A Multi-Valued Boltzmann Machine

C. T. Lin and C. S. G. Lee

Abstract— **The idea of Hopfield network is based on the Ising spin glass model in which each spin has only two possible states: up and down. By introducing stochastic factors into this network and performing a simulated annealing process on it, it becomes a Boltzmann machine which can escape from local minimum states to achieve the global minimum. This paper generalizes the above ideas to *multi-value* case based on the XY spin glass model in which each spin can be in any direction in a plane. Simply using the gradient descent method and the analog Hopfield network, two different analog connectionist structures and their corresponding evolving rules are first designed to transform the XY spin glass model to distributed computational models. These two analog computational models are single-layered connectionist structures and multi-layered Hopfield analog networks. The latter network eases the node (neuron) computational requirement of the former at the expense of more neurons and connections. With the proposed evolving rules, the proposed models evolve according to a predefined Hamiltonian (energy function) which will decrease until it reaches a (perhaps local) minimum. Since these two structures can easily get stuck in local minima, a *multi-valued Boltzmann machine* is proposed which adopts the discrete planar spin glass model for the local minimum problem. Each neuron in the multi-valued Boltzmann machine can only take n discrete directions (states). The stochastic simulated annealing method is introduced to the evolving rules of the multi-valued Boltzmann machine to solve the local minimum problem. The multi-valued Boltzmann machine can be applied to the mobile robot navigation problem by defining proper *artificial magnetic field* on the traverse terrain. This new artificial magnetic field approach for the mobile robot navigation problem has shown to have several advantages over existing graph search and potential field techniques.**

I. Introduction

Spin glass is a model which can be used to investigate the collective properties of physical systems made from a large number of simple elements. The interactions among these elementary components yield collective phenomena, such as stable magnetic orientations and the crystallizing state of metal or alloy annealing. In the *Ising spin glass* model [1]–[3], an Ising spin on a lattice point takes on one of two possible values (directions) (i.e., ± 1 or up and down). By generalizing the Ising spin glass model to a *XY spin glass* model [4], [5], each spin can point to any direction in a plane instead of just two possible directions. Analogous to the Ising spin glass model, Hopfield proposed a well-known Hopfield network [6], [7] which was found to have many useful applications, especially in associative memory and optimization problems [8]. Since a Hopfield network can easily get stuck in a local minimum state in optimization problems, a Boltzmann machine [9] was proposed in which probabilistic transitions and the simulated annealing scheme are incorporated into the Hopfield network. A proper cooling process and a stochastic response function in a Boltzmann machine can prevent the system from getting trapped in a locally optimal state, and it can escape from a local minimum to reach the global minimum.

Manuscript received May 29, 1993; revised May 18, 1994. This work was supported in part by the National Science Foundation under Grant CDR 8803017 to the Engineering Research Center for Intelligent Manufacturing Systems and in part by the National Science Council of Taiwan under Grant NSC 82-0113-E-009-296-T.

C. T. Lin is with the Department of Control Engineering, National Chiao-Tung University, Hsinchu, Taiwan, R.O.C.

C. S. G. Lee is with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907 USA.

IEEE Log Number 9406644.

Based on the XY spin glass model [5], this paper generalizes the concepts of the previous neural network systems to develop a multi-valued Boltzmann machine. Unlike the original Boltzmann machine [9] in which each neuron has only two possible states, the proposed machine allows its neuron to have more than two possible states. Associated with a XY spin glass model is an interaction energy called Hamiltonian which describes the total system energy. The Hamiltonian of a XY spin glass model changes when the spins in the model change their states (directions). To utilize this physical model for a computational problem, a proper Hamiltonian must be defined according to the specific application problem. As the system evolves according to some evolving rules, the predefined Hamiltonian will decrease until it reaches a minimum which corresponds to the solution of the computational problem. So, in our design, both the structure and the evolving rules must be considered for emulating a XY spin glass model.

We first design two continuous-valued connectionist structures with their corresponding evolving rules to simulate the functions of a XY spin glass model. The state of each neuron in these two structures can be in any direction continuously in a plane. The design of these two structures is based on the gradient descent method and the analog Hopfield network [7]. First, a single-layered connectionist structure with a straight-forward evolving rule based on polar coordinates is proposed. This rule is then revised to form two more efficient evolving rules. In this structure, each neuron (computational unit) is required to perform some complex computations, including trigonometric functions. To ease this requirement, we further design another structure, a multi-layered Hopfield analog network, based on the Cartesian XY coordinates. The computational requirement in this structure is simpler at the expense of more neurons (hardware). Although these two structures can realize the functions of a XY spin glass model, they, like the Hopfield network, suffer the local minimum problem. To solve the local minimum problem, a multi-valued Boltzmann machine is proposed and developed according to the discrete planar spin model [5] which is the digitalization of the XY spin glass model. In the discrete planar spin model, each spin can take only n discrete directions in a plane, where n is a positive integer. The discrete planar spin model can simulate the XY spin glass model when n is large. Based on the concept of the discrete planar spin model, each neuron in a multi-valued Boltzmann machine is allowed to have only a limited number of stable states (directions). To utilize the multi-valued Boltzmann machine for a computational problem, a Hamiltonian (energy function) is first defined according to the *a priori* or initial knowledge and the external input information of the computational problem, which correspond to the internal interaction magnetic field and the external local magnetic field in the XY spin glass model, respectively. Then each neuron will change its state properly according to the proposed evolving rule. The stochastic simulated annealing method is applied to the evolving rule such that a neuron can decide its next state stochastically according to the energy gap between two possible states. With the simulated annealing scheme [9], a state change which will increase the total system energy is allowed in the earlier evolving process. This can help the system to jump out of a possible minimum and reach a global minimum.

The proposed multi-valued Boltzmann machine provides a new *artificial magnetic field* approach to the mobile robot navigation problem for solving the collision-free path problem. The obstacle avoidance in the mobile robot navigation problem has been a difficult problem in artificial intelligence, and two basic approaches have been

taken: graph search methods [10]–[14] and potential field methods [15]–[20]. The potential field methods use artificial potential fields applied to the obstacles and goal positions and use the resulting field to influence the path of the mobile robot which is subject to this potential. The artificial magnetic field in a multi-valued Boltzmann machine is the parallel part of the artificial potential in the potential field methods. However, unlike the potential field methods in which proper potential functions need to be defined, the magnetic fields are defined as the link weights among neurons or external inputs to the neurons. In this way, the difficulties for the potential field approaches to navigate a mobile robot to pass around a concave-shaped object or through narrow passages can be solved. More importantly, by introducing the stochastic simulated annealing method to the evolving rule of the multi-valued Boltzmann machine, the local minimum problem, which is the major deficiency of the potential field approach, can be avoided.

In the next section, the XY spin glass model is first introduced. Then based on the gradient descent method and the analog Hopfield network, two analog connectionist structures and their corresponding evolving rules are derived. The multi-valued Boltzmann machine is proposed in Section III. The application of the proposed multi-valued Boltzmann to the mobile robot navigation problem is discussed and illustrated in Section IV. Conclusions are summarized in the last section.

II. XY Spin Glass Computational Models

In this section, the XY spin glass model in [5] is first introduced. Then, to transform this physical model to a computational model for practical applications, we design two analog connectionist structures and their corresponding evolving rules based on the general gradient descent method and the analog Hopfield network. The neurons in these two structures have graded response.

A. XY Spin Glass Model

Spin glass systems are characterized by a set of values for the spin variable $\boldsymbol{s}_i$, by a lattice on which they are defined (we shall consider the one-dimensional case first), and by an interaction energy, Hamiltonian (E). In the Ising spin system, each spin (called "classic spin") has only two possible values (i.e., ± 1) [1]–[3]. The classical XY spin glass system is composed of planar 2-component spins, and each spin can be in any direction [4], [5]. Hence, we can define $\boldsymbol{s}_i = (\cos\theta_i, \sin\theta_i)^T$ to be a classical vector of unit length located on a lattice point i, where θ_i is the state of $\boldsymbol{s}_i$, and superscript "T" denotes vector/matrix transpose.

To transform the physical model in [5] to a computational model for practical applications, the Hamiltonian of a spin glass model comes from two contributions.

$$\begin{aligned} E = E_{\text{total}} &\triangleq E_{\text{exchange}} + E_{\text{field}} \\ &= -\sum_{\langle i,j\rangle} J_{ij}(\boldsymbol{s}_i \cdot \boldsymbol{s}_j) - \sum_i \boldsymbol{f}_i \cdot \boldsymbol{s}_i \end{aligned} \tag{1}$$

where J_{ij} is a nearest-neighbor exchange interaction between spins $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, $\boldsymbol{f}_i$ is the external local magnetic field inserted on the spin $\boldsymbol{s}_i$, and the first summation is taken over all the nearest-neighboring pairs $\langle i,j\rangle$; that is, $J_{ij} \neq 0$. Thus, E_{exchange} is the internal magnetic field produced by the surrounding spins. This magnetic field creates a tendency to align all the spins in a regular way, and results in a regularizing effect of the exchange interaction. This spin-spin interaction represents *a priori* knowledge on relationship between spins, for example, the continuity property. E_{field} is the external local magnetic field which describes the interaction of the spins with the local magnetic field, which tends to align the spins locally within the field. This field represents the external input information in a computational model. Hence, there are two kinds of interaction: the interaction among spins (J_{ij}) and the interaction with the external field ($\boldsymbol{f}_i$). If we let

$$\boldsymbol{h}_i = \sum_j J_{ij}\boldsymbol{s}_j, \tag{2}$$

then $\boldsymbol{h}_i$ represents the internal magnetic field at site i produced by the surrounding spins. Then (1) becomes

$$E = -\sum_i \boldsymbol{h}_i \cdot \boldsymbol{s}_i - \sum_i \boldsymbol{f}_i \cdot \boldsymbol{s}_i = -\sum_i (\boldsymbol{h}_i + \boldsymbol{f}_i) \cdot \boldsymbol{s}_i. \tag{3}$$

Let $\boldsymbol{m}_i = \boldsymbol{h}_i + \boldsymbol{f}_i$. Equation (3) becomes

$$E = -\sum_i \boldsymbol{m}_i \cdot \boldsymbol{s}_i = -\sum_i |\boldsymbol{m}_i| \cos(\delta_i - \theta_i). \tag{4}$$

where $|\boldsymbol{m}_i|$ indicates the magnitude of $\boldsymbol{m}_i$, and δ_i is the angle of $\boldsymbol{m}_i$.

If the XY spin glass model is extended to the two-dimensional case, each spin will occur in a two-dimensional mesh grid lattice. Each spin in a lattice site (i,j) is denoted as $\boldsymbol{s}_{ij}$ and its state is denoted as θ_{ij}. The internal interaction field on $\boldsymbol{s}_{ij}$ from $\boldsymbol{s}_{kl}$ is denoted as $j_{ij,kl}$, and the local external magnetic field on $\boldsymbol{s}_{ij}$ is denoted as $\boldsymbol{f}_{ij}$. The Hamiltonian is then defined as

$$\begin{aligned} E = E_{\text{total}} &\triangleq E_{\text{exchange}} + E_{\text{field}} \\ &= -\sum_{\langle ij,kl\rangle} J_{ij,kl}\boldsymbol{s}_{ij} \cdot \boldsymbol{s}_{kl} - \sum_{ij} \boldsymbol{f}_{ij} \cdot \boldsymbol{s}_{ij}. \end{aligned} \tag{5}$$

The XY spin glass model can provide us with collective computational abilities if the energy function (Hamiltonian) is chosen such that there will be some orientation of one spin which gives the minimum of the potential energy. In the following subsection, based on the XY spin glass model, a single-layered connectionist model and a multi-layered Hopfield analog network are proposed and designed to have such kinds of computational capabilities. These two models are basically the analog connectionist realization of the XY spin glass model, since each computation cell (neuron) in these two models is allowed to point to any direction in a plane.

B. Analog Connectionist Computational Structures for the XY Spin Glass Model

In this subsection, the XY spin glass model is transformed to a computational model for practical applications. For this purpose, two different connectionist structures, a single-layered connectionist structure and a multi-layered Hopfield analog network, are designed using the gradient descent method and the analog Hopfield network [7].

Fig. 1 shows the structure of the proposed single-layered connectionist architecture. The external local magnetic field, $\boldsymbol{f}_{ij}$, is further decomposed into an attractive field $\boldsymbol{g}_{ij}$ and a repulsive field $\boldsymbol{r}_{ij}$ for some application domains while still maintaining its generality. The Hamiltonian for this structure (in (5)) becomes

$$\begin{aligned} E &= E_{\text{exchange}} + E_{\boldsymbol{g}\text{-field}} + E_{\boldsymbol{r}\text{-field}} \\ &= -\sum_{ij}\sum_{kl\in N(ij)} J_{ij,kl}\boldsymbol{s}_{ij} \cdot \boldsymbol{s}_{kl} \\ &\quad - J_{\boldsymbol{g}}\sum_{ij} \boldsymbol{g}_{ij} \cdot \boldsymbol{s}_{ij} - J_{\text{r}}\sum_{ij} \boldsymbol{r}_{ij} \cdot \boldsymbol{s}_{ij} \end{aligned} \tag{6}$$

where $N(ij)$ is the neighbor of (ij) but $(ij) \notin N(ij)$, and $J_{\boldsymbol{g}}$ and $J_{\boldsymbol{r}}$ are scalar constants. A computational unit (i.e., a neuron) corresponds to a spin at a lattice site, and $J_{ij,kl}$ can be viewed as the connection weight between neurons $\boldsymbol{s}_{ij}$ and $\boldsymbol{s}_{kl}$. Through $J_{ij,kl}$, $\boldsymbol{s}_{ij}$ can get its inputs from other neurons. Hence, $\boldsymbol{s}_{kl}$ is called a "neighbor" of $\boldsymbol{s}_{ij}$

if $J_{ij,kl} \neq 0$. $\boldsymbol{g}_{ij}$ and $\boldsymbol{r}_{ij}$ are the weights on the direct input links to neuron $\boldsymbol{s}_{ij}$ from the external world. The connections in Fig. 1 show the eight-nearest-neighbors on a mesh grid lattice. Next, we need to derive the evolving rules of θ_{ij} for each neuron $\boldsymbol{s}_{ij}$. Let

$$\boldsymbol{h}_{ij} = \sum_{kl \in N(ij)} J_{ij,kl} \boldsymbol{s}_{kl}, \tag{7}$$

then (6) becomes

$$E = -\sum_{ij} (\boldsymbol{h}_{ij} + J\boldsymbol{g}\boldsymbol{g}_{ij} + J_{\boldsymbol{r}}\boldsymbol{r}_{ij}) \cdot \boldsymbol{s}_{ij}. \tag{8}$$

Let

$$\boldsymbol{m}_{ij} = \boldsymbol{h}_{ij} + J\boldsymbol{g}\boldsymbol{g}_{ij} + J_r \boldsymbol{r}_{ij}, \tag{9}$$

then

$$E = -\sum_{ij} \boldsymbol{m}_{ij} \cdot \boldsymbol{s}_{ij} = -\sum_{ij} |\boldsymbol{m}_{ij}| \cos(\delta_{ij} - \theta_{ij}) \tag{10}$$

where $|\boldsymbol{m}_{ij}|$ indicates the magnitude of $\boldsymbol{m}_{ij}$, δ_{ij} is the angle of $\boldsymbol{m}_{ij}$, and $0 \leq \delta_{ij} < 2\pi$. Applying the gradient descent method to (10), we have

$$\frac{dE}{dt} = \sum_{ij} \frac{\partial E}{\partial \theta_{ij}} \frac{d\theta_{ij}}{dt}. \tag{11}$$

Since

$$\frac{\partial E}{\partial \theta_{ij}} = -|\boldsymbol{m}_{ij}| \sin(\delta_{ij} - \theta_{ij}), \tag{12}$$

if we let

$$\frac{d\theta_{ij}}{dt} = -k \frac{\partial E}{\partial \theta_{ij}} = k|\boldsymbol{m}_{ij}| \sin(\delta_{ij} - \theta_{ij}), \quad \text{for } k \geq 0, \tag{13}$$

then we have

$$\frac{dE}{dt} = -\sum_{ij} k|\boldsymbol{m}_{ij}|^2 \sin^2(\delta_{ij} - \theta_{ij}) \leq 0.$$

So the energy always decreases or remains unchanged.

Eq. (13) is a feasible evolving rule of θ_{ij}; however, it has two drawbacks. First, the updating magnitude (distance) becomes smaller and smaller when $|\delta_{ij} - \theta_{ij}| > (\pi/2)$. Second, the gradient descent rule fails when $|\delta_{ij} - \theta_{ij}| = \pi$, since $\partial E/\partial \theta_{ij} = 0$ at this critical point. To speed up the convergence and avoid the critical point, we revise the evolving rules according to the following two principles:

1) The evolving rule of θ_{ij} must yield $dE/dt \leq 0$.
2) The updating magnitude (distance) factor, k, must be small as compared to $|\delta_{ij} - \theta_{ij}|$ to avoid over-running.

The first revised evolving rule uses a fixed updating distance:

$$\frac{d\theta_{ij}}{dt} = \begin{cases} k \operatorname{sign}(\sin(\delta_{ij} - \theta_{ij})), & \text{if } |\delta_{ij} - \theta_{ij}| \neq \pi \\ k(\text{or } -k), & \text{if } |\delta_{ij} - \theta_{ij}| = \pi \end{cases} \tag{14}$$

Using (14), (11) becomes

$$dE/dt = -\sum_{ij} k|\boldsymbol{m}_{ij}| |\sin(\delta_{ij} - \theta_{ij})| \leq 0.$$

The second revised evolving rule uses a changing updating distance:

$$\frac{d\theta_{ij}}{dt} = \begin{cases} k \sin(\delta_{ij} - \theta_{ij}), & \text{if } |\delta_{ij} - \theta_{ij}| \leq \pi/2 \\ k(-2 - \sin(\delta_{ij} - \theta_{ij})), & \text{if } -\pi \leq \delta_{ij} - \theta_{ij} \leq -\pi/2 \\ k(2 - \sin(\delta_{ij} - \theta_{ij})), & \text{if } \pi/2 \leq \delta_{ij} - \theta_{ij} \leq \pi. \end{cases} \tag{15}$$

It can be easily shown that $dE/dt \leq 0$ in any of the above cases.

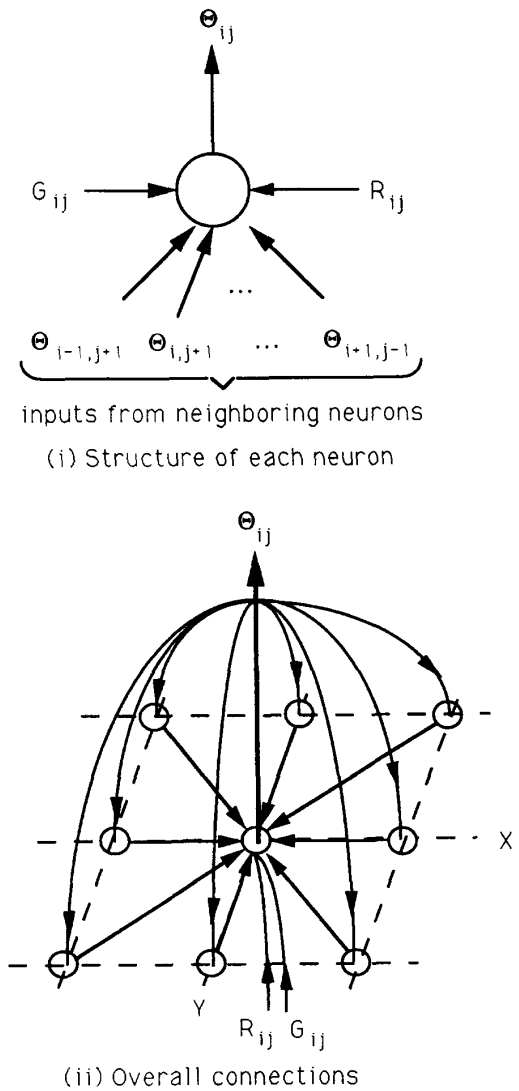

Fig. 1. Single-layered connectionist structure.

Any one of the above equations, (13), (14), or (15) can be used as an evolving rule of the neuron in Fig. 1 to update their θ_{ij}'s. Obviously, each neuron (computational unit) is required to have the capability of computing vector addition and trigonometric functions. To avoid this complex computational requirement, we further design a multi-layered Hopfield analog network by using more neurons with simpler computational functions. This design is based on the existing analog Hopfield network which comes from the concept of Ising spin glass model [1], [2]. Instead of using polar coordinates as we did in the single-layered connectionist structure, Cartesian (X-Y) coordinates are used for the multi-layered Hopfield analog network. We use two individual analog Hopfield networks to process the X coordinate and Y coordinate. These two networks are connected together through the neurons of the third layer. The multi-layered Hopfield analog network provides a simple way to realize the XY spin glass model using existing hardware techniques [25]. The detailed design process and the structure of this network is shown in the Appendix.

When applied to optimization problems like the application in the mobile robot navigation problem, the above proposed structures suffer from the local minimum problem, that is, the system easily gets stuck in a local minimum without further evolution. In this case, a solution of the computational problem cannot be obtained. To obtain a robust computational model, we develop a multi-valued Boltzmann machine which can avoid the local minimum problem. The proposed multi-valued Boltzmann machine is based on the discrete planar XY spin glass model [5] for its structure, and it utilizes stochastic simulated annealing methods [6] to develop its associated evolving rules.

III. A Multi-valued Boltzmann Machine

As pointed out by Tanaka and Edwards [3], [5], there are many local minimum states in the XY spin glass model. It is possible that the above updating algorithms will cause the network to stuck in a local minimum. To solve this problem, stochastic simulated annealing methods are introduced in our algorithm. Before doing this, the concept of "discrete planar spin" must be introduced first. A discrete planar spin $\boldsymbol{s}_i$ at lattice site i can take only n discrete directions in a plane. If $n = 2$, then it becomes an Ising spin glass model, and if $n \to \infty$, then it becomes the above XY spin glass model. The angle between two adjacent directions is given by $\omega_n = 2\pi/n$. To find the local minimum of the energy function in (4), we must study the energy difference for a deformation of the ith spin, $\boldsymbol{s}_i$. Let E_i be the local energy at lattice site i for a given configuration of $\{J_{ij}\}$ and be defined as (see (2)–(4))

$$E_i \triangleq -\boldsymbol{s}_i \cdot \boldsymbol{m}_i. \tag{16}$$

If we change the direction of $\boldsymbol{s}_i$ by the angles $m\omega_n (m = 1, 2, 3, \cdots, n-1)$ from the starting configuration, the local energy after such deformation becomes

$$\begin{aligned} E_i' &= -|\boldsymbol{m}_i| \cos(\theta_i + m\omega_n - \delta_i) \\ &= -(\cos m\omega_n)(\boldsymbol{s}_i \cdot \boldsymbol{m}_i) - (\sin m\omega_n)(\boldsymbol{s}_i \times \boldsymbol{m}_i)_z \end{aligned} \tag{17}$$

where $(\boldsymbol{s}_i \times \boldsymbol{m}_i)_z = -|\boldsymbol{m}_i| \sin(\theta_i - \delta_i)$. The energy difference becomes

$$\begin{aligned} \Delta E_i \triangleq E_i' - E_i &= (1 - \cos m\omega_n)(\boldsymbol{s}_i \cdot \boldsymbol{m}_i) \\ &\quad - (\sin m\omega_n)(\boldsymbol{s}_i \times \boldsymbol{m}_i)_z. \end{aligned} \tag{18}$$

The local minimum state can be defined as $\Delta E_i \geq 0$ for all site i and for $m = 1$ and $n - 1$ (or equivalently, $m = \pm 1$). This is the stability condition against a single-spin deformation by one elementary step in either direction. This condition can be written explicitly as

$$a_n(\boldsymbol{s}_i \cdot \boldsymbol{m}_i) \pm b_n(\boldsymbol{s}_i \times \boldsymbol{m}_i)_z \geq 0 \qquad \text{for all } i, \tag{19}$$

where $a_n = 1 - \cos\omega_n$, and $b_n = \sin\omega_n$. Notice that, when $n \to \infty$, the above condition gives $(\boldsymbol{s}_i \times \boldsymbol{m}_i)_z = 0$ and $(\boldsymbol{s}_i \cdot \boldsymbol{m}_i) \geq 0$. This is equivalent to saying that $\boldsymbol{m}_i = k_i \boldsymbol{s}_i$, where $k_i \geq 0$. That is, the spin $\boldsymbol{s}_i$ always tries to align itself in the direction of $\boldsymbol{m}_i$.

Based on the properties of the discrete planar spin system discussed above, a multi-valued Boltzmann machine is proposed according to the Metroplis algorithm [26]. The Metroplis algorithm states that if the energy gap between two states, θ_i and θ_j of the kth unit, is ΔE_k (where $\Delta E_k = E_{\theta_i} - E_{\theta_j}$), and suppose unit k is in state θ_i, then unit k will change to state θ_j with probability

$$P_k = \frac{1}{1 + e^{-\Delta E_k / T}} \tag{20}$$

where T is a parameter that acts like "temperature." This probability is called *Boltzmann distribution*. The temperature T will decrease in the evolving process to simulate the cooling effect in the annealing process [26]. So the probability of finding the system in any global state will obey a Boltzmann distribution. It is this cooling process and the stochastic state changing that enable the system to escape from a local minimum to achieve the global minimum [26].

To incorporate the Metroplis algorithm into our system, the discrete planar spin model is used. If a particular spin, $\boldsymbol{s}_{ij}$, is currently in state (angle) θ_{ij}, then the local energy defined above can be obtained from (10) as $E_{ij} = -\boldsymbol{s}_{ij} \cdot \boldsymbol{m}_{ij}$. If we change the direction of $\boldsymbol{s}_{ij}$ by an angle $m\omega_n$ $(m = 1, 2, 3, \cdots, n-1)$ from θ_{ij}, then according to (17), the local energy after such deformation becomes

$$E_{ij}' = -(\cos m\omega_n)(\boldsymbol{s}_{ij} \cdot \boldsymbol{m}_{ij}) - (\sin m\omega_n)(\boldsymbol{s}_{ij} \times \boldsymbol{m}_{ij})_z, \tag{21}$$

where $(\boldsymbol{s}_{ij} \times \boldsymbol{m}_{ij})_z = |\boldsymbol{m}_{ij}| \sin(\theta_{ij} - \delta_{ij})$ is the third component of the external product in the spin space. From (18), the energy difference becomes

$$\begin{aligned} \Delta E_{ij} &= E_{ij} - E_{ij}' = -(1 - \cos m\omega_n)(\boldsymbol{s}_{ij} \cdot \boldsymbol{m}_{ij}) \\ &\quad + (\sin m\omega_n)(\boldsymbol{s}_{ij} \times \boldsymbol{m}_{ij})_z \\ &= -(1 - \cos m\omega_n)|\boldsymbol{m}_{ij}| \cos(\theta_{ij} - \delta_{ij}) \\ &\quad + (\sin m\omega_n)|\boldsymbol{m}_{ij}| \sin(\theta_{ij} - \delta_{ij}). \end{aligned} \tag{22}$$

From (20), the probability that spin $\boldsymbol{s}_{ij}$ changes to state (angle) $(\theta_{ij} + m\omega_n)$ is

$$P_{ij} = \frac{1}{1 + e^{-\Delta E_{ij}/T}}. \tag{23}$$

In the case of $n = 8$, (for example, in an eight-nearest-neighbor situation, each spin may point to one of its neighbors), $\omega_n = \pi/4$. If the current angle of $\boldsymbol{s}_{ij}$ is θ_{ij}, then (22) and (23), respectively, become

$$\begin{aligned} \Delta E_{ij}(m) &= -\left(1 - \cos\frac{m\pi}{4}\right)|\boldsymbol{m}_{ij}| \cos(\theta_{ij} - \delta_{ij}) \\ &\quad + \left(\sin\frac{m\pi}{4}\right)|\boldsymbol{m}_{ij}| \sin(\theta_{ij} - \delta_{ij}) \end{aligned} \tag{24}$$

$$P_{ij}(m) = \frac{1}{1 + e^{-\Delta E_{ij}(m)/T}}. \tag{25}$$

So the neuron $\boldsymbol{s}_{ij}$ will change its state stochastically according to these probabilities:

$$\text{Prob}\left(\boldsymbol{s}_{ij} \text{ in } \theta_{ij} + \frac{\pi}{4}\right) = P_{ij}(1),$$

$$\text{Prob}\left(\boldsymbol{s}_{ij} \text{ in } \theta_{ij} + \frac{2\pi}{4}\right) = P_{ij}(2), \cdots,$$

and

$$\text{Prob}\left(\boldsymbol{s}_{ij} \text{ in } \theta_{ij} + \frac{7\pi}{4}\right) = P_{ij}(7).$$

The evolving rule in (24) always has a tendency to decrease the total system energy. However, from (25), when T is a large value (high temperature), a neuron has a higher probability to change to a state which will increase total system energy than when T is a small value (low temperature). So in high temperature, the system tends to escape from local minima, and in low temperature, the system tends to converge to a global minimum. This enables the system to get rid of a local minimum and seek for a global minimum.

IV. Application to Mobile Robot Navigation Problem

In this section, existing approaches to the mobile robot navigation problem are reviewed. This background introduction serves to show the motivation of our application and the advantages of our approach. Then, the use of the proposed multi-valued Boltzmann machine on the mobile robot navigation problem is discussed.

A. Mobile Robot Navigation Problem

Endowing mobile robots with the ability to plan their own collision-free paths is a difficult problem in Artificial Intelligence that has received a great deal of research interest. Two basic approaches have been taken: graph search (global) methods [10]–[14] and potential field (local) methods [15]–[20]. The graph search methods rely on an exhaustive search of the unoccupied configuration space for a continuous path from the start point to the goal point. A chart or graph is first produced showing free spaces, where no collision will occur, and forbidden spaces, where a collision will occur. Based on this graph, a collision-free path is then selected by piecing together the free spaces or by tracing around the forbidden spaces. Global methods generally entail searching for a path in a graph, where the graph is computed from the constraints [10]–[13]. Especially in [13], some route planning algorithms were proposed. These algorithms begin by assigning a cost to each grid cell of a digital terrain map. These costs are usually selected according to some mission criteria, with the highest costs associated with untraversable areas. A search algorithm, such as A^* search [27], is then applied to obtain a value for each grid cell, indicating the minimum cost remaining to get from that cell to the goal. The best incremental step to get to the goal from any given cell is obtained by finding the neighboring cell with the lowest value. In this way, an optimal path can be found from any starting point to the goal. Although the graph search methods can always yield the set of all optimal paths to a goal without local minimum problem, it is rather time consuming. Thus, they are suited only for off-line path planning and cannot be used for real-time obstacle avoidance. Moreover, since the graph search only specifies a predetermined path, this route abstraction loses a lot of information needed in real-time decisionmaking. In [14], a gradient field representation of a traverse terrain is proposed to solve the information-losing problem due to abstraction. In this representation, a proper direction for a mobile robot to move in is indicated for each grid cell of a terrain map. The gradient field is a natural by-product of existing route planning algorithms in [13].

In the potential field methods, information about the local environment is used in real-time to generate a control input for the mobile robot which brings the robot closer to the goal while avoiding nearby obstacles. Among existing potential field methods, the artificial potential method is the most popular and frequently used local method. The idea of using "potential functions" for the specification of robot tasks was pioneered by Khatib [15] in the context of obstacle avoidance; the methodology was also developed independently by Arimoto [16]. The potential field methods use artificial potential fields applied to the obstacles and goal positions and use the resulting field to influence the path of the mobile robot which is subject to this potential. In the artificial potential method, obstacles to be avoided are surrounded by repulsive potential functions and the goal point is surrounded by an attractive well. These potentials are added to form a composite potential that indicates the movement of a robot. In [17], the potential function control was presented in the context of impedance control to avoid a moving obstacle while heading for a fixed goal. The potential function control was enhanced through the use of a "reserve avoidance time" in [18] such that potential functions increase their amplitudes depending on the minimum braking distance to stop a robot from colliding with an obstacle. In [19], a variant of the potential field method was used to produce the appropriate velocity and steering commands for the robot. Although not as thorough as the graph search techniques, the computational speed of the potential field methods and their ease of extension to higher dimensions make them an excellent alternative to the graph search techniques. They provide the necessary framework to deal with changing environments and can be used for real-time obstacle avoidance. However, the artificial potential methods have one major problem—the addition of attractive and repulsive potentials can create local minima in the potential and the mobile robot may get stuck in a local minimum. Moreover, the potential functions also make the mobile robot difficult, if not impossible, to traverse through narrow passages and around concave-shaped obstacles due to the limitation of chosen functions. The local minimum problem was attacked in [20] by creating potential functions based on superquadrics, which closely model a large class of object shapes; however, this approach requires spherically symmetric attractive wells and complex computations.

Several attempts have been made to use the best features of both graph search and potential fields methods. The geometrical solutions for global planning and potential fields for local planning were used in [21], [22]. These benefit from the global planning ability of the graph search methods, but suffer from their same shortcomings as well. In [23], a trial path is chosen and then modified under the influence of the potential field until an appropriate path is found. In [24], a stochastic technique for path planning was proposed. The algorithm incrementally builds a graph connecting the local minima of a potential function and concurrently searches this graph until a goal configuration is attained. The local minimum graph is searched using a depth-first strategy [27] with random backtracking.

In the next subsection, a new approach for the mobile robot navigation problem is presented using the proposed multi-valued Boltzmann machine. The basic concept of this approach comes from the artificial potential field methods. However, several problems of the artificial potential field approach are tackled and solved.

B. Artificial Magnetic Field Approach to Mobile Robot Navigation Problem

The proposed multi-valued Boltzmann machine has been applied to the mobile robot navigation problem using an *artificial magnetic field* approach. The artificial magnetic field is the parallel part of the artificial potential field in the potential field methods. Similar to the potential field approaches, an attractive magnetic field is defined for the goal position and a repulsive magnetic field is defined for an obstacle. However, instead of defining the potential fields as potential functions, the magnetic fields are defined as the link weights among neurons or the external inputs to each neuron in the multi-valued Boltzmann machine. In this way, a magnetic field can be defined to surround any irregular-shaped (including concave-shaped) object accurately. Thus, the proposed approach can navigate a mobile robot to traverse around a concave-shaped object or pass through narrow passages, which are not easy for the potential field methods as mentioned above. More importantly, since the stochastic simulated annealing method is used in the evolving rule of the multi-valued Boltzmann machine, the local minimum problem, which is the major deficiency of the potential field approach, can be avoided.

The 2-D mobile robot navigation problem can be stated as: Given a 2-D transverse map divided into small grid cells, the goal location and the obstacle locations, find the proper moving direction of each grid cell that can navigate the mobile robot to the goal from any "free" grid cell (the grid cell without any obstacle on it) quickly without colliding with any obstacles. This problem can be formulated as follows. Each free grid cell corresponds to a spin, and the state (direction) of a spin (neuron) represents the proper moving direction on the corresponding grid cell. An energy function (Hamiltonian) whose ground state represents the proper spin directions will be defined, and it will then be minimized on the proposed multi-valued

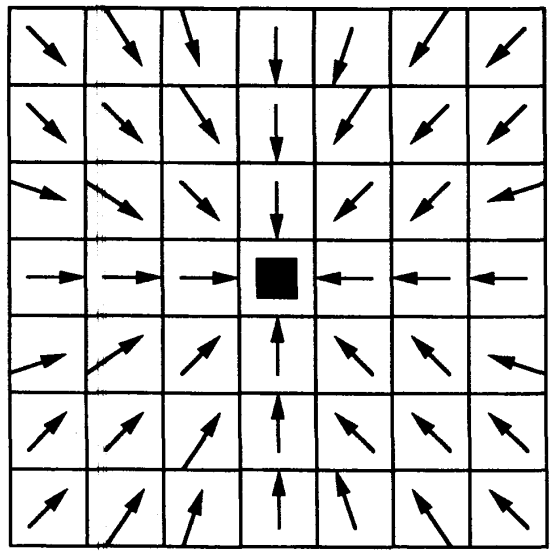

Fig. 2. An example of attractive magnetic field.

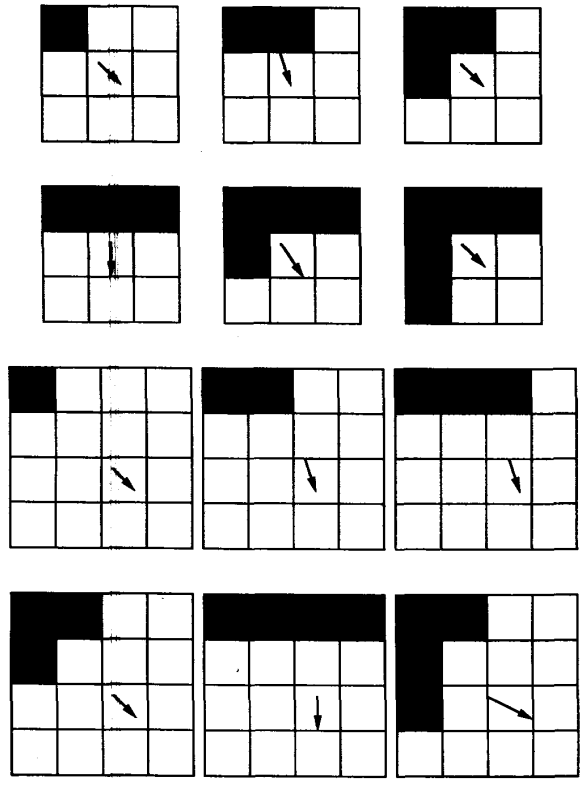

Fig. 3. Examples of repulsive magnetic field.

Boltzmann machine. Our approach is based on building three kinds of (artificial) magnetic fields and a continuity constraint. The continuity constraint assumes that the moving directions of nearby grid cells should be similar to assure the smooth motion of the mobile robot. In a physical system, this is also the case for the magnetic lines

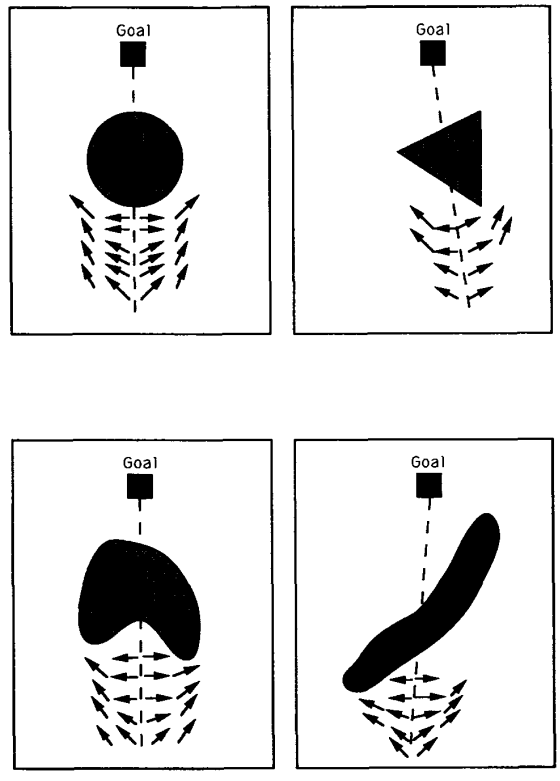

Fig. 4. Examples of breakline-repulsive magnetic field.

which are always continuous in the magnetic field. This constraint forms the internal magnetic field produced by surrounding spins to create a tendency to align all the spins in a regular way and corresponds to the exchange interaction, E_{exchange}, in (1) or (6). For a given goal location, an attractive artificial magnetic field is built to "attract" the mobile robot to this position. An example of the attractive magnetic field is shown in Fig. 2. This can be viewed as the external attractive magnetic field $g-$field in (6). Similarly, on the boundaries of each obstacle, a repulsive magnetic field, $R_{o\text{-field}}$, is built to repulse the mobile robot away from the obstacle. Some examples of such repulsive magnetic field are shown in Fig. 3. Each pattern in Fig. 3 is a "window" or "mask" we used to decide the repulsive field on the center grid cell of the window, where the center grid cell is the cell with an arrow on it in each window. We assume that the resolution of the digital traverse map is not so high to result in "one-cell-width" obstacle or free gap. The windows in Fig. 3 are only for illustration and not intended to be exhaustive. It is obvious that such kind of repulsive magnetic field can be built around any irregular-shaped obstacle. We also provide a breakline-repulsive magnetic field, $R_{b\text{-field}}$, for each obstacle to simulate the breakline situation of the vortex patterns in fluid flow such that a near optimal path (i.e., better direction) to surround the obstacle and reach the goal can always be found for the grid cells surrounding the obstacle. Some examples of the breakline-repulsive magnetic field are shown in Fig. 4. A good chosen breakline-repulsive magnetic field of an obstacle quite depends on the relative positions of the goal and the center of area of the obstacle. This choice is very simple for regular-shaped objects as shown in Fig. 4. The Hamiltonian for this

formulation is

$$\begin{aligned} E_{\text{total}} &= E_{\text{exchange}} + E_{\text{field}} \\ &= E_{\text{exchange}} + E_{\boldsymbol{g}\text{-field}} + E_{\boldsymbol{r}\text{-field}} \\ &= -J_s \sum_{\langle ij,kl \rangle} \boldsymbol{s}_{ij} \cdot \boldsymbol{s}_{kl} - J_{\boldsymbol{g}} \sum_{ij} \boldsymbol{g}_{ij} \cdot \boldsymbol{s}_{ij} \\ &\quad - J_{\boldsymbol{r}} \sum_{ij} \boldsymbol{r}_{ij} \cdot \boldsymbol{s}_{ij} \end{aligned}$$

where $\boldsymbol{r}_{ij} = R_{o-ij} + R_{b-ij}$, and we simply set $J_{ij,kl} = J_s =$ constant. The final result we will get is the magnetic field representation of traverse terrain, or called "magnetic field map," in which the black grids represent the goal and the obstacles, and the arrows on the other grids indicate the proper directions for the mobile robot to move into. Applying a multi-valued Boltzmann machine to the navigation problem, the neighborhood of a grid cell must be first defined. For example, if we use the eight-nearest neighbors, then n is set to 8 in the multi-valued Boltzmann machine. Then, in the ground state, each neuron will point to one of its eight neighbors properly to indicate the next grid for the mobile robot to move into. The magnetic field map obtained in the first simulation is shown in Fig. 5. Here, obstacles with regular and simple shapes are used. In the second simulation, obstacles with more complex shapes are set in the traverse terrain. The resulting magnetic field map is shown in Fig. 6. This simulation illustrates the abilities of the magnetic field approach to deal with very irregular-shaped and concave-shaped objects, which are difficult to deal with for the potential field methods. Fig. 7 shows the simulation result of applying the magnetic field approach to a more realistic planning problem. Here, a higher resolution is used on the digital traverse map and each arrow on this map represents the combined effect of the final outputs of a group of neurons (grid cells). The traverse map used is adopted from the cross-country experiments performed in [14]. The basic mission objective is for the vehicle to get from one location to another while maintaining radio contact at all times. The obstacles to be avoided include gully and rock outcrop. Moreover, the vehicle cannot run into the RF shadow area where the radio contact will be disconnected. The magnetic field representation in Fig. 7 is similar to the gradient field representation shown in [14] which is obtained through precise but time consuming graph search technique.

The simulated annealing method adopted in the multi-valued Boltzmann machine aims to solve the local minimum problem in the corresponding multi-valued Hopfield network. However, the robustness of the proposed method in getting out of the local minimum depends on the selection of the annealing temperatures. Theoretically, slow and smooth annealing procedure (i.e., the sequence of annealing temperatures) can achieve better results, but it will lengthen the convergence time. Hence, there is a tradeoff between the annealing temperatures and the convergence time. Empirical studies showed that a few (about three to eight) discrete decreasing temperatures are enough to obtain satisfactory results [26]. For example, if four decreasing temperatures are used in the annealing process, we can choose these temperatures to be $T = 17.85, 4.46, 1.28, 0.06$, which were suggested in [26] for solving the traveling salesman problem. It is noted that the proposed multi-valued Boltzmann machine is a generalization of the original two-valued Boltzmann machine. Hence, it inherits most of the properties of the original Boltzmann machine on the issues including the convergence, the sensitivity, annealing temperatures, and the ability of getting out of local minima, etc. Ref. [26] provides more detailed and theoretical studies on these important issues.

The multi-valued Boltzmann machine serves as a compass for the mobile robot in its traverse terrain. It suggests to the mobile robot the next direction to move in at any cell of traverse terrain. As compared to most previous work, especially the potential field methods, one significant advantage of the proposed magnetic field approach is its abilities to handle very irregular and concave-shaped obstacles. This is illustrated in Figs. 6 and 7. For example, consider the traverse terrain shown in Fig. 6. If an artificial potential field is used to cover the H-shaped obstacle in the figure, the upper and lower concave regions around this obstacle will become *forbidden regions*. That means the mobile robot is forbidden from entering these regions and thus it cannot start its traverse from any point in these regions. However, like the potential field method, the

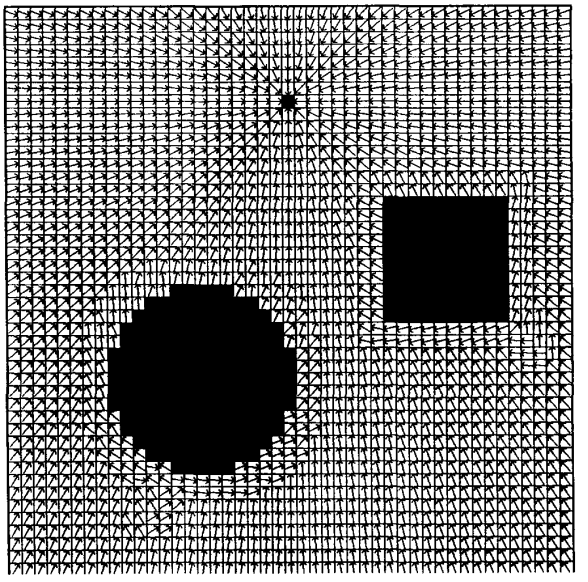

Fig. 5. Artificial magnetic field representation of transverse terrain for mobile robot (i).

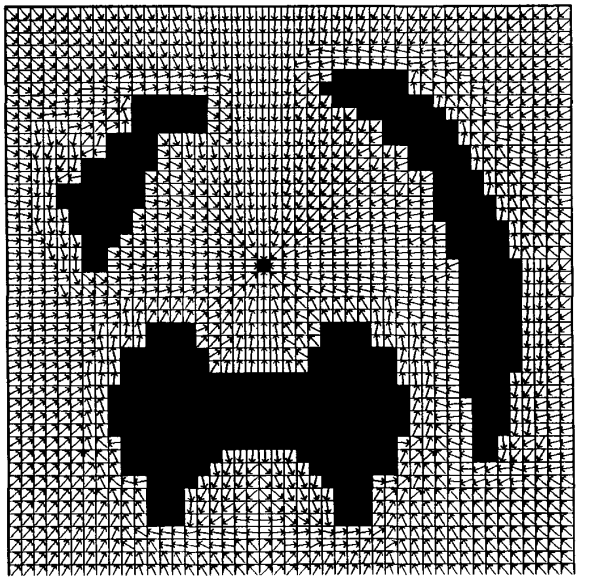

Fig. 6. Artificial magnetic field representation of transverse terrain for mobile robot (ii).

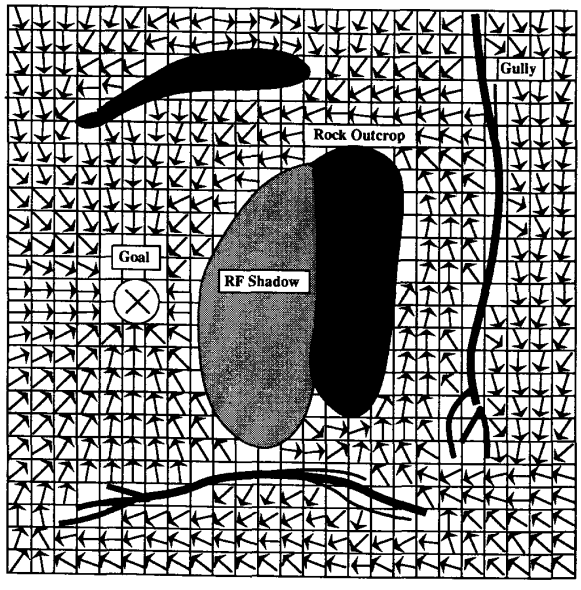

Fig. 7. Artificial magnetic field representation of transverse terrain for mobile robot (iii).

proposed method does not guarantee time-optimal solutions. Parallel to the potential functions to the potential field method, the breakline-repulsive magnetic fileds are the key components to the magnetic field method. Proper choice of these fields are important to the performance of the proposed approach. The breakline-repulsive magnetic fileds used in this paper (see Figs. 3 and 4) are chosen subjectively in an ad-hoc manner. A systematic method to choosing proper and efficient magnetic fileds for the multi-valued Boltzmann machine is an on-going research topic.

The proposed magnetic field approach can be considered as a compatible co-process with higher-level path planning. Complex path planning should be performed in parallel with the magnetic field approach, providing intermediate goals in the event that a solution cannot be obtained by local techniques. In addition to the differences between the magnetic field and the potential field approaches mentioned above, another significant distinction between them is how the resulting information is used. In the potential field approach, the final resulting potential field is used directly to compute the desired motion, while in our approach, the magnetic field representation of a traverse terrain is never used to provide direct control of the vehicle. Instead, they are only an additional source of information provided to a set of real-time decisionmaking processes. So, all features of the environment are kept for intelligent decisionmaking and are not abstracted into a single representational framework.

V. Conclusion

In this paper, two analog connectionist structures are designed to transform the XY spin glass model to computational models for practical applications. These structures are the generalization of the original Hopfield network which is based on the concept of Ising spin glass model. Due to the inevitable local minimum problem of these structures when applied to optimization problems, a multi-valued Boltzmann machine is developed based on the discrete planar XY spin glass model and the stochastic simulated annealing method. This multi-valued Boltzmann machine generalizes the original two-valued Boltzmann machine to a multi-valued machine. To utilize the XY

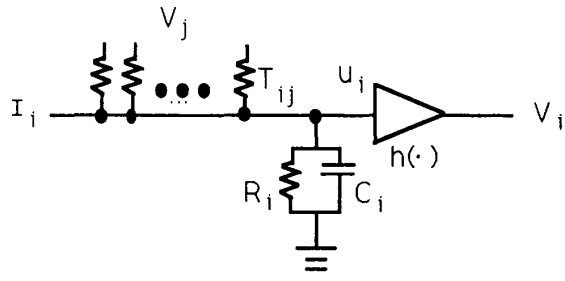

Fig. 8. Elementary analog circuit in multi-layered Hopfield network.

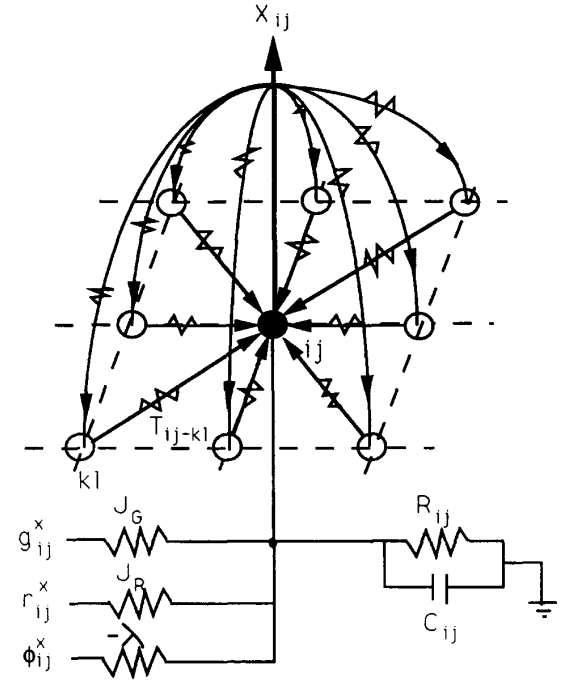

Fig. 9. Node connections within layers 1 and 3 in multi-layered Hopfield network.

spin glass model as a computational machine, a general Hamiltonian (energy function) is defined. This Hamiltonian is minimized by the evolving rules associated with the proposed analog connectionist structures and the multi-valued Boltzmann machine. The proposed multi-valued Boltzmann machine has been applied to the mobile robot navigation problem using an artificial magnetic field approach. Computer simulations were conducted and they verified the validity of the proposed system. The proposed solution to the mobile robot navigation problem has been shown to possess several advantages over the traditional graph search and artificial potential field approaches. Other possible applications of the proposed multi-valued Boltzmann machine can be found in combinatorial optimization problems and fuzzy content addressable memory (fuzzy associative memory). These will be studied closely in the future.

VI. Appendix

To avoid vector addition and trigonometric function computations in the single-layered connectionist structure proposed in Section II, a multi-layered Hopfield analog network is designed in this Appendix

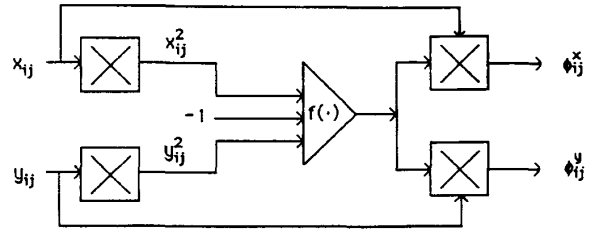

Fig. 10. Analog circuit of layer 2 in multi-layered Hopfield network.

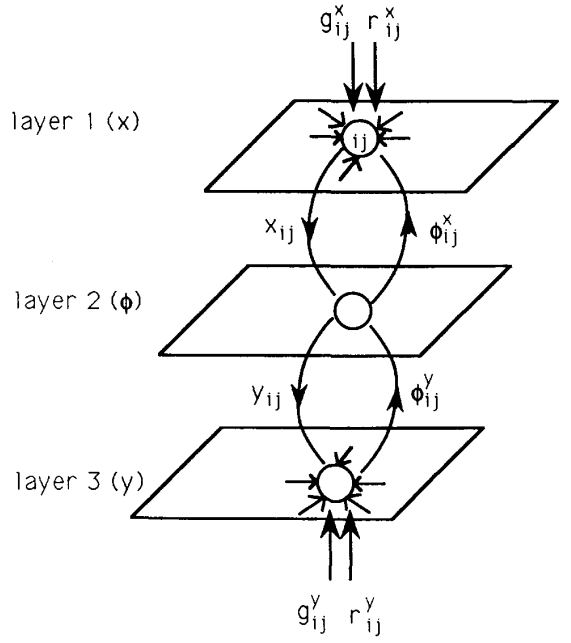

Fig. 11. Overall structure (connections between layers) of multi-layered Hopfield network.

based on Cartesian (X-Y) coordinates instead of polar coordinates. Let $\boldsymbol{s}_{ij} = (x_{ij}, y_{ij}), \boldsymbol{s}_{kl} = (x_{kl}, y_{kl}), \boldsymbol{g}_{ij} = (g_{ij}^x, g_{ij}^y), \boldsymbol{r}_{ij} = (r_{ij}^x, r_{ij}^y)$, where $x_{ij}^2 + y_{ij}^2 = x_{kl}^2 + y_{kl}^2 = 1$. Then (6) becomes

$$\begin{aligned} E = &-\sum_{ij}\sum_{kl\in N(ij)} J_{ij,kl}(x_{ij}x_{kl} + y_{ij}y_{kl}) \\ &- J_{\boldsymbol{g}}\sum_{ij}(g_{ij}^x x_{ij} + g_{ij}^y y_{ij}) - J_{\boldsymbol{r}}\sum_{ij}(r_{ij}^x x_{ij} + r_{ij}^y y_{ij}) \\ &+ \lambda\sum_{ij} F(x_{ij}^2 + y_{ij}^2 - 1) \end{aligned} \tag{A.1}$$

with $dF(z)/dz = f(z)$ and $f(z) = |z|$. The Lagrange multiplier (i.e., the last term) has been added to satisfy the unit length constraint. Let

$$E' = E + \sum_{ij}\frac{1}{R}\int_0^{x_{ij}} h^{-1}(x)\,dx + \sum_{ij}\frac{1}{R}\int_0^{y_{ij}} h^{-1}(y)\,dy. \tag{A.2}$$

where $h^{-1}(\cdot)$ is a monotone increasing function (e.g., a sigmoid function). Using the gradient descent method with respect to x and y coordinates, the rate of change of the Hamiltonian can be obtained as

$$\frac{dE}{dt} = \sum_{ij}\frac{\partial E}{\partial x_{ij}}\frac{dx_{ij}}{dt} \quad \text{and} \quad \frac{dE}{dt} = \sum_{ij}\frac{\partial E}{\partial y_{ij}}\frac{dy_{ij}}{dt} \tag{A.3}$$

where

$$\begin{aligned} \frac{\partial E}{\partial x_{ij}} = &-\sum_{kl\in N(ij)} J_{ij,kl}x_{kl} - J\boldsymbol{g}g_{ij}^x - J_{\boldsymbol{r}}r_{ij}^x \\ &+ \lambda f(x_{ij}^2 + y_{ij}^2 - 1)x_{ij} + \frac{x_{ij}}{R} \\ \frac{\partial E}{\partial y_{ij}} = &-\sum_{kl\in N(ij)} J_{ij,kl}y_{kl} - J\boldsymbol{g}g_{ij}^y - J_{\boldsymbol{r}}r_{ij}^y \\ &+ \lambda f(x_{ij}^2 + y_{ij}^2 - 1)y_{ij} + \frac{y_{ij}}{R}. \end{aligned} \tag{A.4}$$

In order to make the Hamiltonian E decreases (i.e., $dE/dt \le 0$), let

$$C_{ij}\frac{dx_{ij}}{dt} = -\frac{\partial E}{\partial x_{ij}} \quad \text{and} \quad C_{ij}\frac{dy_{ij}}{dt} = -\frac{\partial E}{\partial y_{ij}}, \tag{A.5}$$

where $C_{ij} > 0$. A three-layered, Hopfield-typed, analog circuit is proposed for performing the computation of (A.5). First, considering the circuit in Fig. 8, the function performed by this circuit is

$$C_i\frac{dV_i}{dt} = \sum_j T_{ij}V_j - \frac{V_i}{R_i} + I_i. \tag{A.6}$$

Here we let $h(\cdot)$ be an identity function, since in the energy function ((A.2)), x_{ij} and y_{ij} are already limited by the constraint that $x_{ij}^2 + y_{ij}^2 = 1$. Since $h(\cdot)$ represents the input-output characteristic of an amplifier with negligible response time in an analog circuit, $h(\cdot)$ is close to an identity function when steepness of the response is adjusted to be very small (that is, when $\lambda \ll 1$ in [7]). Using this simple circuit as the basic component, a three-layered, 2-D Hopfield network is designed as shown in Figs. 9 to 11. Layers 1 and 3 are used to generate the outputs x_{ij}'s and y_{ij}'s, respectively. Layer 2 generates the feedback constraints $\phi_{ij}^x (= f(x_{ij}^2 + y_{ij}^2 - 1)x_{ij})$ and $\phi_{ij}^y (= f(x_{ij}^2 + y_{ij}^2 - 1)y_{ij})$. The connections within layer 1 are shown in Fig. 9. Assuming the response time of the constraint neurons (in layer 2) is negligible as compared to the x, y neurons, we have from (A.6),

$$\begin{aligned} C\frac{dx_{ij}}{dt} = &\sum_{kl\in N(ij)} J_{ij,kl}x_{kl} + J\boldsymbol{g}g_{ij}^x + J_{\boldsymbol{r}}r_{ij}^x \\ &- \lambda\phi_{ij}^x + \frac{x_{ij}}{R}, \end{aligned} \tag{A.7}$$

which is exactly what we want in the first equation of (A.5).

Connections within layer 3 are identical to those in layer 1 and they can perform the evolving rule of the second equation in (A.5). Connections in layer 2 are shown in Fig. 10 in which analog multipliers are required to compute ϕ_{ij}^x and ϕ_{ij}^y. The overall structure is shown in Fig. 11 which indicates the connections between different layers. The components in the above structure are all elementary arithmetic analog unit and are all achievable by current analog VLSI techniques [28].

References

[1] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *J. Phys. F: Metal Phys.*, vol. 5, pp. 965–974, 1975.
[2] S. Kirkpatrick and D. Sherrington, "Infinite-ranged models of spin-glasses," *Phys. Rev.*, vol. 17, pp. 4384–4403, 1978.
[3] F. Tanaka and S. F. Edwards, "Analytic theory of the ground state properties of a spin glass: I. Ising spin glass," *J. Phys. F: Metal Phys.*, vol. 10, pp. 2769–2778, 1980.
[4] S. F. Edwards and P. W. Anderson, "Theory of spin glasses: II," *J. Phys. F: Metal Phys.*, vol. 6, pp. 1927–1937, 1976.
[5] F. Tanaka and S. F. Edwards, "Analytic theory of the ground state properties of a spin glass: II. XY spin glass," *J. Phys. F: Metal Phys.*, vol. 10, pp. 2779–2792, 1980.
[6] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Natl. Acad. Sci. USA*, vol. 79, pp. 2554–2558, 1982.

[7] ——, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proc. Natl. Acad. Sci. USA*, vol. 81, pp. 3088–3092, 1984.
[8] J. J. Hopfield and D. W. Tank, "Neural computation of decisions in optimization problems," *Biol. Cybern.*, vol. 52, pp. 141–152, 1985.
[9] G. E. Hinton and T. J. Sejnowski, "Optimal perceptual inference," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Washington DC, pp. 448–453, 1983.
[10] R. A. Brooks, "Solving the find-path problem by good representation of free space," *IEEE Trans. Syst. Man Cyber.*, vol. 13, no. 3, pp. 190–197, 1983.
[11] T. Lozano-Perez, "Spatial planning: a configuration space approach," *IEEE Trans. Comput.*, vol. 32, no. 2, pp. 108–120, 1983.
[12] J. L. Crowley, "Navigation for an intelligent mobile robot," *IEEE J. Robotics Automat.*, vol. 1, no. 1, pp. 31–41, 1985.
[13] J. S. B. Mitchell, D. W. Payton, and D. M. Keirsey, "Planning and reasoning for autonomous vehicle control," *Int. J. Intell. Syst.*, vol. 2, pp. 129–198, 1987.
[14] D. W. Payton, J. K. Rosenblatt, and D. M. Keirsey, "Plan guided reaction," *IEEE Trans. Syst. Man Cyber.*, vol. 20, no. 6, pp. 1370–1382, 1990.
[15] O. Khatib, "Real-time obstacle avoidance for manipulators and mobile robots," *Int. J. Robotics Res.*, vol. 5, no. 1, pp. 90–98, 1986.
[16] F. Miyazaki and S. Arimoto, "Sensory feedback based on the artificial potentials for robots," *Proc. 9th IFAC*, Budapest, Hungary, 1984.
[17] J. R. Andrews and N. Hogan, "Impedance control as a framework for implementing obstacle avoidance in a manipulator," in *Control of Manufacturing Processes and Robotic Systems*, David E. Hardt and Wayne J. Book, Eds. New York: ASME, 1983, pp. 243–251.
[18] B. H. Krogh, "A generalized potential field approach to obstacle avoidance control," in *SME Conf. Proc. Robotics Research: The Next Five Years and Beyond*, Bethlehem, PA, Aug. 1984.
[19] R. C. Arkin, "Motor schema-based mobile robot navigation," *The Int. J. Robotics Res.*, pp. 92–112, Aug. 1989.
[20] R. Volpe and P. Khosla, "Manipulator control with superquadric artificial potential functions: theory and experiments," *IEEE Trans. Syst. Man Cyber.*, vol. 20, no. 6, pp. 1423–1436, 1990.
[21] B. H. Krogh and C. E. Thorpe, "Integrated path planning and dynamic steering for autonomous vehicles," *Proc. 1986 IEEE Int. Conf. on Robotics and Automation*, pp. 1664–1669, San Francisco, 1986.
[22] P. Tournassoud, "A strategy for obstacle avoidance and its application to multi-robot systems," *Proc. 1986 IEEE Int. Conf. on Robotics and Automation*, pp. 1224–1229, San Francisco, 1986.
[23] C. W. Warren, "Global path planning using artificial potential fields," *Proc. 1989 IEEE Int. Conf. on Robotics and Automation*, pp. 316–321, 1989.
[24] J. Barraquand and J.-C. Latombe, "A Monte-Carlo algorithm for path planning with many degrees of freedom," *Proc. 1990 IEEE Int. Conf. on Robotics and Automation*, pp. 1712–1717, 1990.
[25] C. Koch, J. Luo, and C. Mead, "Computing motion using analog and binary resistive networks," *IEEE Computer*, pp. 52–63, March 1988.
[26] E. Aarts and J. Korst, *Simulated annealing and Boltzmann machine*, John Wiley & Sons, 1989.
[27] N. J. Nilsson, *Problem Solving Methods in Artificial Intelligence*. New York: McGraw-Hill, 1971.
[28] C. Mead, *Analog VLSI and Neural Systems*, Addison-Wesley, 1989.

Controlling a Computer via Facial Aspect

Philippe Ballard and George C. Stockman

Abstract— **Control of a computer workstation via face position and facial gesturing would be an important advance for people with hand or body disabilities as well as for all users. Steps toward realization of such a system are reported here. A computer system has been developed to track the eyes and the nose of a subject and to compute the direction of the face. Face direction and movement is then used to control the cursor. Test results show that the resulting system is usable, although several improvements are needed.**

I. INTRODUCTION

We demonstrate a mode of communication whereby the user controls the cursor and selects a desired menu item with his/her face. Our primary goal is to enhance both the work and living environment of disabled persons, but we believe that development of a system viable for all persons would have the most impact.

Eye-tracking systems contained in helmets, glasses, or other devices have already been built to aim guns in aircraft and to study the scanning process of reading [18]. In our project, we wanted to demonstrate feasibility of a face tracking device contained within the computer workstation itself, keeping the user completely free of any devices or wires. A secondary goal is to also avoid having to use cosmetics to provide enhanced facial features. Because a long-term goal is to recognize gestures made with the eye and mouth, we want a system which will image the entire face and not just the eyes. An objective benchmark is to provide a choice of symbols from a set of 64 possible choices within a time of 2 seconds. This cycle time would provide a viable communication alternative for many patients of the Artificial Languages Lab at Michigan State University, according to its Director, Dr. John Eulenberg.

Much of the work on faces in computer vision is not directly related to face direction detection, but some of the techniques used in face recognition, such as feature extraction, location of templates and matching algorithms can be applied to face communication systems.

Work by Wagner and Galiana [17] shows how template matching can be used to track eye movements. In another related work, Schmandt, Ackerman and Hindus [13] showed how window navigation tasks usually performed with a mouse can be controlled by voice. Pentland and Mase [11] present a lip reading system that may be used to augment any speech recognition system to get better accuracy in noisy environments. Although lip reading is not directly related to face direction detection, it presents similar constraints for light requirements and camera positioning. Moreover, success at lip reading can yield symbols for communication.

Several other papers describe useful algorithms to locate the face within a noisy background. Govindaraju, Srihari and Sher [8] developed procedures to locate human faces in newspaper photographs. Their approach is based on cost minimization of feature graphs. Turk and Pentland [16] find the face by analyzing frame differencing in motion under the hypothesis that people are constantly moving. In an early paper, Baron [3] shows how to locate the eyes in a face. His approach is based on the standardization of the image size. Yuille,

Manuscript received June 11, 1993; revised June 3, 1994.
The authors are with the Pattern Recognition & Image Processing Laboratory, Department of Computer Science, Michigan State University, East Lansing, MI 48824 USA.
IEEE Log Number 9406646.