STEREO CORRESPONDENCE USING THE HOPFIELD NEURAL NETWORK OF A NEW ENERGY FUNCTION

JUN JAE LEE, JAE CHANG SHIM and YEONG HO HA†
Department of Electronic Engineering, Kyungpook National University, Taegu 702-701, Korea

(Received 8 September 1993; in revised form 31 March 1994; received for publication 20 April 1994)

Abstract — This paper presents an approach using a Hopfield neural network to the stereo correspondence problem for extracting the 3D structure of a scene. The stereo correspondence problem can be defined in terms of finding a disparity map that satisfies three competing constraints: similarity, smoothness and uniqueness. In order to solve the stereo correspondence problem using a Hopfield neural network, these constraints are transformed into the form of an energy function, whose minimum value corresponds to the best solution of the problem, on the Hopfield network. In the process of mapping the constraints into energy function, the energy functions are derived so that the network ensures Hopfield's convergence rule. Stereo correspondence is then carried out through the network evolving energy surface to find the minimum energy corresponding to the solution of the problem. The examples for random-dot stereograms and real images are shown in the experiment, illustrating how the proposed network works.

Stereo correspondence—Similarity—Smoothness—Uniqueness—Hopfield neural network

I. INTRODUCTION

The primary goal of stereo vision is to obtain 3D depth information of objects from two images of intensity values taken from different viewpoints. The key step in the stereo vision is the correspondence process which finds the corresponding points between the left and right images. Correspondence is, however, itself difficult and computationally expensive.¹,²

There are two conventional methods for stereo correspondence: region-based and feature-based methods according to the nature of correspondence primitives. Region-based methods use correlation among intensity patterns in the local neighborhood of a pixel in one image with intensity patterns in the corresponding neighborhood of a pixel in the other image.³,⁴ These methods obtain a dense array of disparity but are poor in very large smooth area. Feature-based methods use symbolic features such as edges and edge segments. These methods are computationally more efficient as well as better in performance than the region-based methods in natural images, but an interpolation technique is needed to obtain the dense array of disparity.⁵—⁸

In the correspondence process based on the above methods, relaxation processes are generally used to find the best matches.⁵,⁶,⁷,⁹ In the relaxation process, the intention is to obtain an assignment of each point or region to a value corresponding to disparity in a manner consistent with certain predefined constraints. The process is parallel and cooperative in nature because the hypotheses in such a process comprise a parallel and cooperative computation mechanism. There are two types of relaxation processes for stereo correspondence: probabilistic-based and optimization-based processes. In the probabilistic-based process,⁷,¹⁰ the initial probabilities computed from similarity in the feature values surrounding the match points are updated iteratively to impose global consistency. These iterative procedures are continued until either the probabilities reach a steady state or a certain termination condition is satisfied. In each iteration of relaxation process, the probability value must be adjusted according to current probability value and neighbor information, and a normalization process is also needed. In the optimization-based processes,⁴,¹¹ stereo correspondence is carried out by a minimizing energy function where the energy function is formulated from the constraints. It represents a mechanism for the propagation of constraints among neighboring match-points for the removal of ambiguity of multiple stereo matches in an iterative manner. The optimal solution is ground state, that is, the state (or states) of the lowest energy. One such technique is Barnard's stochastic approach, called simulated annealing, which is time consuming, making it impractical.

As the Hopfield network turns out to be appropriate to perform the relaxation process because this network tends to reach a stable state by minimizing energy function when the network evolves,¹²—¹⁴ many researchers have recently been tried to use this model for vision applications such as stereo correspondence, image restoration, and graph partition.¹¹,¹⁴—¹⁹ The stereo correspondence can be viewed as an optimization problem that satisfies three competing constraints: (1)
contour, and the inhibitory connection for uniqueness exists only among candidate neurons corresponding to edge points. Furthermore, the network is computationally very efficient if extended in analog form.

REFERENCES


About the Author—JUN JAE LEE received the B.S. and M.S. in electronic engineering from Kyungpook National University, Taegu, Korea, in 1986 and 1991, respectively, and is currently a Ph.D. student in the Department of Electronic Engineering Kyungpook National University. He worked for Kyungpook National University as a research assistant from September 1991 to July 1993. His main interests are in image processing, computer vision and neural networks.

About the Author—JAE CHANG SHIM received the B.S., M.S. and Ph.D. degrees in electronic engineering from Kyungpook National University, Taegu, Korea, in 1987 and 1990, and 1993, respectively. He worked for Seoul National University as a research assistant from December 1989 to September 1990, and joined IBM T. J. Watson research center as a visiting scientist under a joint study program of Korea Research Foundation from January to April 1993. In March 1994, he joined the Department of Computer Engineering of Andong National University as a full-time instructor. His main interests are in image processing and computer vision.

About the Author—YEOH HO HA received the B.S. and M.S. degrees in electronic engineering from Kyungpook National University, Taegu, Korea, in 1976 and 1978, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Texas at Austin, Austin, Texas, 1985. In March 1986 he joined the Department of Electronic Engineering of Kyungpook National University, as an Assistant Professor, and is currently an Associate Professor. His main interests are in digital signal processing, image processing and computer vision.
images of amplitude bias resulting from different viewpoints. To eliminate amplitude bias, the normalization technique can be used, and to reduce noise and recover a smooth and continuous surface, median filtering and a polynomial fitting technique can be applied, respectively.

4. CONCLUSION

An approach to stereo correspondence using the Hopfield neural network is presented. To solve the stereo correspondence problem using the Hopfield neural network, we mapped the three constraints proposed by Marr and Poggio into the energy function of Hopfield's network. We formulated a neural algorithm by a region-based method in 3D form, to improve and enhance Zhou and Chellappa's method. In their method, two of the three constraints were mapped to energy function but the other constraint, which is the uniqueness constraint, remained unmapped. It also has the disadvantage of no convergence, because it does not satisfy Hopfield's convergence rule, due to self-feedback interconnection strength resulting from the formulation of the smoothness constraint. Therefore we approached with three constraints for correspondence and Hopfield's convergence rule. To do this, we first formulated an energy function to satisfy the similarity constraint; and second, formulated an energy function to satisfy the smoothness constraint by setting the output state of the inactive neuron as $-1$ instead of 0 to remove the self-feedback interconnection strength; and third, derived an energy function to satisfy the uniqueness constraint under $(1, -1)$ states.

To avoid ambiguous matches, our network can be also applied to feature-based methods to obtain correct matches among candidates corresponding to the extracted features, where the excitatory connection for smoothness exists only along neurons corresponding to the edge...
maximum disparity, \( D \), is 4. Therefore the number of neurons corresponding to candidate matches is 5 at a point \((i, j)\) of the image. A cross-correlation measure, sum of square differences of pixel intensity within local window was used as similarity measure. The final state depends on the initial state. Since the threshold \( \theta_{i,j,k} \) is selected so that the least of \( \{ \sum w(i,j) - I_k(i,j+k) \}^2 \) for all \( k \) becomes negative and the others positive, only one neuron is active among the candidate neurons. In the sparse random-dot stereogram, all possible match points including false targets within the prescribed disparity range initially exist as more than one when \( \theta_{i,j,k} \) is given as 0 for all \( i, j, k \). Figure 3 shows how the network gradually organizes itself into correct solution as iteration increases, where black elements represent the disparity at image points. More than one neuron is initially active among neurons corresponding to the image point \((i, j)\). However, the interaction between neurons using inhibitory strength forces each neuron that represents a mismatch to become inactive. We used \( A = 1, B = 5, C = 20, \theta = 0, 3 \times 3 \) and \( 5 \times 5 \) windows for \( W \) and \( S \), respectively. A dense array of disparity is generated with only 11 iterations in Fig. 3(c). Errors are mostly restricted to the regions of disparity discontinuity. The disparity field is still good. Figure 4(c) is the resulting disparity map of Fig. 2 after 14 iterations. The disparity values are plotted in 3D form. We used \( A = 1, B = 5, C = 20, \theta = 0, 3 \times 3 \) and \( 5 \times 5 \) windows for \( W \) and \( S \), respectively.

With natural images, the surfaces of the scene are complex. It is usually very difficult to obtain the surface accurately. Experiments have been conducted on the Pentagon image which has a relatively large amount of information on local structure, and the bear image with little local structure. Figure 5 shows an aerial view of the Pentagon in Washington, DC. The image is \( 256 \times 256 \), obtain by decimating the original image of \( 512 \times 512 \times 8 \) bits and the disparity ranges from \(-5\) to \(4\) pixels. Figure 6 shows a stereo pair obtained by translating CCD camera in the laboratory. The image is \( 200 \times 200 \times 8 \) bits, and the disparity is \( 0 \) for background, \(3~4\) for ball, and \(7~8\) for bear. Figure 7 shows a 3D plot of the disparity surface for natural images. With the Pentagon image, we used \( A = 1, B = 5, C = 40, 5 \times 5 \) and \( 7 \times 7 \) windows for \( W \) and \( S \), respectively, and with the bear image, \( A = 1, B = 5, C = 50, 7 \times 7 \) and \( 5 \times 5 \) windows are used. The window size for smoothness in the bear image is larger than in the Pentagon because the former has more smooth disparity than the latter. \( \theta \) is set to a little higher than the minimum value among similarity terms corresponding to candidates at point \((i, j)\). Therefore, more than one neuron is initially active among the neurons corresponding to the image point \((i, j)\), and the initial state and its consecutive states are truly different from that shown in Fig. 7. If \( \theta \) is larger, many neurons are active. From the experimental results, the performance is better when the number of active neurons in the initial state is greater than in the final state. There are trade-offs between window sizes for excitatory connections. The larger the window size, the more isolating points are removed, but blurring occurs. It is not easy to determine the constants \( A, B \) and \( C \). If stereo images are noise-free, i.e. ideal images, the constants \( B \) and \( C \) have no meaning and the corresponding result is dependent on only the similarity term. If they have a lot of noise, the constant \( B \) is relatively larger, and \( C \) is also larger to satisfy uniqueness.

The experimental results for real images show some mismatches due to the similarity measure. For this reason, preprocessing is needed to cope with images of a wide variety of scenes, images corrupted by a certain amount of noise due to the electronic imaging sensor, film granularity, and quantization error, and stereo
Fig. 5. Pentagon images. (a) Left image. (b) Right image.
Fig. 3. Correspondence results on each layer according to iteration. The black elements represent the disparity at each image point. Thus the black pixels in the left-most image mean disparity of zero, and going to the right, the disparities are 1, 2, 3 and 4. (a) Initial state, (b) iteration 1, (c) iteration 11.

Fig. 4. Three-dimensional plot of the disparity map for decorrelated random-dot stereogram.
where \( \delta_{ij} \) is the Dirac delta function. From (15), the similarity, smoothness and uniqueness constraints are mapped to part of input, the excitatory connection to neighboring neurons, and the inhibitory connection to both candidate neurons corresponding to disparity values and the part of input, respectively.

2.4. Stereo correspondence

The correspondence process is achieved through calculating the net input which provides excitatory and inhibitory support, and input. Each neuron \((i, j, k)\) randomly and asynchronously receives inputs through interconnection strength from neurons and bias input. The output of the neuron is then calculated by threshold, and the solution is not taken until the network settles into a steady state where the neuronal activities remain constant. The initial state is set only with input bias corresponding to similarity constraint. Due to the occlusion, noise and other distortions, a number of neurons will become active, but the interaction between neurons causes them to enforce the uniqueness constraint as an iteration increases.

The algorithm can be summarized:

**Step 1.** Set the initial state of neurons by (3) which is substituted \( u_{i,j,k} \) for \( I_{i,j,k} \).

**Step 2.** Randomly pick up a neuron \((i,j,k)\).

**Step 3.** Calculate its net input by (2).

**Step 4.** Decide the new state of each neuron according to step 1.

**Step 5.** Check the changes of states after searching all neurons. If they have changed, go to the step 2.

**Step 6.** Output the final state.

3. EXPERIMENTAL RESULTS AND DISCUSSION

We have tested the proposed network for random-dot stereograms and natural stereo images. Experiments have been performed on 10\% and 50\% density random-dot stereograms. Figure 1 shows a four-level wedding cake of a 128 \times 128 random-dot stereogram composed of 10\% white and 90\% black pixels. Figure 2 shows the decorrelated stereogram. The original stereogram is 50\% density random dots. In the right image, 20\% of the dots were decorrelated at random. Cross-eye fusion by the human visual system can immediately bring this structure under examination, despite the patterns with sparse features and noise between left and right eye images. These examples are designed to test the capability of the proposed network even in the worse situations. Intensity values of the white and black pixels in the figures are 255 and 0, respectively. The background has zero disparity, and each successive layer has a disparity of 1, 2 and 3 pixels. The allowable

Fig. 1. A 10\% density random-dot stereogram. (a) Left image, (b) right image.

Fig. 2. A 50\% density random-dot stereogram. In the right image, 20\% of dots were decorrelated at random. (a) Left image, (b) right image.
Table 1. Energy as a function of $x$ according to $D$

<table>
<thead>
<tr>
<th>$D \times x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>3</td>
<td>-2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>-3</td>
<td>-1</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>-2</td>
<td>-4</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>1</td>
<td>-5</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>24</td>
<td>6</td>
<td>-4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>35</td>
<td>13</td>
<td>-1</td>
<td>2</td>
<td>-5</td>
</tr>
<tr>
<td>7</td>
<td>48</td>
<td>22</td>
<td>4</td>
<td>-8</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>63</td>
<td>33</td>
<td>11</td>
<td>-3</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>80</td>
<td>46</td>
<td>20</td>
<td>-8</td>
<td>-10</td>
</tr>
<tr>
<td>10</td>
<td>99</td>
<td>61</td>
<td>31</td>
<td>-5</td>
<td>-11</td>
</tr>
</tbody>
</table>

Minimum energy state

The uniqueness term must be connected by inhibitory connection strength, i.e. the first term must have its minimum value when $V_i$ is 1, and each of all candidate neurons $V_j$ excluding $k$ is $-1$. However, the energy of the first term has a minimum value in the case that the number of active neurons and inactive neurons is the same. Therefore, another term, i.e. the second one, enforcing negative energy, which is contributed by input bias, is needed to add to this term. Even though it does so, (8) always is not minimum when only the output of one neuron is 1. To show which state is the minimum energy according to variable $D$, let $x$ be the number of neurons whose output is 1. The minimum energy state (region in the box) then is changed at a constant rate according to $D$ as shown in Table 1.

If the ratio of the first term to the second one is adjusted appropriately, (8) can be reformulated so that the function has its minimum value only when $x$ is 1 without regard to $D$. Let the energy function $E$ be in the form:

$$E = \sum_{k=0}^{D} v_k \sum_{x=k}^{D} v_x + \beta \sum_{k=0}^{D} v_k$$

Equation (9) can be rewritten by the function of $x$.

$$E(x) = 2x(2x + 1 - x) + \beta x$$

Arranging (10) in descending order gives

$$E(x) = 2x^2 - (4\beta + 4x)x + \beta$$

By differentiating (11) and substituting $x$ for 1 to have a minimum value when $x$ is 1, we obtain

$$\beta = 2\alpha(D + 1)$$

Table 2. Reconstructed energy when $x$ is 1

<table>
<thead>
<tr>
<th>$D \times x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>-4</td>
<td>0</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-4</td>
<td>-8</td>
<td>-4</td>
<td>8</td>
<td>-28</td>
</tr>
<tr>
<td>4</td>
<td>-10</td>
<td>-14</td>
<td>-10</td>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>5</td>
<td>-18</td>
<td>-22</td>
<td>-18</td>
<td>-6</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>-28</td>
<td>-32</td>
<td>-28</td>
<td>-16</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>-40</td>
<td>-44</td>
<td>-40</td>
<td>-28</td>
<td>-8</td>
</tr>
<tr>
<td>8</td>
<td>-54</td>
<td>-58</td>
<td>-54</td>
<td>-42</td>
<td>-22</td>
</tr>
<tr>
<td>9</td>
<td>-70</td>
<td>-74</td>
<td>-70</td>
<td>-58</td>
<td>-38</td>
</tr>
<tr>
<td>10</td>
<td>-88</td>
<td>-92</td>
<td>-88</td>
<td>-76</td>
<td>-56</td>
</tr>
</tbody>
</table>

Minimum energy state

Equation (9) satisfies the uniqueness constraint as long as $\beta = 2\alpha(D + 1)$, and Table 2 shows the energy value calculated when $x = 1$.

Let $x$ be 1, and extending (8) over the entire region gives

$$E = \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \left\{ \sum_{k=0}^{D} v_{i,j,k} \sum_{x=k}^{D} v_{i,j,n} + 2(D - 1) \sum_{k=0}^{D} v_{i,j,k} \right\}$$

Combining (5), (6), and (13), the total energy function is given as follows:

$$E = A \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \left\{ (f_i(l, j) - f_k(l, j + k))^2 - \theta_{i,j,k} \right\} V_{i,j,k}$$

$$- B \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \sum_{k=0}^{D} V_{i,j,k}$$

$$- C \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \left\{ \sum_{k=0}^{D} V_{i,j,k} \sum_{x=k}^{D} V_{i,j,n} \right\}$$

$$+ 2(D - 1) \sum_{k=0}^{D} V_{i,j,k}$$

where the first term, the second term, and the third term are similarity constraint, smoothness constraint, and uniqueness constraint, respectively. The constants $A$, $B$ and $C$ determine relative importance to achieve the best result.

2.3. Determination of neural parameter

By comparing the terms in (14) with the corresponding terms in (1), we can determine the interconnection strengths and bias inputs as follows:

$$T_{i,j,k,m,n} = 2B \sum_{s \neq s} \delta_{(i,s,j,m,s)} + \delta_{k,n}$$

$$- 2C \sum_{s \neq s} \delta_{i,s} \delta_{j,m}(1 - \delta_{k,n})$$

$$I_{i,j,k} = -A \{(f_i(l, j) - f_k(l, j + k))^2 - \theta_{i,j,k} \}$$

$$- 2C(D + 1)$$

(15)
$D$ is the allowable maximum disparity. The state of each neuron (active or inactive) in the network represents a possible match between an interesting point in the left image and one in the right image.

The energy function of the Hopfield neural network in 3D form is as follows:\(^\text{(1)}\)

$$E = -\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sum_{k=1}^{N_z} \sum_{m=1}^{D} \sum_{n=1}^{D} T_{i,j,k,l,m,n} V_{i,j,k} V_{l,m,n}$$

$$-\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sum_{k=1}^{D} I_{i,j,k} V_{i,j,k}$$

where $V_{i,j,k}$ is the activity of firing rate of $i,j$th neuron, $T_{i,j,k,l,m,n}$ the interaction between two neurons, and $I_{i,j,k}$ input bias to each neuron. The neuron has only two states: fully active, i.e. $V_{i,j,k} = 1$; and inactive, i.e. $V_{i,j,k} = 0$ or $-1$. In our model, we set the inactive state as $-1$. If $V_{i,j,k}$ is 1, this means that disparity value is $k$ at the image point $(i,j)$. The dynamics are discrete in time and asynchronous. Therefore, the net input for neuron $(i,j,k)$ is given by

$$u_{i,j,k} = \sum_{l=1}^{N_x} \sum_{m=1}^{N_y} \sum_{n=1}^{D} T_{i,j,k,l,m,n} V_{l,m,n} + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sum_{k=1}^{D} I_{i,j,k}$$

and the output of the neuron $(i,j,k)$ is determined from

$$V_{i,j,k} = \begin{cases} 1, & \text{if } u_{i,j,k} > 0 \\ -1, & \text{otherwise.} \end{cases}$$

### 2.2. Neural network modeling of energy function

To solve the stereo correspondence problem with the neural network, the neural parameters $T_{i,j,k,l,m,n}$ and $I_{i,j,k}$ should be determined. For any optimization problem using the Hopfield neural network, these parameters embody all characteristics of the problem. Once both $T_{i,j,k,l,m,n}$ and $I_{i,j,k}$ are fixed, and both the energy surface and the local minimum points are completely defined, the network flows into the stable state, where the corresponding energy is minimized. However, for the network to ensure the convergence to stable minima, symmetrical interconnection strength and no self-feedback are required,\(^\text{(2)}\), i.e.

$$T_{i,j,k,l,m,n} = T_{l,m,n,i,j,k}$$

$$T_{i,j,k,l,m,n} = 0.$$  

(4)

Our objective is now to find an energy function for the Hopfield neural network which will increase performance when the relaxation process proceeds, with satisfying the convergence of (4). The energy function creates typically the form of energy surface and determines many minima that represent a valid solution to the problem. Therefore the performance of the system depends on the energy function. In this paper, we approach correspondence with three constraints proposed by Marr and Poggio and the convergence rules of (4).

(1) An energy function satisfying similarity constraint.

In an attempt to satisfy the similarity constraint, we formulate a function whose energy has its minimum value when all regions of two images are matched as close as possible in the least square sense:

$$E = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sum_{k=1}^{D} \left( \left( f_{L}(i,j) - f_{R}(i,j+k) \right)^2 \right) - \theta_{i,j,k} V_{i,j,k}$$

(5)

where $f_{L}$ and $f_{R}$ are the primitives of the left and right images, respectively, for correlation. As primitives, intensity or intensity derivative patterns, etc.,\(^\text{14-71}\) can be used in a small window. Since $(f_{L}(i,j) - f_{R}(i,j+k))^2$ is always positive, we add threshold $\theta_{i,j,k}$ so that the least of $(f_{L}(i,j) - f_{R}(i,j+k))^2 - \theta_{i,j,k}$ for all $k$ becomes negative, and others positive. The energy then has its minimum value if the neuron $V_{i,j,k}$ corresponding to the least square of feature difference between $I_{i}$ and $I_{R}$ is 1, and the others are 0 and $-1$.

(2) An energy function satisfying smoothness constraint.

In an attempt to satisfy the smoothness constraint, the energy function is formulated to have its minimum value in the case that, if the disparity of a point of the image is $k$, then the disparity of its neighboring points is $k$. The energy function is formulated by depending on whether the inactive state of neuron is 0 or $-1$. Therefore, we formulate the energy function for each case and determine the inactive state of neuron which is appropriate to our objectives.

Case 1: active state = 1, inactive state = 0

$$E = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sum_{k=0}^{D} \sum_{S=0}^{N_x \times N_y} \left( V_{i-j,k} - V_{i-j+k} \right)^2$$

(6)

where $S$ is an index set excluding $(0,0)$ for all neighbors in a window centered at point $(i,j)$. This model proposed by Zhou has self-feedback, i.e. $T_{i,j,k,i,j,k} \neq 0$. As a result, the energy function $E$ does not always decrease monotonically with transition,\(^\text{(2)}\)

Case 2: active state = 1, inactive state = $-1$

$$E = -\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sum_{k=0}^{D} \sum_{S=0}^{N_x \times N_y} V_{i-j,k} V_{i-j+k}$$

(7)

Equation (7) has its minimum value if $V_{i,j,k}$ is 1 and $V_{i,j+k}$ are 1, or if $V_{i,j,k}$ is $-1$ and $V_{i,j+k}$ are $-1$. Therefore, the output of neuron is suitable for the second case rather than the first one. The uniqueness constraint will be derived under this assumption in the next section.

(3) An energy function satisfying uniqueness constraint.

The network for stereo correspondence should reflect the uniqueness constraint, i.e. only one neuron should be active (its state is 1) among the hypothesis of disparity value.

To satisfy this constraint, the energy function should be formulated to have its minimum when only one neuron is active: 1 and the rest are off: $-1$. First, we construct a function of form:

$$E = \sum_{k=0}^{D} V_{k} + \sum_{k=0}^{D} V_{k}$$

(8)
similarity, matched points have similar local features; (2) smoothness, disparity values change smoothly, except at a few depth discontinuities; and (3) uniqueness, each point in an image should be assigned at most one disparity value. Based on these constraints, we want to find the best correspondence so that each pixel in the left image matches each pixel in the right image. In order to solve the stereo correspondence problem using the Hopfield neural network, these constraints are transformed into the form of an energy function, whose minimum value corresponds to the best solution of the problem, on the Hopfield network. Then, the matching is achieved by minimizing the cost function (constraint optimization), where all the constraints on the solution can be explicitly included in the cost function. The energy function creates typically the form of energy surface and determines many minima that represent valid solutions to the problem. Therefore the performance of the system depends on the energy.

In previous works, Zhou and Chellappa\cite{11} and Nasrabadi and Choo\cite{19} formulated a neural algorithm in 3D form which is constructed by a region-based method, and in 2D form by a feature-based method, respectively. In Zhou and Chellappa's network, two of the three constraints were mapped to energy function but the other constraint, which is the uniqueness constraint, remained unmapped. Max-net was introduced in their methods to solve this problem. This network, however, has a disadvantage of no convergence because it does not satisfy Hopfield's convergence rule, due to self-feedback interconnection strength resulting from the formulation of the smoothness constraint. In Nasrabadi and Choo's method, the neural network is used to extract the correct matches between extracted features using geometric and uniqueness constraints which are mapped on a 2D binary Hopfield neural network. But the similarity constraint is not mapped. This system works well without similarity constraints in the case of robot vision whose target isolates and identifies different objects. To cope with the industrial process for greater precision, a dense array of disparity is needed. To do this, a similarity constraint should be included.

In this paper, a neural algorithm is developed by a region-based method in 3D form to obtain dense depth information. Here, we intend to improve Zhou and Chellappa's network in two ways. First, the reason of no convergence resulting from formulating the energy function representing smoothness constraint is illustrated and a new energy function is proposed to solve that problem. In Zhou and Chellappa's model, the energy of the system does not always decrease monotonically with transition, due to self-feedback of interconnection strength. To overcome this problem, they used a deterministic decision rule, requiring the check of energy, which is to take a new state only if the energy due to state change is decreased. As a result, it is time consuming and difficult to be implemented in hardware. Instead, the proposed energy function allows the network to flow into a stable state spontaneously, where the energy function corresponding to the solution of the problem is minimized. Second, the improvement of their work is concerned with the uniqueness constraint, which should be included to obtain the best solution. The similarity and smoothness constraints on solution are simply formulated by cost (energy) function. However, the uniqueness constraint is not that simple. To solve this problem ensuring convergence of the network, we derived a new energy function having its minimum value when each point in the left image has only a unique match in the right image. The interconnection strengths and bias input are determined from the formulated energy function. Stereo correspondence then is carried out through the network evolving energy surface to find the minimum energy corresponding to the solution of the problem. Since the energy function of the network is derived with symmetrical interconnection strengths and no self-feedback, the network defined by the proposed energy function convergence to a stable state and also is suitable to implement in hardware. Random-dot stereograms and natural image examples are given to demonstrate the effectiveness of the proposed correspondence network.

### 2. Stereo Correspondence Using the Hopfield Neural Network

The stereo correspondence problem is defined in terms of finding a disparity map that satisfies three competing constraints\cite{12}(1) similarity, matched points have similar local features; (2) smoothness, disparity values change smoothly, except at a few depth discontinuities; and (3) uniqueness, each point in an image should be assigned at most one disparity value. Based on these constraints, we want to find the best correspondence so that each pixel in the left image matches each pixel in the right image. Let us assume that the epipolar lines are parallel to the horizontal lines of image. The correspondence problem then reduces to the assignment of a single horizontal disparity to each pixel in the left image.

To solve the stereo correspondence problem using the Hopfield network, one must decide first the representation scheme which allows the outputs of the neurons to be interpreted as a solution of the problem, define second an energy function whose minimum value corresponds to the best solution of the problem, and finally derive the values of interconnection weights and the input bias from the defined energy function. After setting up input bias values, one can start to evolve the system to get the solution.

#### 2.1. The neural representation scheme for correspondence

For stereo correspondence, the representation scheme consists of \(N_r \times N_c \times (D + 1)\) mutually interconnected binary neurons in 3D form, where \(N_r\) and \(N_c\) are the row and column sizes of the image, respectively, and \(D\) is the maximum disparity. For each neuron \(i\), \(j\), and \(k\) at a possible 3D index location, the neuron represents a left image pixel with three states:

- \(y_{i,j,k} = 0\) represents no disparity.
- \(y_{i,j,k} = 1\) represents a possible disparity.
- \(y_{i,j,k} = 2\) represents a possible disparity.

The energy function of the network is given by:

\[
E = -\sum_{i,j,k} T_{i,j,k} y_{i,j,k} + \sum_{i,j,k} I_{i,j,k} y_{i,j,k}
\]

where \(T_{i,j,k}\) are the interconnection strengths and \(I_{i,j,k}\) are the biases.

#### 2.2. Energy function

The energy function for the Hopfield network and the proposed solution to the stereo correspondence problem is given by:

\[
E = -\sum_{i,j,k} T_{i,j,k} y_{i,j,k} + \sum_{i,j,k} I_{i,j,k} y_{i,j,k}
\]

where \(T_{i,j,k}\) are the interconnection strengths and \(I_{i,j,k}\) are the biases.

The energy function is defined such that the energy is decreased as the network evolves, and the network converges to a state where the energy is minimized. The energy function is also designed such that it ensures the uniqueness constraint, where each pixel in the left image has only a unique match in the right image. The interconnection strengths and bias input are determined from the formulated energy function. Stereo correspondence then is carried out through the network evolving energy surface to find the minimum energy corresponding to the solution of the problem.