Landmark-based Shape Recognition by A Modified Hopfield Neural Network

Nirwan Ansari and Kuowei Li
Center for Communications and Signal Processing
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
University Heights
Newark, New Jersey 07102
USA

Abstract — A new method is introduced to achieve partial shape recognition by means of a modified Hopfield neural network. Given a scene consisting of partially occluded objects, a model object in the scene is hypothesized by matching the landmarks of the model with those in the scene. A local shape measure, known as the sphericity of a triangular transformation, is used as a measure of similarity between two landmarks. The hypothesis of a model object in a scene is completed by matching the model landmarks with the scene landmarks. The matching task is performed by a Modified Hopfield Neural Network. The location of the model in the scene is estimated with a least squares fit among the matched landmarks. A heuristic measure is then computed to decide if the model is in the scene.

I. INTRODUCTION

The problem of recognizing partially occluded objects is of considerable interest in the field of industrial automation, especially in robotic applications where multiple objects, touching objects, or overlapping objects cause many problems in identifying and locating the objects in the workcell of a robot. This problem has intrigued many researchers in the areas of computer vision.

The approach proposed in this paper inherits the merit of the two approaches [1],[2]. That is, it inherits the parallelism of the Hopfield Neural Network [1] and the merit of scale invariance [2]. Similar to [2], we use landmarks as our shape features. In order to characterize the similarity between two features, we use a function known as sphericity [2] to discriminate the dissimilarity between two landmarks. The matching technique is based on Hopfield Neural Network [3]-[5]. The extraction of landmarks and the sphericity will be briefly described in Section 2. In contrast to [1], our approach is not sensitive to scale variation. Our approach to partial shape recognition by means of a modified Hopfield method will be discussed in details in Section 3. Finally, experimental results will be presented in Section 4, along with a discussion on the merits of our approach as compared to [1],[2].

II. LANDMARK EXTRACTION AND SPHERICITY

For the purpose of shape recognition, much of the visual data perceived by a human being is highly redundant. It has been suggested from the viewpoint of the human visual system [6] that some dominant points along an object contour are rich in information content and are sufficient to characterize the shape of the object. These dominant points of an object are called the landmarks of the object.

In this paper, we shall only consider points with high curvatures as landmarks. The landmarks along an object boundary are obtained [7] by successively smoothing the boundary with a Gaussian filter with various widths, w, until the extreme curvature points do not change (their number remains the same and their locations deviate little) for a range of w. Landmarks obtained for a library of objects by this approach are shown in Figure 1.

After extracting the shape features from a model and a scene, some sort of similarity or dissimilarity measure must be used to quantify the difference between the shape features. In this paper, we use the sphericity [8] as the similarity or dissimilarity measure.

Figure 1: The landmarks of a library of objects obtained based on the cardinal curvature points. (a) wire stripper. (b) wrench. (c) specialty plier. (d) needle-nose plier. (e) wire cutter. (f) spacecraft. (g) Island of Borneo. (h) Island of Halmahera. (i) Island of Luzon. (j) Island of Mindanao. (k) Island of New Guinea. (l) Island of Sulawesi.

ISSN# 0-7803-0233-8/91 $1.00©1991 IEEE
The sphericity of a triangular transformation which maps a triangle to another triangle is a measure of similarity between the two triangles. Under the triangular transformation, the inscribed circle of one triangle is mapped onto an inscribed ellipse of the other triangle. The sphericity is defined as the ratio of the geometric mean to the arithmetic mean of the lengths of the principal axes of the inscribed ellipse. If the two triangles are similar, the sphericity is one. The less similar the two triangles are, the smaller is the value of the sphericity. If the vertices of one triangle are considered as the coordinates of three consecutive landmarks belonging to a model, and the vertices of the other triangle as those belonging to a scene, the sphericity is thus a measure of the similarity between the set of three model landmarks and the scene landmarks. The triangular transformation is uniquely defined by an affine transform [9]. It can be shown [8] that the sphericity of an affine transform is invariant when the affine transform undergoes a similarity transformation.

III. THE NEURAL NETWORK APPROACH TO SHAPE RECOGNITION

The hypothesis of a model object in a scene is completed by matching the model landmarks with the scene landmarks. The landmark matching task is performed by the Modified Hopfield Neural Network. The location of the object in the scene is then estimated by a least squares fit among the matched landmarks. A heuristic measure based on the least squared error of the fit is finally used to verify the hypothesis.

A. The Partial Shape Recognition Problem

For partial shape recognition, we only require knowledge of the positions of the landmarks of the object in the image. It is important to impose a consistent ordering of the landmarks and arrange landmarks in a pre-defined order reflecting the shape and geometry of the object. Given a scene consisting of M landmarks, and a model consisting of N landmarks, the hypothesis of this model object in the scene is determined by how well the model landmarks are matched to the scene landmarks. The landmark matching task is thus to find the correspondence between the model landmarks and the scene landmarks. Each correspondence between a model landmark and a scene landmark constitutes a “landmark correspondence pair.” To map the landmark matching task onto the Hopfield Neural Network, we represent each “landmark correspondence pair” by the value of a neuron, i.e., the correspondence pair is established if the neuron has a value of 1. Since there are N model landmarks and M scene landmarks, we need a network with NxM neurons to achieve the landmark matching task. The network is represented by a matrix in which each matrix entry denoted by $V_{x,y}$ represents a neuron. Each neuron takes on binary values 1 or 0. The first and second subscript of $V_{x,y}$ denote the row index and the column index, respectively. Consider a scene with two overlapping object (wire stripper and wrench) as shown in Figure 2. The scene consists of 12 landmarks. To match landmarks of the wire stripper (Figure 1(a)) which consists of 6 landmarks to the scene, we need a network with 6x12 neurons, as shown in Figure 3. The row index corresponds to a model landmark while the column index corresponds to a scene landmark. In this example, shown in Figure 3(c), neurons $V_{2,11}, V_{6,12}, V_{4,11}, V_{4,12}$ and $V_{6,12}$ are turned on "$1's$," indicating that model landmarks 2, 3, 4, 5, and 6 match scene landmark 11, 12, 1, and 2, respectively. Note that, as shown in Figure 3(c), which shows the final (steady) state of the network, each row or column only have at most one neuron turned on. This is due to the physical constraint that one model landmark can match to only one scene landmark, and vice versa.

![Figure 2: A scene which consists of a wire stripper and a wrench overlapping each other. Each landmark is labeled and indicated by an "x".](image)

![Figure 3: Network status during the landmark matching process. (a) initial state. (b) the state after the preprocessing step. (c) the final state.](image)
B. The Neural Network Architecture

To enable the N x M neural network to perform the landmark matching task, the network is described by an energy function in which the lowest energy state corresponds to the "best" set of landmark correspondence pairs. The network must satisfy the following constraints and assumptions:

1. Each model landmark can match no more than one scene landmark.

2. Each scene landmark can match no more than one model landmark.

3. The lowest energy should favor the best set of corresponding pairs.

The modified Hopfield energy function is given by:

\[ E = \sum_{X=1}^{N} \sum_{Y=1}^{M} V_{X,Y} + \sum_{X=1}^{N} \sum_{Y=1}^{M} V_{X,Y} + \sum_{X=1}^{N} \sum_{Y=1}^{M} V_{X,Y} \]

\[ + \left( \sum_{X=1}^{N} \left( - \sum_{X=1}^{N} V_{X,Y} \right)^2 + \sum_{Y=1}^{M} \left( 1 - \sum_{Y=1}^{M} V_{X,Y} \right)^2 \right) \]

\[ - \sum_{X=1}^{N} \sum_{Y=1}^{M} \left( W_{X,Y} F_{1}(V_{X,Y}) + W_{X,Y} F_{2}(S_{X,Y}(X+1),Y+(i+1)) \right) \]  

The first three terms are used to enforce the first two constraints. The fourth term which is used to impose the third condition represents the strength of the interconnection among the neurons. Note that \( W_{1} + W_{2} = 1 \).

\[ F_{1}(V_{X,Y}) = \begin{cases} 1 & \text{if } V_{X,Y} = 1 \\ 0 & \text{if } V_{X,Y} = 0, \text{ and} \end{cases} \]  

\[ F_{2}(S_{X,Y}(X+1),Y+(i+1)) = \begin{cases} 1 & \text{if } S_{X,Y}(X+1),Y+(i+1) \geq \theta \end{cases} \]  

\[ V_{X,Y} \] is the neuron value at the current state. \( V_{X,Y} = 1 \) at the current state implies that at the current state the network hypothesizes that model landmark \( X \) likely matches some scene landmark \( i \). Thus, \( F_{1}(1) = 1 \) favors a lower energy as opposed to \( F_{1}(0) = 1 \). \( S_{X,Y}(X+1),Y+(i+1) \) is the sphericity value derived from the triangular transformation mapping model landmarks \( X,X+1 \) to scene landmarks \( i,j+1 \), respectively. This sphericity value indicates how the neighboring landmarks of the model and the scene "support" the hypothesis that model landmark \( X \) matches scene landmark \( i \). It is like a taly system; all neighboring model and scene landmarks can "vote" for the match between model landmark \( X \) and scene landmark \( i \). If the sphericity value is greater than \( \theta \), there is a supporting evidence that model landmark \( X \) matches scene landmark \( i \). Hence, the function \( F_{2} \) yields a value of 1 favoring a lower energy.

The original Energy Function of Hopfield Neural Network [4] is given below:

\[ E = -1/2 \sum_{i} T_{ij}V_{i} - \sum_{i} I_{i} \]

where \( V_{i} \) indicates the output of neuron \( i \), \( T_{ij} \) indicates interconnection strength from neuron \( j \) to \( i \), and \( I_{i} \) is the external input to neuron \( i \).

Note that Equation (1) can be rewritten into an equation similar to the energy function of the original Hopfield Neural Network. Using double subscripts to index neurons, we can rewrite Equation (4) as:

\[ E = -1/2 \sum_{X,Y} T_{X,Y}V_{X,Y} - \sum_{X} I_{X} \]  

(5)

Let the connection matrix which indicates the interconnection among neurons in the network be:

\[ T_{X,Y} = \delta_{XY} - \delta_{ij} + \left\{ W_{1} F_{1}(V_{X,Y}) + W_{2} F_{2}(Sy_{X,Y}(X+1),Y+(i+1)) \right\} \]

where \( \delta_{XY} \) and \( \delta_{ij} \) are Kronecker delta function; i.e., \( \delta_{XY} = 1 \), if \( X=Y \), and 0, otherwise. Thus Equation (1), the energy function, can be rewritten as:

\[ E = -\sum_{X,Y} T_{X,Y}V_{X,Y} + \sum_{X} (1 - \sum_{Y} V_{X,Y})^{2} + \sum_{X} \left( 1 - \sum_{Y} V_{X,Y} \right)^{2} \]

To minimize the energy function defined by Equation (6) is equivalent to minimizing:

\[ -\sum_{X,Y} T_{X,Y}V_{X,Y} - \sum_{X} \sum_{i} V_{X,i} \]

(7)

which has a similar form of Equation (5). It can be shown that the network will converge to the steady state (minimum) energy if we use the following updating rule:

\[ V_{X,i} = \begin{cases} 1 & U_{X,i} > 0, \\ 0 & U_{X,i} < 0, \\ \text{unchanged} & U_{X,i} = 0 \end{cases} \]

(8)

where \( U_{X,i} \) denotes the input to the neuron \( (X,i) \), and

\[ U_{X} = \sum_{Y \neq X} \sum_{j \neq i} (T_{X,Y} + T_{Y,X})V_{Y,j} + 4 \]

(9)

where \( T_{X,Y} = \delta_{XY} - \delta_{ij} + W_{1} F_{1}(V_{X,Y}) + W_{2} F_{2}(Sy_{X,Y}(X+1),Y+(i+1)) \). Note that \( T_{X,Y} + T_{Y,X} \).

In vision, many problems can be modeled as optimization problems. Our shape recognition procedure can be summarized below:

1. The initial states of neurons are set to 1 or 0 depending on the following two conditions:

- sphericity value \( S_{X,Y}(X+1),Y+(i+1) \) indicating the local compatibility between the \( X \)th model landmark and the \( i \)th scene landmark.
- curvature matching function \( C_{X,i} \), defined by:

\[ C_{X,i} = \frac{C_{X} - C_{i}}{|C_{X}| + |C_{i}|} \]

where \( C_{X},C_{i} \) are the curvature of model landmark \( X \) and scene landmark \( i \), respectively.

\[ V_{X,i} = \begin{cases} 1 & \text{if } C_{X,i} \leq \theta_{1} \text{ or } S_{X,Y}(X+1),Y+(i+1) \geq \theta_{2} \\ 0 & \text{else} \end{cases} \]

where \( \theta_{1} \) and \( \theta_{2} \) are two thresholds. We use both
sphericity value and curvature matching function to determine the initial states of network in order to alleviate the disadvantage of either sphericity or curvature matching function. For the example, the initial state of the network is shown in Figure 3(a).

Owing to occlusion, noise and other distortions, some landmarks of the object may be missing in the scene and some extraneous landmarks may be introduced in the scene. As shown in Figure 3(a), there are six mismatched neurons being activated. To expedite the matching process, we introduce a preprocessing step. That is, all activated neurons which are isolated are set to be inactive. Thus, an active neuron is said to be isolated if none of its diagonally \((-45^\circ)\) adjacent neurons are activated. For example, in the initial network shown in Figure 3(a), the activated neurons \((2,7)(1,12)(6,1)\) are said to be isolated. After this preprocessing step the network for the above example is shown in Figure 3(b).

(2) Compute the input value to each neuron according to Equation (9). \(W_1\) is usually small because the previous hypothesis (previous iteration) may not be correct. \(W_2\) is usually larger than \(W_1\) because \(F_3\) indicates the support from the neighboring landmarks.

(3) Update each neuron according to Equation (14). As shown earlier, such updating rule will guarantee the network to converge.

(4) Repeat Steps (2) and (3) until convergence. The final state of the network indicates the "landmark correspondence pairs." Final state of the network for the earlier example is shown in Figure 3(c).

(5) We next estimate the location of the model object in the scene which will be discussed in the next section.

(6) A match error is finally computed to validate the match.

C. Location Estimation and Matching Verification

After having determined the landmarks of a model that match well with those in the scene, we next estimate the location of the model object in the scene. Location of the object in the scene is estimated by finding a coordinate transformation consisting of translation, rotation, and scaling that maps the matched landmarks of the model to the corresponding scene landmarks in a least squares sense. A score based on the least squared error of the mapping is used to quantify the overall goodness of the match between the model and the scene.

The least squared error only quantifies how well a portion of the model landmarks match with the corresponding scene landmarks. It does not, however, account for the overall goodness of match. Denote \(\varepsilon\) as the least squared error derived from the matched pairs of landmarks between the model and the scene, we use the following heuristic measure \([2]\) which penalizes incomplete matching of the landmarks of the model:

\[
\varepsilon = \begin{cases} 
(1.0 + \frac{n-3}{k-2})\log_2(\frac{n-3}{k-2})\varepsilon & \text{for } k \geq 3 \\
\infty & \text{for } k = 0, 1, 2. 
\end{cases}
\]

(12)

where \(n\) is the total number of landmarks of the model, \(k\) is the number of model landmarks that match the scene landmarks, and \(\varepsilon = e/(k(\text{scale factor}))\), i.e., \(\varepsilon\) is the normalized least squared error.

The scale factor is derived from the coordinate transformation. The heuristic measure, \(\varepsilon\), is referred to as the match error. In the earlier example, five of six of the model landmarks match with those in the scene, and the match error value is 0.81. The hypothesis of the model in the scene is finally determined by the value of the match error - with a small error, the hypothesis is accepted. The results of mapping the wire stripper and wrench into the scene are shown in Figure 4.

Figure 4: The result of mapping the wire stripper and wrench into the scene shown in Figure 2 by the least square coordinate transformation.

IV. EXPERIMENTAL RESULTS

We have performed various experiments. Here, we shall only present another example which consists of four overlapping objects as shown in Figure 5. Compared to their respective models, objects in the scene have been scaled and rotated. Compared to their respective model landmarks, landmarks are missing and extraneous landmarks are introduced in the scene. With respect to each model, those landmarks in the scene not belonging to the model are eventually made inactive by our algorithm. The results of matching each model object of the library with the scene are summarized in Table 1. The model that correctly match with the scene have the smallest match errors. Figure 6 shows the result of mapping the spacecraft, wire stripper, Sulawesi and Luzon into the scene shown in Figure 5.

V. CONCLUSIONS

There has been an upsurge of research in neural networks. As an application, we have modified Hopfield Neural Network to perform partial shape recognition. It can be shown that when the connection matrix is not symmetric, the network also converges with a different updating rule. It has been demonstrated through experimental results that
our proposed approach can handle occlusion quite well. The performance depends on the quality of the extracted scene landmarks, and the number of correct landmarks in a scene that are detectable. As compared to [1], which cannot handle scale variations, our approach can recognize objects with different scales. As compared to [2], in which the search is sequential, our approach can be implemented in parallel.

<table>
<thead>
<tr>
<th>Model</th>
<th>Matched/Total</th>
<th>Match Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>wrench</td>
<td>1/6</td>
<td></td>
</tr>
<tr>
<td>needle-nose plier</td>
<td>0/4</td>
<td></td>
</tr>
<tr>
<td>wire cutter</td>
<td>2/6</td>
<td></td>
</tr>
<tr>
<td>specially plier</td>
<td>0/6</td>
<td></td>
</tr>
<tr>
<td>wire stripper</td>
<td>5/6</td>
<td>0.83</td>
</tr>
<tr>
<td>Bomeco</td>
<td>0/7</td>
<td></td>
</tr>
<tr>
<td>Halmahera</td>
<td>2/8</td>
<td></td>
</tr>
<tr>
<td>Luzon</td>
<td>4/18</td>
<td>206</td>
</tr>
<tr>
<td>Mindanao</td>
<td>2/13</td>
<td></td>
</tr>
<tr>
<td>New Guinea</td>
<td>6/11</td>
<td>5.4</td>
</tr>
<tr>
<td>Sulawesi</td>
<td>7/9</td>
<td>0.5</td>
</tr>
<tr>
<td>spacecraft</td>
<td>5/7</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Table 1. The summary of the results of matching the library objects with the scene shown Figure 5. Objects which are in the scene are indicated by "*".

VI. REFERENCES


