Continuous-time Relaxation Labeling Processes *

Andrea Torsello **Marcello Pelillo** †

Dipartimento di Matematica Applicata e Informatica

Università Ca' Foscari di Venezia

Via Torino 155, 30173 Venezia Mestre, Italy

{atorsell,pelillo}@dsi.unive.it

---

*Submitted to *Pattern Recognition*

†Author to whom correspondence should be addressed

Abstract

We study the properties of two new relaxation labeling schemes described in terms of differential equations, and hence evolving in countinuous time. This contrasts with the customary approach to defining relaxation labeling algorithms which prefers discrete time. Continuous-time dynamical systems are particularly attractive because they can be implemented directly in hardware circuitry, and the study of their dynamical properties is simpler and more elegant. They are also more plausible as models of biological visual computation. We prove that the proposed models enjoy exactly the same dynamical properties as the classical relaxation labeling schemes, and show how they are intimately related to Hummel and Zucker's now classical theory of constraint satisfaction. In particular, we prove that, when a certain symmetry condition is met, the dynamical systems' behavior is governed by a Liapunov function which turns out to be (the negative of) a well-known consistency measure. Moreover, we prove that the fundamental dynamical properties of the systems are retained when the symmetry restriction is relaxed. We also analyze the properties of a simple discretization of the proposed models, which is useful in digital computer implementations. Simulation results are presented which confirm the effectiveness of the proposed models.

Keywords: Relaxation labeling processes, differential equations, dynamical systems, consistency.

1 Introduction

Relaxation labeling processes are a popular class of parallel, distributed computational models aimed at solving (continuous) constraint satisfaction problems, instances of which arise in a wide variety of computer vision and pattern recognition tasks [1, 9]. Almost invariably, all the relaxation algorithms developed so far evolve in discrete time, i.e., they are modeled as difference rather than as differential equations. The main reason for this widespread practice is that discrete-time dynamical systems are simpler to program and simulate on digital computers. However, continuous-time dynamical systems are more attractive for several reasons. First, they can more easily be implemented in parallel, analog circuitry (see, e.g., [4]). Second, the study of their dynamical properties is simplified thanks to the power of differential calculus, and proofs are more elegant and more easily understood. Finally, from a speculative standpoint, they are more plausible as models of biological computation [7].

Recently, there has been some interest in developing relaxation labeling schemes evolving in continuous time. In particular, we cite the work by Stoddart [16] motivated by the Baum-Eagon inequality [12], and the recent work by Li *et al.* [10] who developed a new relaxation scheme based on augmented Lagrangian multipliers and Hopfield networks. Yu and Tsai [18] also used a continuous-time Hopfield network for solving labeling problems. All these studies, however, are motivated by the assumption that the labeling problem is formulated as an energy-minimization problem, and a connection to standard theories of consistency [8] exists only when the compatibility coefficients are assumed to be symmetric. This is well-known to be a restrictive and unrealistic assumption. When the symmetry condition is relaxed the labeling problem is equivalent to a variational inequality problem, which is indeed a generalization of standard optimization problems [8].

In this paper, we study the dynamical properties of two simple relaxation labeling schemes which evolve in continuous time, each being described in terms of a system of coupled differential equations. The systems have been introduced in the context of evolutionary game theory, to model the evolution of relative frequencies of species in a multi-population setting [17], and one of them has also recently been proposed by Stoddart *et al.* [16], who studied its properties only in a very restricted case (i.e., symmetric compatibility coefficients). Both schemes are considerably simpler that Hummel and Zucker's continuous-time model [8] which requires a complicated projection operator. Moreover, the first scheme has no normalization phase, and this makes it particularly attractive for practical hardware implementations. Since our models automatically satisfy the constraints imposed by the structure of the labeling problem, they are also much simpler than Yu and Tsai's [18] and Li *et al.*'s [10] schemes, which have to take constraints into account either in the form of penalty functions or Lagrange multipliers.

The principal objective of this study is to analyze the dynamics of these relaxation schemes and to relate them to the classical theory of consistency developed by Hummel and Zucker [8]. We show that all the dynamical properties enjoyed by standard relaxation labeling algorithms do hold for ours. In particular, we prove that, when symmetric compatibility coefficients are employed, the models have a Liapunov function which rules their dynamical behavior, and this turns out to be (the negative of) a well-known consistency measure. Moreover, and most importantly, we prove that the fundamental dynamical properties of the systems are retained when the symmetry restriction is relaxed. We also study the properties of a simple discretization of the proposed models, which is useful in digital computer implementations. Some simulation results are presented which show

how the models behave in practice and confirm their validity.

The outline of the paper is as follows. In Section 2, we briefly review Hummel and Zucker's consistency theory, which is instrumental for the subsequent development. In Section 3 we introduce the models and in Section 4 we present the main theoretical results, first for the symmetric and then for the non-symmetric case. Section 5 describes two ways of discretizing the models, and proves some results. In Section 6 we present our simulation results, and Section 7 concludes the paper.

2 Consistency and its properties

The labeling problem involves a set of objects $B = \{b_1, \cdots, b_n\}$ and a set of possible labels $\Lambda = \{1, \cdots, m\}$. The purpose is to label each object of B with one label of Λ. To accomplish this, two sources of information are exploited. The first one relies on *local* measurements which capture the salient features of each object viewed in isolation; classical pattern recognition techniques can be practically employed to carry out this task. The second source of information, instead, accounts for possible interactions among nearby labels and, in fact, incorporates all the contextual knowledge about the problem at hand. This is quantitatively expressed by means of a real-valued four-dimensional matrix of compatibility coefficients $R = \{r_{ij}(\lambda, \mu)\}$. The coefficient $r_{ij}(\lambda, \mu)$ measures the strength of compatibility between the hypotheses "b_i has label λ" and "b_j has label μ:" high values correspond to compatibility and low values correspond to incompatibility. In our discussion, the compatibilities are assumed to be nonnegative, i.e., $r_{ij}(\lambda, \mu) \geq 0$, but this seems not to be a severe limitation because all the interesting concepts involved here exhibit some sort of "linear invariance" property [12]. In this paper, moreover, we will not be concerned with the crucial problem of how to derive the compatibility coefficients. Suffice it to say that they can be either determined on the basis of statistical grounds [11, 15] or, according to a more recent standpoint, adaptively learned over a sample of training data [14, 13].

The initial local measurements are assumed to provide, for each object $b_i \in B$, an m-dimensional vector $\bar{p}_i^0 = (p_i^0(1), \cdots, p_i^0(m))^T$ (where "T" denotes the usual transpose operation), such that $p_i^0(\lambda) \geq 0$, $i = 1 \ldots n$, $\lambda \in \Lambda$, and $\sum_\lambda p_i^0(\lambda) = 1$, $i = 1 \ldots n$. Each $p_i^0(\lambda)$ can be regarded as the initial, non-contextual degree of confidence of the hypothesis "b_i is labeled with label λ." By simply concatenating $\bar{p}_1^0, \bar{p}_2^0, \cdots, \bar{p}_n^0$ we obtain a weighted labeling assignment for the objects of B that will be denoted by $\bar{p}^0 \in \mathbb{R}^{nm}$. A relaxation labeling process takes as input the initial labeling assignment $\bar{p}^0$ and iteratively updates it taking into account the compatibility model R.

At this point, we introduce the space of weighted labeling assignments:

$$\mathbb{K} = \left\{ \bar{p} \in \mathbb{R}^{nm} \;\middle|\; p_i(\lambda) \geq 0,\ i = 1 \ldots n,\ \lambda \in \Lambda \ \text{ and } \ \sum_{\lambda=1}^{m} p_i(\lambda) = 1,\ i = 1 \ldots n \right\}$$

which is a linear convex set of $\mathbb{R}^{nm}$. Every vertex of $\mathbb{K}$ represents an *unambiguous* labeling assignment, that is one which assigns exactly one label to each object. The set of these labelings will be denoted by $\mathbb{K}^*$:

$$\mathbb{K}^* = \left\{ \bar{p} \in \mathbb{K} \mid p_i(\lambda) = 0 \text{ or } 1,\ i = 1 \ldots n,\ \lambda \in \Lambda \right\} .$$

Moreover, a labeling $\bar{p}$ in the interior of $\mathbb{K}$ (i.e., $0 < p_i(\lambda) < 1$, for all i and λ) will be called *strictly ambiguous*.

Now, let $\bar{p} \in \mathbb{K}$ be any labeling assignment. To develop a relaxation algorithm that updates $\bar{p}$ in accordance with the compatibility model, we need to define, for each object $b_i \in B$ and each label $\lambda \in \Lambda$, what is called a *support* function. This should quantify the degree of agreement between the hypothesis that b_i is labeled with λ, whose confidence is expressed by $p_i(\lambda)$, and the context. This measure is commonly defined as follows:

$$q_i(\lambda; \bar{p}) = \sum_{j=1}^{n} \sum_{\mu=1}^{m} r_{ij}(\lambda, \mu) p_j(\mu) \ . \tag{1}$$

Putting together the instaces $q_i(\lambda; \bar{p})$, for all the $p_i(\lambda)$, we obtain an nm-dimensional support vector that will be denoted by $\bar{q}(\bar{p})$.[1]

The following updating rule

$$p_i^{t+1}(\lambda) = \frac{p_i^t(\lambda) q_i^t(\lambda)}{\sum_\mu p_i^t(\mu) q_i^t(\mu)} \tag{2}$$

where $t = 0, 1, \ldots$ denotes (discrete) time, defines the original relaxation labeling operator of Rosenfeld, Hummel, and Zucker [15], whose dynamical properties have recently been clarified [12]. In the following discussion we shall refer to it as the "classical" relaxation scheme.

We now briefly review Hummel and Zucker's theory of constraint satisfaction [8] which commences by providing a general definition of consistency. By analogy with the unambiguous case, which is more easily understood, a weighted labeling assignment $\bar{p} \in \mathbb{K}$ is said to be *consistent* if

$$\sum_{\lambda=1}^{m} p_i(\lambda) q_i(\lambda; \bar{p}) \geq \sum_{\lambda=1}^{m} v_i(\lambda) q_i(\lambda; \bar{p}) \ , \quad i = 1 \ldots n \tag{3}$$

[1]Henceforth, when it will be clear from context, the dependence on $\bar{p}$ will not be stated.

for all $\bar{v} \in \mathbb{K}$. Furthermore, if strict inequalities hold in (3), for all $\bar{v} \neq \bar{p}$, then $\bar{p}$ is said to be *strictly consistent*. It can be seen that a necessary condition for $\bar{p}$ to be strictly consistent is that it is an unambiguous one, that is $\bar{p} \in \mathbb{K}^*$.

In [8], Hummel and Zucker introduced the *average local consistency*, defined as

$$A(\bar{p}) = \sum_{i=1}^{n} \sum_{\lambda=1}^{m} p_i(\lambda) q_i(\lambda) \tag{4}$$

and proved that when the compatibility matrix R is symmetric, i.e., $r_{ij}(\lambda, \mu) = r_{ji}(\mu, \lambda)$ for all i, j, λ, μ, then any local maximum $\bar{p} \in \mathbb{K}$ of A is consistent. Basically, this follows immediately from the fact that, when R is symmetric, we have $\nabla A(\bar{p}) = 2\bar{q}$, $\nabla A(\bar{p})$ being the gradient of A at $\bar{p}$. Note that, in general, the converse need not be true since, to prove this, second-order derivative information would be required. However, by demanding that $\bar{p}$ be strictly consistent, this does happen [12].

3 Continuous-time relaxation labeling processes

The two relaxation labeling models studied in this paper are defined by the following systems of coupled differential equations:

$$\frac{d}{dt} p_i(\lambda) = p_i(\lambda) \left(q_i(\lambda) - \sum_{\mu} p_i(\mu) q_i(\mu) \right) \tag{5}$$

and

$$\frac{d}{dt} p_i(\lambda) = p_i(\lambda) \frac{q_i(\lambda) - \sum_{\mu} p_i(\mu) q_i(\mu)}{\sum_{\mu} p_i(\mu) q_i(\mu)} \ . \tag{6}$$

For the purpose of the present discussion, $q_i(\lambda)$ denotes the linear support as defined in equation (1). As a matter of fact, many of the results proved below do not depend on this particular choice. More generally, the only requirements are that the support function be nonnegative and, to be able to grant the existence and unicity of the solution of the differential equations, that it be of class C^1 [6].

In the first model we note that, although there is no explicit normalization process in the updating rule, the assignment space $\mathbb{K}$ is invariant under dynamics (5). This means that any trajectory starting in $\mathbb{K}$ will remain in $\mathbb{K}$. To see this, simply note that:

$$\sum_{\lambda} \frac{d}{dt} p_i(\lambda) = \sum_{\lambda} p_i(\lambda) \left(q_i(\lambda) - \sum_{\mu} p_i(\mu) q_i(\mu) \right) = 0$$

which means that the interior of $\mathbb{K}$ is invariant. The additional observation that the boundary too is invariant completes the proof. The same result can be proven for the other model as well,

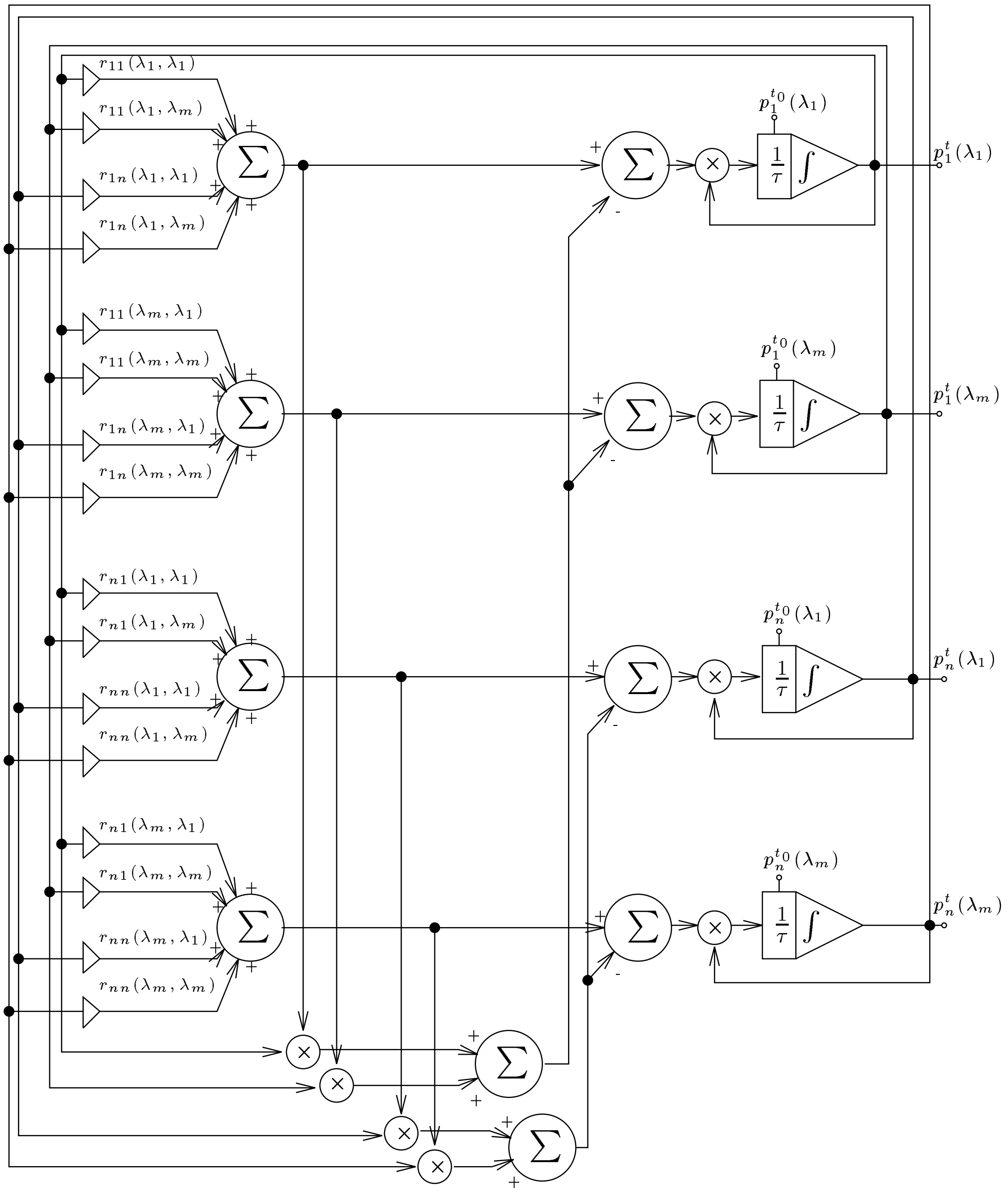

Figure 1: Scheme of a circuit implementation for the standard model (5)—see, e.g., [4] for a description of the symbols adopted.

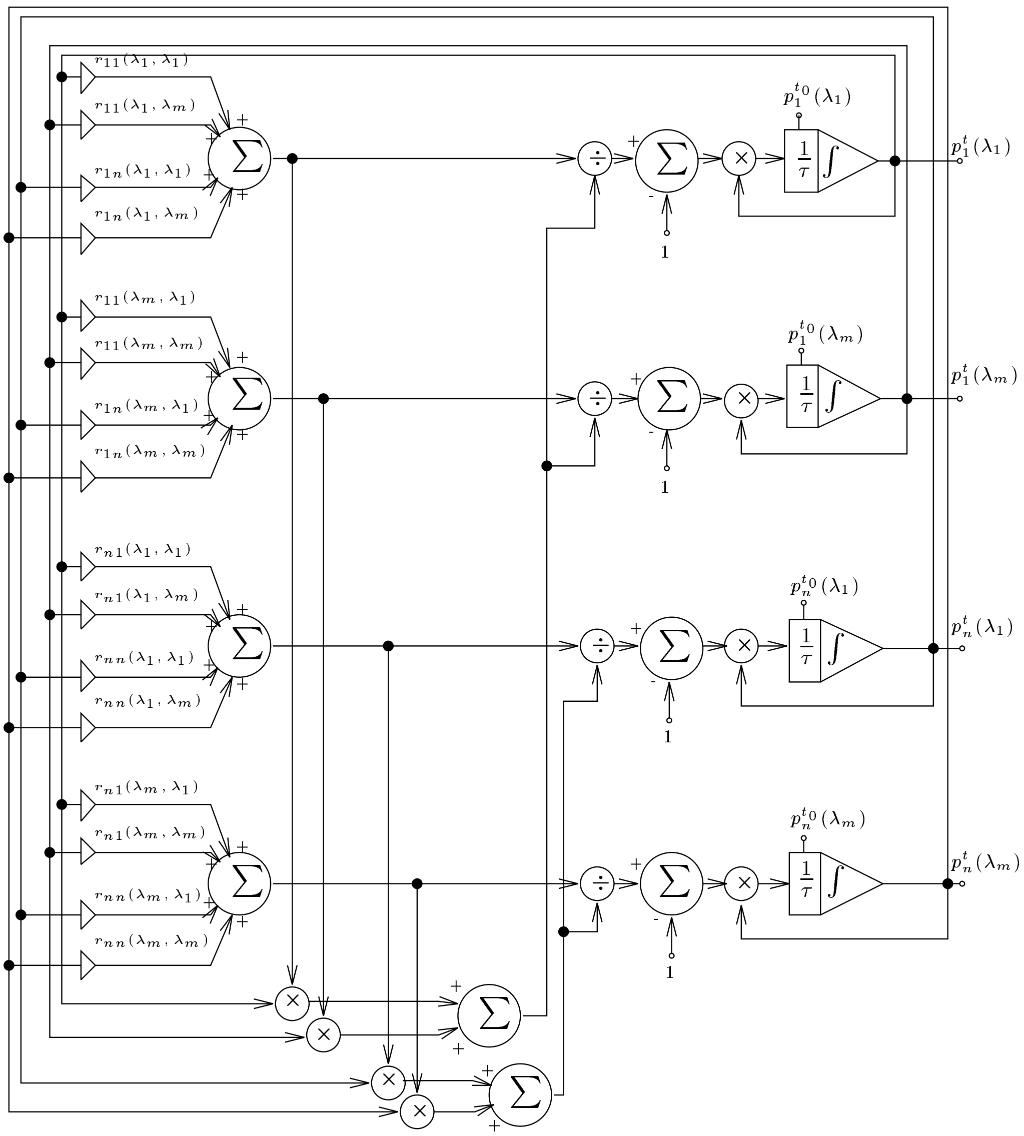

Figure 2: Scheme of a circuit implementation for the normalized model (6)—see, e.g., [4] for a description of the symbols adopted.

following basically the same steps. The lack of normalization makes the first model, which we call the *standard* model, more attractive than Hummel and Zucker's projection-based scheme [8], since it makes it more amenable to hardware implementations and more acceptable biologically. The interest in the other model, called the *normalized* model and also studied by Stoddart [16], derives from the fact that, in a way, it is the continuous-time translation of the classical Rosenfeld-Hummel-Zucker relaxation scheme [15]. This will be clearer when we show the discretizations of the models. We note that, using a linear support function (1), the dynamics of the models is invariant under a rescaling of the compatibility coefficients $r_{ij}(\lambda,\mu)$. That is, if we define a set of new compatibility coefficients $r^*_{ij}(\lambda,\mu) = \alpha r_{ij}(\lambda,\mu) + \beta$, with $\alpha > 0$ and $\beta \geq 0$, the orbit followed by the model remains invariable, while the speed at which the dynamics evolve changes by a factor α.

As stated in the Introduction, one of the most attractive features of continuous-time systems is that they are readily mapped onto hardware circuitry. Figures 1 and 2 show a circuit implementation for the standard and the normalized models, respectively. As clearly expected, the standard model leads to a more economic implementation.

The fixed (or equilibrium) points of our dynamical systems are characterized by $\frac{d}{dt}\bar{p} = 0$ or, more explicitly, by $p_i(\lambda)\left[q_i(\lambda) - \sum_\mu p_i(\mu)q_i(\mu)\right] = 0$ for all $i = 1\ldots n$, $\lambda \in \Lambda$. This leads us to the condition

$$p_i(\lambda) > 0 \Rightarrow q_i(\lambda) = \sum_\mu p_i(\mu)q_i(\mu) \tag{7}$$

which is the same condition we have for the Rosenfeld-Hummel-Zucker and Hummel-Zucker models.

The next result follows immediately from a characterization of consistent labelings proved in [12, Theorem 3.1].

Proposition 3.1 *Let $\bar{p} \in \mathbb{K}$ be consistent. Then $\bar{p}$ is an equilibrium point for the relaxation dynamics (5) and (6). Moreover, if $\bar{p}$ is strictly ambiguous the converse also holds.*

This establishes a first connection between our continuous-time relaxation labeling processes and Hummel and Zucker's theory of consistency.

4 The dynamical properties of the models

In this section we study the dynamical properties of the proposed dynamical systems. Specifically, we show how our continuous-time relaxation schemes are intimately related to Hummel and Zucker's theory of consistency, and enjoy all the dynamical properties which hold for the classical discrete-time scheme (2), and Hummel and Zucker's projection-based model.

Before going into the technical details, we briefly review some instrumental concepts in dynamical systems theory; see [6] for details. Given a dynamical system, an equilibrium point $\bar{x}$ is said to be *stable* if, whenever started sufficiently close to $\bar{x}$, the system will remain near to $\bar{x}$ for all future times. A stronger property, which is even more desirable, is that the equilibrium point $\bar{x}$ be *asymptotically stable*, meaning that $\bar{x}$ is stable and in addition is a *local attractor*, i.e., when initiated close to $\bar{x}$, the system tends towards $\bar{x}$ as time increases. One of the most fundamental tools for establishing the stability of a given equilibrium point is known as the Liapunov's direct method. It involves seeking a so-called *Liapunov* function, i.e., a continuous real-valued function defined in state space which is nonincreasing along a trajectory.

4.1 Symmetric compatibilities

We present here some results which hold when the compatibility matrix R is symmetric, i.e., $r_{ij}(\lambda,\mu) = r_{ji}(\mu,\lambda)$, for all $i,j = 1\dots n$ and $\lambda,\mu \in \Lambda$.

Lemma 4.1 *If the compatibility matrix R is symmetric, then for all $\bar{p} \in \mathbb{K}$ we have*

$$\bar{q}(\bar{p}) \cdot \frac{d}{dt}\bar{p} \geq 0 \ ,$$

where "$\cdot$" represents the inner product operator, for both the standard and normalized relaxation schemes (5) and (6).

Proof: Let $\bar{p}$ be an arbitrary labeling assignment in $\mathbb{K}$. For the standard model we have:

$$\begin{aligned}\bar{q}(\bar{p}) \cdot \frac{d}{dt}\bar{p} &= \sum_{i,\lambda} q_i(\lambda)p_i(\lambda)\left(q_i(\lambda) - \sum_{\mu} p_i(\mu)q_i(\mu)\right) \\ &= \sum_i \left[\sum_\lambda p_i(\lambda)q_i^2(\lambda) - \left(\sum_\lambda p_i(\lambda)q_i(\lambda)\right)^2\right]\end{aligned}$$

Using the Cauchy-Schwartz inequality we obtain, for all $i = 1\dots n$,

$$\begin{aligned}\left(\sum_\lambda p_i(\lambda)q_i(\lambda)\right)^2 &= \left(\sum_\lambda \sqrt{p_i(\lambda)} \cdot \sqrt{p_i(\lambda)q_i^2(\lambda)}\right)^2 \\ &\leq \sum_\lambda p_i(\lambda) \cdot \sum_\lambda p_i(\lambda)q_i^2(\lambda) = \sum_\lambda p_i(\lambda)q_i^2(\lambda)\end{aligned}$$

Hence, since $\sum_\lambda p_i(\lambda)q_i^2(\lambda) \geq \left(\sum_\lambda p_i(\lambda)q_i(\lambda)\right)^2$, we have $\bar{q}(\bar{p}) \cdot \frac{d}{dt}\bar{p} \geq 0$.

The proof for the normalized model is the same as the above; we should just remember that:

$$\bar{q}(\bar{p}) \cdot \frac{d}{dt}\bar{p} = \sum_i \frac{\sum_\lambda q_i(\lambda)p_i(\lambda)\left(q_i(\lambda) - \sum_\mu p_i(\mu)q_i(\mu)\right)}{\sum_\mu p_i(\mu)q_i(\mu)}$$

□

A straightforward consequence of the previous lemma is the following important result, which states that, in the symmetric case, the average local consistency is always increasing along the trajectories of our dynamical systems.

Theorem 4.2 *If the compatibility matrix R is symmetric, we have*

$$\frac{d}{dt}A(\bar{p}) \geq 0$$

for all $\bar{p} \in \mathbb{K}$. In other words, $-A$ is a Liapunov function for the relaxation models (5) and (6).

Proof: Assuming $r_{ij}(\lambda, \mu) = r_{ji}(\mu, \lambda)$, we have:

$$\begin{aligned}\frac{d}{dt}A(\bar{p}) &= \sum_{i\lambda}\sum_{j\mu} r_{ij}(\lambda,\mu)p_j(\mu)\frac{d}{dt}p_i(\lambda) \\ &= \bar{q}(\bar{p}) \cdot \frac{d}{dt}\bar{p} \geq 0\end{aligned}$$

□

As far as the normalized scheme is concerned, this result has been proven by Stoddart [16]. By combining the previous result with the fact that strictly consistent labelings are local maxima of the average local consistency (see [12, Proposition, 3.4]) we readily obtain the following proposition.

Theorem 4.3 *Let $\bar{p}$ be a strictly consistent labeling and suppose that the compatibility matrix R is symmetric. Then $\bar{p}$ is an asymptotically stable stationary point for the relaxation labeling processes (5) and (6) and, consequently, is a local attractor.*

Therefore, in the symmetric case our continuous-time processes have exactly the same dynamical properties as the classical Rosenfeld-Hummel-Zucker model [12] and the Hummel-Zucker projection-based scheme [8].

4.2 Arbitrary compatibilities

In the preceding subsection we have restricted ourselves to the case of symmetric compatibility coefficients and have shown how, under this circumstance, the proposed continuous-time relaxation schemes are closely related to the theory of consistency of Hummel and Zucker. However, although symmetric compatibilities can easily be derived and asymmetric matrices can always be made symmetrical (i.e., by considering $R + R^T$), it would be desirable for a relaxation process to work also when no restriction on the compatibility matrix is imposed [8]. This is especially true when the relaxation algorithm is viewed as a plausible model of how biological systems perform visual computation [19].

Despite the fact that in the asymmetric case no functional to be maximized can be found, we now show that the proposed relaxation dynamical systems still perform useful computations in this case, and their connection with the theory of consistency continues to hold.

The main result is the following:

Theorem 4.4 *Let $\bar{p} \in \mathbb{K}$ be a strictly consistent labeling. Then $\bar{p}$ is an asymptotically stable equilibrium point for the continuous-time relaxation labeling schemes defined in equations (5) and (6).*

Proof: To prove the theorem we have to show that a strictly consistent labeling $\bar{p}$ is a stable equilibrium point, and also a local attractor for the systems. We shall only prove the latter; the stability part of the statement can easily be proved by adapting the arguments of the proof of Theorem 6.4 in [12], and therefore we omit the details.

We begin by recalling that a strictly consistent assignment is necessarily non-ambiguous. That is:

$$p_i(\lambda) = \begin{cases} 0 & \text{if } \lambda \neq \lambda(i) \\ 1 & \text{if } \lambda = \lambda(i) \end{cases} = \delta_{\lambda\lambda(i)}$$

where δ is the Kronecker delta, i.e., $\delta_{xy} = 1$ if $x = y$, and $\delta_{xy} = 0$ otherwise, and $\lambda(i)$ is the unique consistent assignment for object i. Furthermore we have $q_i\big(\lambda(i)\big) > q_i(\lambda)$ for all $\lambda \neq \lambda(i)$.

The first step in proving the theorem is to rewrite the model in the following way:

$$\frac{d}{dt}\bar{p} = F(\bar{p})$$

where, for all $i = 1 \dots n$ and $\lambda \in \Lambda$,

$$F_i(\lambda)(\bar{p}) = p_i(\lambda)\left(q_i(\lambda) - \sum_\mu p_i(\mu)q_i(\mu)\right)$$

for the standard model, and

$$F_i(\lambda)(\bar{p}) = p_i(\lambda)\left(\frac{q_i(\lambda)}{\sum_\mu p_i(\mu)q_i(\mu)} - 1\right)$$

for the normalized model.

Let $DF(\bar{p})$ be the differential of F in $\bar{p}$. We show that the real part of each eigenvalue of $DF(\bar{p})$ is negative if $\bar{p}$ is strictly consistent.

We first prove the theorem for the standard model. Deriving F with respect to $p_j(\rho)$, we have:

$$\frac{\partial F_i(\lambda)}{\partial p_j(\rho)}(\bar{p}) = \delta_{ij}\delta_{\lambda\rho}\left(q_i(\lambda) - \sum_\mu p_i(\mu)q_i(\mu)\right) + p_i(\lambda)\left(\frac{\partial q_i(\lambda)}{\partial p_j(\rho)} - \delta_{ij}q_i(\rho) - \sum_\mu p_i(\mu)\frac{\partial q_i(\mu)}{\partial p_j(\rho)}\right) \quad (8)$$

If we arrange the assignment vector in the following way:

$$p = (p_1(\lambda_1), p_1(\lambda_2), \dots, p_1(\lambda_m), p_2(\lambda_1), \dots, p_n(\lambda_m))^T$$

and define the matrices $C_{ij} = (C_{ij}(\lambda,\mu))_{\lambda,\mu}$ as $C_{ij}(\lambda,\mu) = \frac{\partial F_i(\lambda)}{\partial p_j(\mu)}$, the differential takes the form:

$$DF = \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{pmatrix}$$

We can demonstrate that, if $\bar{p}$ is strictly consistent, $C_{ij} = 0$ if $i \neq j$. In fact, we have:

$$\begin{aligned} \frac{\partial F_i(\lambda)}{\partial p_j(\rho)}(\bar{p}) &= p_i(\lambda)\left(\frac{\partial q_i(\lambda)}{\partial p_j(\rho)} - \sum_\mu p_i(\mu)\frac{\partial q_i(\mu)}{\partial p_j(\rho)}\right) \\ &= \delta_{\lambda\lambda(i)}\left(\frac{\partial q_i(\lambda)}{\partial p_j(\rho)} - \frac{\partial q_i\big(\lambda(i)\big)}{\partial p_j(\rho)}\right) = 0 \end{aligned}$$

In this case the differential takes the form:

$$DF = \begin{pmatrix} C_{11} & & 0 \\ & \ddots & \\ 0 & & C_{nn} \end{pmatrix}$$

Analyzing the matrices C_{ii} we can see that these too take a particular form on strictly consistent assignments. In fact we have:

$$\begin{aligned}\frac{\partial F_i(\lambda)}{\partial p_i(\rho)}(\bar{p}) =& \delta_{\lambda\rho}\left(q_i(\lambda) - \sum_\mu p_i(\mu)q_i(\mu)\right) + p_i(\lambda)\left(\frac{\partial q_i(\lambda)}{\partial p_i(\rho)} - q_i(\rho) - \sum_\mu p_i(\mu)\frac{\partial q_i(\mu)}{\partial p_i(\rho)}\right)\\ =& \delta_{\lambda\rho}\left(q_i(\lambda) - q_i\big(\lambda(i)\big)\right) + \delta_{\lambda\lambda(i)}\left(\frac{\partial q_i(\lambda)}{\partial p_i(\rho)} - q_i(\rho) - \frac{\partial q_i\big(\lambda(i)\big)}{\partial p_i(\rho)}\right)\\ =& \delta_{\lambda\rho}\left(q_i(\lambda) - q_i\big(\lambda(i)\big)\right) - \delta_{\lambda\lambda(i)}q_i(\rho)\end{aligned}$$

As we can notice, the non-zero values of C_{ii} are on the main diagonal and on the row $C_{ii}(\lambda\rho)$ with $\lambda = \lambda(i)$. Thus the eigenvalues of C_{ii} are the elements on the main diagonal. These are:

$$\begin{cases} q_i(\lambda) - q_i\big(\lambda(i)\big) & \text{for } \lambda \neq \lambda(i),\\ -q_i\big(\lambda(i)\big) & \text{otherwise.}\end{cases} \tag{9}$$

Since $\bar{p}$ is strictly consistent, $q_i(\lambda) < q_i\big(\lambda(i)\big)$ so all the eigenvalues are real negative and not lower than $-q_i\big(\lambda(i)\big)$. This tells us that the assignment is a "sink," and hence a local attractor for the dynamical system [6].

Now we can prove the theorem for the normalized model. The fundamental steps to follow are the same as for the standard model; we mainly have to derive the new values for the partial derivates:

$$\begin{aligned}\frac{\partial F_i(\lambda)}{\partial p_j(\rho)}(\bar{p}) =& \frac{\delta_{ij}\delta_{\lambda\rho}q_i(\lambda) + p_i(\lambda)\frac{\partial q_i(\lambda)}{\partial p_j(\rho)}}{\sum_\mu p_i(\mu)q_i(\mu)}\\ &- \frac{p_i(\lambda)q_i(\lambda)\left(\delta_{ij}q_i(\rho) + \sum_\mu p_i(\mu)\frac{\partial q_i(\mu)}{\partial p_j(\rho)}\right)}{\left(\sum_\mu p_i(\mu)q_i(\mu)\right)^2} - \delta_{ij}\delta_{\lambda\rho}\\ =& \delta_{ij}\delta_{\lambda\rho}\frac{q_i(\lambda)}{q_i(\lambda(i))} + \delta_{\lambda\lambda(i)}\frac{\frac{\partial q_i(\lambda)}{\partial p_j(\rho)}}{q_i(\lambda(i))} - \delta_{ij}\delta_{\lambda\lambda(i)}\frac{q_i(\lambda)q_i(\rho)}{q_i(\lambda(i))^2}\\ &- \delta_{\lambda\lambda(i)}\frac{q_i(\lambda)\frac{\partial q_i(\lambda(i))}{\partial p_j(\rho)}}{q_i(\lambda(i))^2} - \delta_{ij}\delta_{\lambda\rho}\\ =& \delta_{ij}\delta_{\lambda\rho}\frac{q_i(\lambda)}{q_i(\lambda(i))} - \delta_{ij}\delta_{\lambda\lambda(i)}\frac{q_i(\rho)}{q_i(\lambda(i))} - \delta_{ij}\delta_{\lambda\rho}\end{aligned}$$

Like for the standard model, we have $C_{ij} = 0$ for $i \neq j$, and the matrices C_{ii} are non-zero only on the main diagonal and on the row related to the assignment $\lambda(i)$. Once more, then, the eigenvalues are equal to the elements on the main diagonal. These are:

$$\begin{cases} \frac{q_i(\lambda) - q_i(\lambda(i))}{q_i(\lambda(i))} & \text{for } \lambda \neq \lambda(i),\\ -1 & \text{otherwise.}\end{cases} \tag{10}$$

Thus the eigenvalues are all real negative and not lower than -1, i.e., strictly consistent assignments are sinks, and hence local attractors, for system (6). $\square$

This is the analog to the fundamental local convergence result of Hummel and Zucker [8, Theorem 9.1], which is also valid for the classical relaxation scheme (2) [12, Theorem 6.4]. Note that, unlike Theorem 4.3, no restriction on the structure of the compatibility matrix is imposed here.

5 Discretizing the models

In order to simulate the behavior of the models on a digital computer, we need to make them evolve in discrete rather than continuous time steps. Two well known techniques to approximate differential equations are the Euler method and the Runge-Kutta method. With the Euler method we have:

$$p_i^{t+h}(\lambda) = p_i^t(\lambda) + hF_i^t(\lambda)(\bar{p}) \tag{11}$$

where h is the step size. This equation has the advantage that it can be calculated in a very efficient way, so it is the ideal candidate for our simulations. We will prove that, given a certain integration step h, this model has all the dynamical properties shown for the continuous models it approximates.

In order to determine the difference in global behavior between the continuous models and the discrete approximations, we also use a finer discretization model: the IV grade Runge-Kutta method. This has been done on the assumption that this model would have a global dynamic behavior very similar to that of the continuous models. We have chosen the following Runge-Kutta scheme:

$$p_i^{t+h}(\lambda) = p_i^t(\lambda) + \frac{1}{6}k_1(i,\lambda) + \frac{2}{6}k_2(i,\lambda) + \frac{2}{6}k_3(i,\lambda) + \frac{1}{6}k_4(i,\lambda)$$

where the coefficients k_1, k_2, k_3, k_4 represent:

$$\begin{cases} k_1(i,\lambda) = hF_i(\lambda)(\bar{p}) \\ k_2(i,\lambda) = hF_i(\lambda)\left(\bar{p} + \frac{1}{2}\bar{k}_1\right) \\ k_3(i,\lambda) = hF_i(\lambda)\left(\bar{p} + \frac{1}{2}\bar{k}_2\right) \\ k_4(i,\lambda) = hF_i(\lambda)(\bar{p} + \bar{k}_3) \end{cases}$$

We will prove that the models discretized with Euler's method are well defined; that is, they map points in the assignment space $\mathbb{K}$ onto $\mathbb{K}$. Euler's scheme applied to our standard relaxation model (5) gives:

$$p_i^{t+h}(\lambda) = p_i^t(\lambda) + hp_i^t(\lambda)\left(q_i^t(\lambda) - \sum_\mu p_i^t(\mu)q_i^t(\mu)\right)$$

We note that when h equals 1 the process is identical to the one recently proposed by Chen and Luh [2, 3]. However, their model imposes strict constraints on the compatibility coefficients to insure that $\mathbb{K}$ be invariant with respect to iterations of the process. It can however be proven that, if an appropriate integration step h is chosen, it is not necessary to impose such constraints.

It is easy to prove that $\sum_\lambda p_i(\lambda)$ always equals 1:

$$\sum_\lambda p_i^{t+h}(\lambda) = 1 + h\left(\sum_\lambda p_i^t(\lambda)q_i^t(\lambda) - \sum_\lambda p_i^t(\lambda)\sum_\mu p_i^t(\mu)q_i^t(\mu)\right) = 1$$

But we have to prove that the iteration of the process never leads to negative assignments.

Proposition 5.1 *Let $h \leq 1/q_i(\lambda;\bar{p})$ for all $i, \lambda, \bar{p}$. If we call E the function generated applying Euler's scheme to the model, then if $\bar{p} \in \mathbb{K}$, we have $E_i(\lambda)(\bar{p}) \geq 0$.*

Proof: We have:

$$\begin{aligned} p_i^{t+h}(\lambda) \geq & p_i^t(\lambda) + hp_i(\lambda)\left(q_i^t(\lambda) - \sum_\mu p_i^t(\mu)\frac{1}{h}\right) \\ = & p_i^t(\lambda) + hp_i^t(\lambda)\left(q_i^t(\lambda) - \frac{1}{h}\right) \geq p_i^t(\lambda) - hp_i^t(\lambda)\frac{1}{h} = 0 \end{aligned}$$

which proves the proposition. □

If we use the linear support function (1), the integration step can be:

$$h \leq \frac{1}{\max_{i\lambda}\left\{\sum_j \max_\mu r_{ij}(\lambda,\mu)\right\}}$$

It can readily be seen that this model also corrects deviation from the assignment space, provided that $p_i^t(\lambda) \geq 0$. In fact, given $\sum_\lambda p_i^t(\lambda) = 1 + \varepsilon$ we have:

$$\begin{aligned} \sum_\lambda p_i^{t+h}(\lambda) = & \sum_\lambda p_i^t(\lambda) + h\sum_\lambda p_i^t(\lambda)\left(q_i(\lambda) - \sum_\mu p_i^t(\mu)q_i^t(\mu)\right) \\ = & (1+\varepsilon) + h\left[\left(1 - \sum_\lambda p_i^t(\lambda)\right)\sum_\mu p_i^t(\mu)q_i^t(\mu)\right] \\ = & 1 + \varepsilon - \varepsilon h\sum_\mu p_i^t(\mu)q_i^t(\mu) \end{aligned}$$

Passing to the normalized model, Euler's scheme gives us:

$$p_i^{t+h}(\lambda) = (1-h)p_i^t(\lambda) + h\frac{p_i^t(\lambda)q_i^t(\lambda)}{\sum_\mu p_i^t(\mu)q_i^t(\mu)}$$

As can easily be seen, with $h = 1$, this is the same equation that defines the classical model. Thus for $h = 1$ the model is well defined.

With an h lower than 1 the resulting assignment is a convex linear combination of $\bar{p}$ and of the assignment resulting from applying one iteration of the classical method to $\bar{p}$. Since the assignment space $\mathbb{K}$ is convex, the resulting assignment will also be in $\mathbb{K}$.

We can see that this model is also numerically stable. In fact, with $h = 1$, if we have $p_i(\lambda) \geq 0$, the model corrects any deviation from $\mathbb{K}$ in one step. On the other hand, with $h < 1$, if $\sum_\lambda p_i^t(\lambda) = 1+\varepsilon$, we have:

$$\begin{aligned}\sum_\lambda p_i^{t+h}(\lambda) =&\sum_\lambda(1-h)p_i^t(\lambda) + \sum_\lambda h\frac{p_i^t(\lambda)q_i^t(\lambda)}{\sum_\mu p_i^t(\mu)q_i^t(\mu)} = (1-h)(1+\varepsilon)+h\\ =&(1-h\epsilon)\end{aligned}$$

That is, the iteration of the model reduces the deviation from $\mathbb{K}$ at every step.

It is easy to prove that strictly consistent assignments are local attractors for these discrete models. In order to do this we must note that the differential of E is $I + hDF$; so, given an eigenvalue a of DF, there is an eigenvalue of DE equal to $1 + ha$. Furthermore, this property defines all eigenvalues of DE. As we have seen in (9), the eigenvalues of DF calculated for the standard model are all not lower than $\max_i\Big\{-q_i\big(\lambda(i)\big)\Big\}$ and all strictly lower than 0; so, for any integration step lower than $1/q_i(\lambda)$ for all i and all λ, we have , for any eigenvalue b of DE, $b = 1 + ha \geq 1 - h\frac{1}{h} = 0$ and $b = 1 + ha < 1$. Thus, strictly consistent assignments are hyperbolic attractors for the system [5]. The eigenvalues of DF calculated for the normalized model are all not lower than -1 and all strictly lower than 0 (10); so, for $h \leq 1$ the eigenvalues of DE are all not lower than 0 and all strictly lower than 1. Thus, in this case as well, strictly consistent assignments are hyperbolic attractors for the system.

6 Experimental results

In order to determine the practical behavior of the proposed models, and to determine possible differences in global behavior between our continuous-time models and the classical one, we conducted two series of experiments. Our goal was to verify that the models exhibit the same behavior

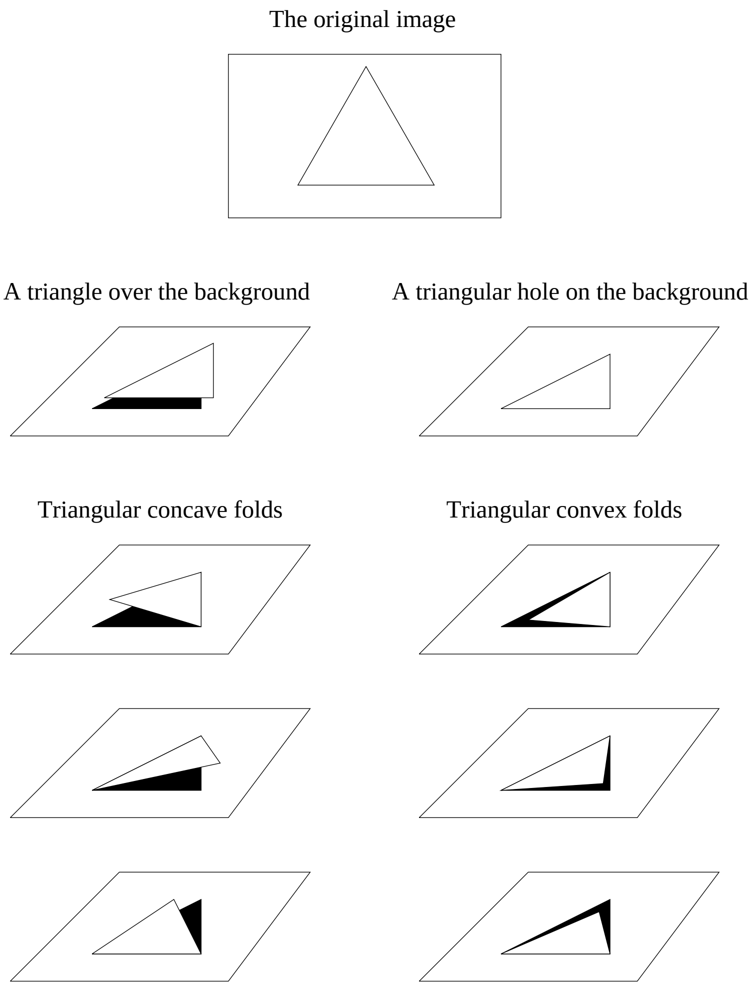

Figure 3: A triangle and its possible interpretations.

as the classical relaxation scheme (2). The experiments were conducted using both the Euler and the Runge-Kutta discretizations described in the previous section.

The first set of simulations were conducted over the classical "triangle" problem introduced as a toy example in the seminal paper by Rosenfeld, Hummel and Zucker [15]. The problem is to label the edges of a triangle as convex, concave, right- or left-occluding. Here, only eight possible labelings are possible (see Figure 3 and [15] for details). The compatibility coefficients used were the same as those given in [15]. As a first control we verified whether the models' behaviors differ, starting from the eight initial assignments given in [15]. From these starting points all the models gave the same sets of classifications. After this preliminary test, we generated 100 random assignments and used them as starting points for each model. The iterations were stopped when the sum of Kullback's I-directed divergence between two successive assignments was lower than 10^{-7}.

Model	Iterations
classical (eq. (2))	79.1
standard, discretized with Euler's scheme	118.8
normalized discretized with Runge-Kutta scheme	81.4
standard, discretized with Runge-Kutta scheme	87.2

Table 1: Average number of iterations for the triangle labeling problem.

All the models converged to a non-ambiguous assignment. Moreover the Euler discretizations of our dynamics gave the same results as the classical model for all initial assignments, while the Runge-Kutta discretizations gave a different result only for one initial assignment. This single assignment was reached with the highest number of iterations of all the assignments generated. This is probably due to the symmetry of the problem: a similar problem can be seen with a uniform probability distribution among assignments. The iteration of each model should converge to the *a priori* probability of each classification, that is 3/8 for each occluding edge and 1/8 for convex or concave edges. What really happens is that the assignments start by heading towards the *a priori* distribution, but, after a few iterations, they head towards a non-ambiguous assignment. This happens because the *a priori* probability is not a hyperbolic attractor for the system. It is possible that a similar problem affected the only initial assignment that gave different results: the models headed toward different non-ambiguous solutions from a unique non-hyperbolic equilibrium that separates the orbits. The average number of iterations that the models needed to reach the stopping criterion is shown in Table 1.

The second set of simulations was carried out by generating random sets of compatibility matrices. This is the set of tests which most effectively point out differences in the dynamic behavior of the models. Since there was no underlying scheme on the pattern of compatibility coefficients, we do not expect the models to converge to a non-ambiguous assignment each time. In fact, in our experiments, the classical model converged to a non-ambiguous assignment only 22% of the time. The aim of this set of tests was to verify whether, when the classical model converges to a non-ambiguous assignment, the other models converge to the same assignment. Ten random coefficients matrices were generated for this experiment and for each matrix the various models were started from ten random assignments. Hence we made a hundred tests for each model. The assignment space dimension was five objects ($n = 5$)and three labels ($m = 3$). The stopping criterion was the same as the previous set of experiments. Here, the classical model (2) converged to a non-ambiguous assignment 22 times. The Runge-Kutta discretization of both models converged

Model	Iterations
classical (eq. (2))	294.0
standard, discretized with Euler's scheme	643.1
normalized discretized with Runge-Kutta scheme	260.1
standard, discretized with Runge-Kutta scheme	324.5

Table 2: Average number of iterations for the random compatibility experiments.

to the same assignments 20 times, while Euler discretization of thge standard model reported the same assignments 17 times. Table 2 reports the average number of iterations needed to reach the stopping criterion.

7 Conclusions

In this paper we have presented and analyzed two relaxation labeling processes. In contrast with the standard approach, these models evolve through continuous-time rather than discrete-time dynamics. This fact permits the design of analog hardware implementations, makes the study of its properties simpler and more elegant, and makes the model more plausible biologically. We have analyzed the dynamical behavior of the models and shown how it is intimately related to Hummel and Zucker's classical theory of consistency. We have proven that the models enjoy exactly the same dynamical properties which have already been proven for the classical processes. The dynamics of the models discretized through Euler's scheme has also been studied. Experimental results confirm the validity of the proposed models.

Acknowledgments. Part of this work was done while the second author was visiting the Center for Computational Vision and Control, Yale University, USA. Funding for this research has been partly provided by the Consiglio Nazionale delle Ricerche, Italy.

References

[1] D. H. Ballard and C. M. Brown, *Computer Vision.* Prentice-Hall, Englewood Cliffs, NJ, 1982.

[2] Q. Chen and J. Y. S. Luh. Ambiguity reduction by relaxation labeling. *Pattern Recognition*, 27(1):165–180, 1994.

[3] Q. Chen and J. Y. S. Luh. Relaxation labeling algorithm for information integration and its convergence. *Pattern Recognition*, 28(11):1705–1722, 1995.

[4] A. Cichoki and R. Unbehauen. *Neural Networks for Optimization and Signal Processing.* Wiley, 1993.

[5] R. L. Devaney. *An Introduction to Chaotic Dynamical Systems.* Addison Wesley, 1989.

[6] M. W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra.* Academic Press, New York, 1974.

[7] J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA*, 81:3088–3092, 1984.

[8] R. A. Hummel and S. W. Zucker. On the foundations of relaxation labeling processes. *IEEE Trans. Pattern Anal. Machine Intell.*, 5:267–287, 1983.

[9] J. Kittler and J. Illingworth. Relaxation labeling algorithms–A review. *Image Vision Comput.*, 3:206–216, 1985.

[10] S. Z. Li, W. Y. C. Soh, and E. K. Teoh. Relaxation labeling using augmented Lagrange-Hopfield method. *Pattern Recognition*, 31(1):73–81, 1998.

[11] S. Peleg and A. Rosenfeld. Determining compatibility coefficients for curve enhancement relaxation processes. *IEEE Trans. Syst. Man Cybern.*, 8:548–555, 1978.

[12] M. Pelillo. The dynamics of nonlinear relaxation labeling processes. *J. Math. Imaging Vision*, 7(4):309–323, 1997.

[13] M. Pelillo and A. M. Fanelli. Autoassociative learning in relaxation labeling networks. *Pattern Recognition Lett.*, 18(1):3–12, 1997.

[14] M. Pelillo and M. Refice. Learning compatibility coefficients for relaxation labeling processes. *IEEE Trans. Pattern Anal. Machine Intell.*, 16(9):933–945, 1994.

[15] A. Rosenfeld, R. A. Hummel, and S. W. Zucker. Scene labeling by relaxation operations. *IEEE Trans. Syst. Man Cybern.*, 6(6):420–433, 1976.

[16] A. J. Stoddart, M. Petrou, and J. Kittler. On the foundations of probabilistic relaxation with product support. *J. Math. Imaging Vision*, to appear.

[17] J. W. Weibull, *Evolutionary Game Theory.* MIT Press, Cambridge, MA, 1995.

[18] S.-S. Yu and W.-H. Tsai. Relaxation by the Hopfield neural network. *Pattern Recognition*, 25(2):197–209, 1992.

[19] S. W. Zucker, A. Dobbins, and L. Iverson. Two stages of curve detection suggest two styles of visual computation. *Neural Computat.*, 1:68–81, 1989.

Biographical sketches

Andrea Torsello received the "Laurea" degree with honors in Computer Science from the University of Venice, Italy, in 1997. His research interests are in pattern recognition and computer vision.

Marcello Pelillo was born in Taranto, Italy, on June 1, 1966. He received the "Laurea" degree with honors in Computer Science from the University of Bari, Italy, in 1989.

From 1988 to 1989 he spent about a year at the IBM Scientific Center in Rome, where he was involved in studies on natural language and speech processing. In 1991 he joined the Department of Computer Science at the University of Bari, Italy, as Assistant Professor. Since 1995 he is with the Department of Applied Mathematics and Computer Science at the University of Venice, Italy. He held visiting positions at Yale University (USA), McGill University (Canada), the University of Vienna (Austria), and the Univeristy of York (UK).

His research interests are in the areas of computer vision, neural networks and pattern recognition, where he has published over 50 papers in refereed journals and conference proceedings. He has been the program co-chair of the *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition* held in Venice, Italy, in May 1997, and he is Guest Co-editor of a special issue of the journal *Pattern Recognition* devoted to this theme. He is member of the IEEE Computer Society and the Pattern Recognition Society.