxation schemes [5] where only the merging of lines can be handled if multiple objects i,j are allowed to get the same final label λ.

The introduction of non-normalized fuzzy relaxation models based on optimization of a function $G(\mathbf{p})$ also solves the problems stated in connection with the existing fuzzy models: Optimization frameworks result in update rules of the same form as equation (6). Thus the current assignment weight $p_{i\lambda}^k$ plays a special role in its updating process and both *increases and decreases* are possible.

B. Algebraic Methods

Again we will restrict attention to a particular modeling tool: fuzzy rule bases as they are used for fuzzy control tasks. Figure 2 can immediately be interpreted in terms of qualitative control rules to calculate the change $\Delta p_{i\lambda}^k$ of the assignment weight $p_{i\lambda}^k$.($n = card(O)$, $m = card(\Lambda)$).

- IF $p_{j_1\mu_1}^k$ IS *high* AND $R_{ij\lambda\mu_1}$ IS *high*
  THEN $\Delta p_{i\lambda}^k$ IS *positive.*
- IF $p_{j_1\mu_1}^k$ IS *low* AND $R_{ij\lambda\mu_1}$ IS *high*
  THEN $\Delta p_{i\lambda}^k$ IS *negative.*
  ⋮ ⋮
- IF $p_{j_n\mu_m}^k$ IS *high* AND $R_{ij_n\lambda\mu_m}$ IS *high*
  THEN $\Delta p_{i\lambda}^k$ IS *positive.*
- IF $p_{j_n\mu_m}^k$ IS *low* AND $R_{ij_n\lambda\mu_m}$ IS *high*
  THEN $\Delta p_{i\lambda}^k$ IS *negative.*

Rule base 1: Calculation of $\Delta p_{i\lambda}^k$

We have neglected the rules leading to no change for brevity. Various realizations of the above rule base can be conceived in the domain of fuzzy control. Since figure 2 describes the local behavior during one relaxation step the fuzzy rule base can be evaluated locally for each object i. The resulting $\Delta p_{i\lambda}^k$ have to be fused with $p_{i\lambda}^k$ to yield the next assignment weights

$$\mathbf{p}_i^{k+1} := \mathbf{f}(\mathbf{p}_i^k, \Delta\mathbf{p}_i^k).$$

The fusion functions f_λ must lead to values $p_{i\lambda}^{k+1}$ in $[0,1]$ which are increasing with both $p_{i\lambda}^k$ and positive $\Delta p_{i\lambda}^k$ and decreasing with negative $\Delta p_{i\lambda}^k$. This fusion operator $\mathbf{f}$ plays a crucial task in the updating process and must be devised very carefully in order not to change the updating behavior that is indicated by $\Delta\mathbf{p}_i^k$. For example, whenever optimization schemes are applicable, through the gradient we know the "best" local updating vector $\Psi(\mathbf{p}_i^k, \mathbf{q}_i^k)$ lying in feasible space and pointing towards increases of $G(\mathbf{p})$. If the fusion operator $\mathbf{f}$ is not designed carefully it may happen that even with the "best" updating vector $\mathbf{q}_i^k = \nabla_i G(\mathbf{p}^k)$ put into $\mathbf{f}$, the fusion operator $\mathbf{f}$ produces a step in the wrong direction. A possible, cautious way to proceed would be to apply the same updating rule as was used for gradient back-projection

$$\mathbf{p}_i^{k+1} := \mathbf{p}_i^k + \alpha\Psi(\mathbf{p}_i^k, \Delta\mathbf{p}_i^k)$$

now back-projecting $\Delta\mathbf{p}_i^k$ calculated with the rule base. This updating puts all the burden to the calculation of $\Delta\mathbf{p}_i^k$ and ensures afterwards that the result is not altered except to achieve a point in feasible space $\mathbf{K}^n$ or $\mathbf{K}^f$. In the following we separate again the discussion of the two cases of normalized and non-normalized relaxation models.

Normalized fuzzy models have to compete with the results obtained by Hummel and Zucker [2] for general, normalized updating schemes. There are two possible generalizations conceivable: On the one hand it may be that the consistency criterion which is fundamental for the work in [2] could be replaced or enhanced by other terms. On the other hand it may be that for certain applications only linguistic descriptions of support or resistance can be found. Both situations may still lead to a solution in terms of fuzzy rule bases.

For non-normalized fuzzy models it has to be observed that the rules in figure 2 and Rule-base 1 are necessary conditions for a relaxation scheme but they will usually not be sufficient. The problem being, that these rules describe how to achieve consistency but do not describe how to reduce ambiguity. In the normalized case this is done partially by the normalization constraint as has already been explained. For non-normalized models the rule base approach has to include rules on how to deal with conflicting evidence. Since objects can have high assignment weights for different labels it might be that the same object j is through the labeling (j, μ_1) and $R_{i\lambda j\mu_1}$ indicating strongly to decrease $p_{i\lambda}^k$, while the labeling (j, μ_2) together with $R_{i\lambda j\mu_2}$ points to an increase. This problem arises here in full severity because we have no normalization constraint and we have to hand-craft all fusion operators and rule bases. For particular situations solutions have already been found [15].

VII. Summary, Conclusions and Outlook

We have re-examined the current classification of relaxation labeling schemes and have found it inappropriate to equate fuzzy models with the category non-normalized relaxation schemes. This is the first time an approach has been undertaken to define and defend normalized fuzzy relaxation models. The presented ideas establish a new framework for fuzzy relaxation approaches be they normalized or non-normalized.

For both types of fuzzy relaxation models two possible implementations have been discussed: optimization based and rule based. Insights and concrete guidelines for new approaches have been presented. The optimization approach leads to theoretically well justified fuzzy relaxation schemes with new, geometrically motivated definitions of measures of ambiguity. The rule based approach offers the possibility to extend the domain of relaxation to problems defined in purely linguistic terms. Explicit implementations and experiments with rule based relaxation on the task of matching curves and lines can be found in [15]. Ongoing work is undertaken to apply the presented ideas to various labeling problems.

We gratefully acknowledge support by : FWF S7003, Task 3-1, BMWV (Gz. 601.574/2-IV/B/9/96) and ÖAD, Amadeus, Proj. 19.

References

[1] A.Rosenfeld, R.A.Hummel, and S.W.Zucker, "Scene labeling by relaxation operations," *IEEE Transactions on Systems Man and Cybernetics*, vol. 6, no. 6, pp. 420–433, 1976.

[2] R.A.Hummel and S.W.Zucker, "On the foundations of relaxation labeling processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, no. 3, pp. 267–287, May 1983.

[3] A.Rosenfeld and A.C.Kak, *Digital Picture Processing*, vol. 2, Academic Press,Inc., San Diego, California 92101, 2nd edition, 1982.

[4] J.Kittler and J.Illingworth, "Relaxation labeling algorithms - A review," *Image and Vision Computing*, vol. 3, no. 4, pp. 206–216, 1985.

[5] S.Z.Li, "Matching: Invariant to translations, rotations and scale changes," *Pattern Recognition*, vol. 25, no. 6, pp. 583–594, 1992.

[6] M.Pelillo, "The Dynamics of Nonlinear Relaxation Labeling Processes," *Journal of Mathematical Imaging and Vision*, , no. 7, pp. 309–323, 1997.

[7] O.D.Faugeras and M.Berthod, "Improving consistency and reducing ambiguity in stochastic labeling: An optimization approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 3, no. 4, pp. 412–424, 1981.

[8] R.M. Haralick, "Decision making in context," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, no. 4, pp. 417–428, 1983.

[9] W.J. Christmas, J.Kittler, and M.Petrou, "Structural Matching in Computer Vision Using Probabilistic Relaxation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 749–764, 1995.

[10] H. Ogawa, "Labeled Point Pattern Matching by Fuzzy Relaxation," *Pattern Recognition*, vol. 17, no. 5, pp. 569–573, 1984.

[11] H.Ogawa, "A fuzzy relaxation technique for partial shape matching," *Pattern Recognition Letters*, , no. 15, pp. 349–355, 1994.

[12] H.S.Ranganath and L.J.Chipman, "Fuzzy relaxation approach for inexact scene matching," *Image and Vision Computing*, vol. 10, no. 9, pp. 631–640, Nov. 1992.

[13] Z.Shao and J.Kittler, "Fuzzy Non-Iterative ARG Labeling with Multiple Interpretations," in *Proceedings of ICPR*, 1996.

[14] N.R.Pal and J.C.Bezdek, "Measuring Fuzzy Uncertainty," *IEEE Transactions on Fuzzy Systems*, vol. 2, no. 2, pp. 107–118, May 1994.

[15] Hermann Borotschnig, David Sinclair, and Axel Pinz, "Fuzzy Graph Tracking," in *Proc. 5th International Symposium on Intelligent Robotic Systems*, H.I.Christensen and C.Bräutigam and C.Ridderström, Ed., Centre for Autonomous Systems, KTH Stockholm, 1997, pp. 91–101, SIRS97 Organizing Comittee.