Stereoscopic neuro-vision system for part recognition in intelligent assembly system

Yingen Xiong and Haifeng Hu
Department of Radio and Electronics of Zhongshan University
Guangzhou 510275 P. R. China

ABSTRACT

In this paper, a new 3D recognition method for intelligent assembly system is presented. In this method neural network technology is used to provide new methodologies for solving difficult computational problems in 3D recognition processes. The method can be divided into two parts. In the first part, phase based stereo matching techniques are used to finds the correspondence between left and right image in stereo image pair. The Hopfield neural network is adopted to implement the stereo matching process. A suitable architecture of neural network is established, so that the computation can be implemented efficiently in parallel. A 3D object reconstruction neural network is constructed by using BP neural network. With the results of stereo matching, the 3D configuration and shape can be reconstructed. In the second part, the feature vector of 3D object is constructed by using 3D moment and its invariant. With the results obtained in first parts, ART2 neural network is adopted for neural network classifier. With the ART2 neural network classifier, the 3D objects can be recognized and classified. The method is tested with both synthetic and real parts in intelligent assembly system. Good results are obtained. It is proved through the experiments and actual applications that the method presented in this paper is correct and reliable. It is very suitable for intelligent assembly system.

Keyword: Local phase, Stereo matching, 3D object recognition, Neural network, Intelligent assembly system

1. INTRODUCTION

As the development of manufacturing system is evolving towards that of intelligent manufacturing system, the development of assembly system is evolving towards that of intelligent assembly system. In intelligent assembly environment, a vision system for part identification and recognition is required to be both fast and robust[1]. The system must perform an accurate and unambiguous part recognition in a noise environment and meet the requirements of continuous operations performed at high speeds. In intelligent assembly system, the part recognition processes require not only to process 2D images, but also to recognize 3D objects. The recognition for 2D images has many successful techniques, but for 3D object recognition, it still has many problems to be solved. In 3D object recognition stage, it is very important to reconstruct 3D object model from 2D images[2]. Conventional approaches require an elaborate system calibration to determine the relationship between the two cameras and the global coordinate system[3,4]. This relationship is very complicated, and it needs long time to perform the process. If the system errors must be considered, an iterating process must be used. So the conventional approaches for these problems can not meet the needs of intelligent assembly.

As in other fields, neural networks are useful to intelligent assembly system. They can provide some very useful ability for 3D part recognition. Their learning ability can get knowledge and it is easy to realize parallel operation[5]. The 3D part recognition processes require both fast and robust and need a lot of knowledge about the assembly system. We can construct parallel operation using neural networks for increasing the recognition speed, so that it can meet the needs of continuous operation. By using the leaning ability of neural networks, we can get the knowledge from the assembly system, so that the knowledge can be used again. However, neural network has some problems that need to be overcome[6-8], we must need an
effective learning method. An improve learning method was presented by Paul and Xiong[7] to overcome these problems, and the results are very encouraging. Stereo vision is a technique for building a three-dimensional (3D) description of a scene observed from several viewpoints. In the passive stereo vision, usually two cameras are used to observe the same scene from two slightly different viewpoints and the stereo image pairs are obtained. After the left image and right image in the stereo image pair are matched. The three dimensional description of the scene can be built. Unfortunately, the stereo matching problem is very difficult. Many of problems encountered in stereo matching in the spatial domain can be avoided by treating the stereo matching in frequency domain by using local phase[8].

The Binary Adaptive Theory (ART1) paradigm[9] and the Analog Adaptive Resonance Theory (ART2) paradigm[10] offer the capability for unsupervised learning. These networks are two layer networks capable of learning on-line using a competitive learning scheme. The on-line learning capability allows the network to perform classification without off-line training session. The ART classifier is very useful for 3D object recognition in intelligent assembly system.

In this paper a new method for 3D object recognition in intelligent assembly system is presented. The neural network is almost used in very process of 3D object recognition. The learning capability of neural network is used to learn the complicated relationship between image and spatial coordinate system. The memory capability is used to store the relationship learned by the network in the parallel capability is used to speed to 3D object recognition processes. In order to solving the difficult problem of stereo matching. We treat the stereo matching problem in frequency domain instead of in spatial domain by using local phase, and use Hopfield neural network to implement the stereo matching processes. For intelligent assembly system the classifier is very important. We used ART2 neural network to construct classifier. A 3D neuro-vision system is developed to implement the whole processes for 3D object recognition. The system has been tested on both synthetic images and real parts. Good results are obtained. It is proved through these actual applications that the method and the system presented in this paper are correct and effective.

2. THE ARCHITECTURE OF NEURO-VISION SYSTEM

Different from 2D recognition, 3D object recognition is more difficulty. It need 3D information of object. So during the recognition process, we need reconstruct 3D information from the stereo image pair first. By using this 3D information we can reconstruct the 3D configuration and range map of the object. And then we can perform classification process. We construct the architecture of neuro-vision system shown in figure 1. In this neuro-vision system the original image pairs are preprocessed first, so the noise can be removed. The local phase is calculated by using windowed Fourer transformation. In order to using epipolar line to provide useful constraint in stereo matching process, we estimate epipolar line between left image and right image of stereo image pair. By using local phase we can construct stereo matching process. The Hopfield neural network is used to perform the stereo matching process. So energy function is developed. After stereo matching process, we can obtain the corresponding points between left image and right image in stereo image pair. By using the results of stereo matching the 3D object configuration is reconstructed with BP neural network and the range map of the object is obtained. The feature vector are constructed using three dimensional moment and its invariant. The object is represented by its feature vector. At last we can perform 3D object recognition process by using ART2 neural network classifier. In the following section we will describe the method in detail.

2.1 Epipolar line estimation

A back propagation neural network is utilized to estimate the corresponding epipolar line in the right image for a given epipolar line of left image. The architecture of the neural network is given as Fig 2. This neural network has four layers. The input layer has three neurons for $x$ and $y$ values of left and $x$ value of right image. It has two hidden layers, Every hidden layer has...
fifteen neurons. The output layer has one neuron for the corresponding $y$ value of right image. Through training by using training examples of different parts, the network will learn the relationship between left image and right image in the stereo image pair. So we do not need any formula to set up the relationship. This is also one of the advantages of neural network.

Figure 1. The Architecture of Neuro-Vision System

Figure 2. The epipolar line estimation neural network
2.2 Local phase calculation

The windowed Fourier transform can be used to analyze the properties of signal, and the disparity of images can be expressed with the shift of local phase.

For an infinite signal \( f(x) \), the windowed Fourier transform is defined as

\[
F(\omega, x) = \int_{-\infty}^{\infty} \exp(j\omega t)g(t)f(x-t)dt
\]

where: \( g(x) \) is window function, and

\[
g(x) = \begin{cases} 
1 & \text{if} \ |x| \leq M/2 \\
0 & \text{otherwise} 
\end{cases}
\]

\( M \) is the parameter of a rectangular window

\[
h(x) = g(x) \exp(j\omega x)
\]

\( h(x) \) is the windowed Fourier kernel

From equation (1) and (2), we obtain

\[
F(\omega, x) = \int_{-M/2}^{M/2} f(x-t) \exp(j\omega t) dt
\]

\[
= -\int_{-M/2}^{M/2} f(x+s) \exp(-j\omega s) ds = R(x) + jI(x)
\]

where \( R(x) \) is the real component of \( F(\omega, x) \)

\[
R(x) = \int_{-M/2}^{M/2} f(x-t) \cos(\omega t) dt
\]

\( I(x) \) is the imaginary of \( F(\omega, x) \)

\[
I(x) = \int_{-M/2}^{M/2} f(x-t) \sin(\omega t) dt
\]

So the magnitude of \( F(\omega, x) \) is given as

\[
|F(\omega, x)| = [R^2(x) + I^2(x)]^{1/2}
\]

and the windowed Fourier phase is defined as

\[
\varphi(x) = \arg(F(\omega, x)) = \arctan(2I(x), R(x))
\]

where \( \arctan(2y, x) \) is the same as \( \arctan(y/x) \) except that it maps to the range \( (-\pi, \pi] \).

Before the windowed Fourier transform is used for local phase, it is important to remove high frequency noise. Usually, the Gaussian filter is used.

\[
n(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{x^2}{2\sigma^2}\right)
\]

The signal becomes

\[
F_n(\omega, x) = h(x) * [n(x) * f(x)] = [h(x) * (n(x))] * f(x)
\]

The windowed Fourier transform can be extended to two dimensional situation. The Fourier kernel is expressed with two dimensional function, and the two dimensional window function is defined. The two dimensional windowed Fourier kernel can also be obtained by using the convolution of the Fourier kernel with the window function. After the signal \( f(x, y) \) convolutes with two dimensional windowed Fourier kernel, the two dimensional Fourier phases can be obtained.

In the spatial domain, we use feature or texture of images to perform the stereo matching process. In frequency domain, the stereo correspondence problem can be formulated in terms of phase matching, that is, determining the correspondence points required so that the local phases of left and right images become equal, i.e. if \( x_l \) is the point on left image and \( x_r \) is the
correspondence point on right image, we have
\[ \varphi_l(x_j) = \varphi_r(x_r) \] (11)

2.3 Phase based stereo matching

Instead of matching features (such as edges or zero-crossings) of the image in spatial, we solve stereo matching problem in frequency domain. According to the property of phase, the corresponding relationship between left image and right image can be found.

In previous work, the quasi-linear property of phase is used to find the relationship between phase and disparity, and an iteration process is used to find the disparity map. However, if the phase linearity is damaged in some conditions, some problems will occur. Instead of finding disparity map, the corresponding relationship of points between left image and right image solved directly in this paper. As mentioned if \( x_j \) point on the left image is correspondence to \( x_r \) point on the right image, they have the same local phase \( \varphi_l(x_j) = \varphi_r(x_r) \). By using this property, the matching function can be defined to fine correspondence points directly.

In order to solve stereo matching problem using Hopfield network, we must define an energy function whose minimum value corresponds to the best solution of the problem. The energy function is constructed by using the constraints that they must be satisfied in the stereo matching process. There are five constraints: uniqueness, epipolar line, ordering, continuity and geometric constraints. By using these constraints we can construct the energy function used in Hopfield network for stereo matching as following.

![Diagram](image)

Figure 3. A neural network architecture for stereo matching
\[ E = \frac{A}{2} \sum_{i} \sum_{X} \sum_{X_{s}} \sum_{x} V_{i,x,j} V_{i,x,j} \]
\[ + \frac{B}{2} \sum_{i} \sum_{X} \sum_{x} \frac{1}{N_{nor}} \sum_{j} \left[ \varphi_{l}(X,i) - \varphi_{l}(j,i) \right]^{2} V_{i,x,j} V_{i,x,j} \]
\[ + \frac{C}{2} \sum_{i} \sum_{X} \sum_{x} \frac{1}{N_{nor}} \left[ \sum_{r \in X} \sum_{j} V_{i,x,j} V_{i,x,j} + \sum_{r \in X} \sum_{j} V_{i,x,j} V_{i,x,j} \right] \]
\[ - \frac{D}{2} \sum_{i} \sum_{X} \sum_{x} \frac{1}{N_{nor}} \left[ \sum_{r \in X} \sum_{j} V_{i,x,j} V_{i,x,j} + \sum_{r \in X} \sum_{j} V_{i,x,j} V_{i,x,j} \right] \]
\[ + \frac{E}{2} \sum_{i} \sum_{X} \sum_{x} d_{i,x,j} V_{i,x,j} + \sum_{i} \sum_{X} \sum_{x} 100000 V_{i,x,j} \]
\[ d_{i,x,j} > \varepsilon \]

where

- \( l \) is the \( l \)th layer.
- \( X \) and \( Y \) represent the \( X \)th and \( Y \)th pixel along an epipolar line in the left image.
- \( i \) and \( j \) are the \( i \)th and \( j \)th pixel along an epipolar line in the right image.
- \( d_{i,x,j} \) is the value of matching function for candidate pair \((X,i)\) in the \( l \)th layer.
- \( s \) is the parallelogram region.
- \( \varepsilon \) is the threshold value.
- \( N_{nor} \) is a number for normalization.

By following the above energy function, the interconnection of neural networks are established. All the major constraint are integrated in this framework.

### 2.4 Three Dimensional Object Reconstruction

We can construct a four layers back propagation neural network to perform the operation. The input layer has four neurones for \( x \) and \( y \) values of the matched points on the left and right images of a stereo image pair. The network has two hidden layers. Each contains 50 neurons for storing the complex relationship between 2D image and 3D space. The output layer has 3 neurons for the 3D points. The architecture of the neural network is given as Figure 4.

![Figure 4 The architecture of 2D to 3D point conversion neural network](image-url)
By feeding the matched points into the input layer, the neural network can output the 3D points of the object. If multiple networks are available points can be converted in parallel.

2.5 Feature vector of 3D object

As many other circumstances, in intelligent assembly environment, it is required that vision system perform part recognition regardless of spatial position, orientation. The associated feature vector for each 3D part must be normalized against shaft, rotation, and scale change transformations. So selection of appropriate feature vectors of 3D parts is one of the most important task for 3D part recognition by using neural network approaches. Obviously, features which are invariant with respect to target and environmental variations (such as translation, rotation, scale) are of more interest than non-invariant features.

Moment invariants have the property of being invariant under the above transformations. 2D moment and its invariants have been successfully used in pattern recognition. 3D moment invariants was first suggested by Sajjadi and Hall[11]. In this research, 3D moment and its invariants are defined as the feature vectors for 3D part recognition. On one hand, it is expected that recognition can be performed on parts with arbitrary placement and orientation within the scan area. On the other hand, the feature vectors will be more suitable for neural network operation.

The 3D ordinary moments \( M_{lmn} \) of 3D object characterized by a \( \rho(x, y, z) \) density function are define as follows:

\[
M_{lmn} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^l y^m z^n \rho(x, y, z) dx dy dz
\]  

(13)

where:

\( \rho(x, y, z) \) is density function.

\((l, m, n)\) is the order of \( M_{lmn} \).

Here the geometrical attributes of an object are the principal features to be used in shape recognized. The \( \rho(x, y, z) \) function could be simplified by take a constant value on surface points and zero elsewhere

\[
\rho(x, y, z) = \begin{cases} 
1 & (x, y, z) \in R \\
0 & (x, y, z) \notin R 
\end{cases}
\]  

(14)

where \( R \) is the range map of object obtained through 3D reconstruction.

So the first order moments as shown below:

\[
\begin{aligned}
x_c &= M_{100} / M_{000} \\
y_c &= M_{010} / M_{000} \\
z_c &= M_{001} / M_{000}
\end{aligned}
\]  

(15)

The three second order invariant used here as defined as

\[
\begin{aligned}
J_1 &= M_{200} + M_{020} + M_{002} \\
J_2 &= M_{200} M_{200} + M_{200} M_{002} + M_{020} M_{000} - M_{110}^2 - M_{110}^2 - M_{001}^2 \\
J_3 &= M_{200} M_{020} M_{002} - M_{002} M_{110}^2 + 2 M_{110} M_{110} M_{101} - M_{110} M_{101} - M_{101} M_{020} M_{002} - M_{200} M_{002}^2
\end{aligned}
\]  

(16)

The other two features are shown below

\[
\begin{aligned}
J_4 &= J_2^2 / J_2 \\
J_5 &= J_3^1 / J_3^1
\end{aligned}
\]  

(17)

The volume \( (M_{000}) \) and average height \( (M_{001} / M_{000}) \) of the range map are also used in the creation of the vector. So the feature vector can be constructed as follow.
\[
F = \{f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}\}^T \\
= \{x_c, y_c, z_c, J_1, J_2, J_3, J_4, J_5, M_{000}, M_{001}, M_{000}/M_{000}\}^T
\]

### 2.6 Three dimensional part recognition by 3D moment invariant

For a given part, the 3D moments can be calculated through the above process. The relative feature vector of the part can be constructed with these 3D moments and their invariants. Different parts have different feature vectors. The following process is to recognize or classify this part.

The part recognition means to classify an input feature vector into an existing class (i.e., it is recognized as previously seen object), or create a new class for storing a new object. For this purpose, a suitable classifier must be constructed.

Neural network has learning and memory ability. It can learn the knowledge for a new part that it has not seen before and store the part in a memory. When the this part appears again, the neural network can recognize it and classify it into a relative class. This is the neural network classifier. The neural network classifier also has the ability of parallel processes. Comparing with the conventional classifier, neural network classifier has more advantages in part recognition.

There are several kinds of neural network classifiers. The back propagation neural network paradigm has been most widely used as classifier. But this classifier must be trained off-line. The Adaptive Resonance Theory (ART1 and ART2) neural network paradigms offer many desirable characteristics for use in 3D part recognition in intelligent assembly environment. ART classifier can learn on-line and has the ability to select the coarseness of the recognition through the setting of a parameter.

The Adaptive Resonance Theory neural network paradigm was proposed by Carpenter. The architecture of neural network classifier is given as Fig 5. This is a two layers neural network. The lower layer is an input layer. The up layer is an output layer.

![Figure 5 ART2 neural network classifier](image)

According to ART approach, the algorithm used in 3D part recognition is constructed as following:

**Step 1.** Present a new feature vector \( F = \{f_i\} \) to input neurons.

**Step 2.** Use bottom-up processing to obtain a weighted sum

\[
y_i = \sum B_{ij} f_i
\]

where \( B_{ij} \) is bottom-up weight.
Step 3. Use the MAXNET procedure to find the upper-level neuron with the largest $y_j$ value.

Step 4. Verify that $x$ truly belongs to the $j$th class by performing topdown processing, that is, form the weighted sum

$$\sum_i T_{ji}f_i$$

Then $F$ belong to $j$th cluster if

$$\sum_i T_{ji}f_i/\|F\| > \delta$$

where

$\delta$ is a vigilance parameter.

$\|F\|$ is the norm of $F$ vector.

If so, proceed to step 5, otherwise, go to step 6.

Step 5. Update $B_j$ and $T_{ji}$ for that specific $j$ and all $i$.

Step 6. Since $F$ does not belong to the one neuron that was the most likely, deactivate that neuron and go back to step 2. To start another class center.

When a feature vector is presented to the network, it is normalized by the inner neurons of the ART2 input layer. According to above algorithm, if the part has been seen before by the ART2 classifier, it will be recognized and classified to existing class. If the part is a new for the ART2 classifier, a new class will be created and the part will be stored in the memory of ART2 neural network. In following section, we will describe the implementation of the ART2 neural network classifier in part recognition processes.

3. IMPLEMENTATION AND EXAMPLES

During the process of 3D object recognition in neuro-vision system, the 3D object configuration must be reconstructed from stereo image pair. By using the results, the range map can be constructed and the feature vectors representing the objects can be calculated. So the 3D object recognition process can be performed by using ART2 neural network classifier. The following section we study the application of neuro-vision system in 3D object recognition.

A data set with feature vectors taken from real scan was created from four object. This test indicated whether the network would be able to distinguish between the feature vectors. The feature vector includes volume, average height, $J_4$ and $J_5$ shown in table 1.

<table>
<thead>
<tr>
<th>name of part</th>
<th>volume</th>
<th>average height</th>
<th>$J_4$</th>
<th>$J_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>gear</td>
<td>312960</td>
<td>51.45</td>
<td>35.01</td>
<td>0.0021</td>
</tr>
<tr>
<td>shaft</td>
<td>212166</td>
<td>93.15</td>
<td>13.79</td>
<td>0.00039</td>
</tr>
<tr>
<td>bolt</td>
<td>172154</td>
<td>69.27</td>
<td>145.19</td>
<td>0.0005</td>
</tr>
<tr>
<td>nut</td>
<td>193167</td>
<td>46.31</td>
<td>41.29</td>
<td>0.0032</td>
</tr>
</tbody>
</table>

We use ART2 neural network to classify these objects. The results are shown as table 2. The first column of table 2 indicates the iteration number in the test run. The next two columns represent the name and number of the input vector chosen for input to the network during that iteration. The last two column indicate the class name “net Match” and location “match Vector” in the network where the vector is either stored or matched.

From the results shown in table 2 we can see that the network correctly classified the feature vector in every case. In all of the test runs, the network correctly classified the feature vector.
4. CONCLUSION

In this paper, a new 3D object recognition method for intelligent assembly system is presented. The neural network technology is used to provide new methodologies for solving difficult computational problems in 3D recognition processes. Different from 2D recognition process, 3D object recognition needs more information. In order to obtain 3D information, in the first part of the paper, 3D object reconstruction process is used to reconstruct 3D configuration and shape of the object. Phase-based stereo matching techniques are used to find the correspondence between left and right image in stereo image pair. The Hopfield neural network is adopted to implement the stereo matching process. A three dimensional object reconstruction neural network is constructed by using BP neural network. With the results of stereo matching, the 3D configuration and shape can be reconstructed. In the second part, With the 3D information the feature vector of 3D object is constructed by using 3D moment and its invariant. ART2 neural network is adopted for neural network classifier. With the ART2 neural network classifier, the 3D objects can be recognized and classified. In the last part of this paper, the method is tested with both synthetic and real part. First, the synthetic and real image is used to test the 3D reconstruction function. The 3D configuration and shape can be reconstructed well from stereo image pair. Second the real parts are used to test the 3D recognition process. The network correctly classified the feature vector in every case. It is proved through the experiments and actual applications that the method presented in this paper is correct and reliable. From the results of stereo matching. The approach presented in this paper can be used in computer vision and intelligent assembly system.

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Input object</th>
<th>Input vector</th>
<th>Network class</th>
<th>Network vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>nut</td>
<td>4</td>
<td>nut</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>bolt</td>
<td>3</td>
<td>bolt</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>shaft</td>
<td>2</td>
<td>shaft</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>bolt</td>
<td>3</td>
<td>bolt</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>nut</td>
<td>4</td>
<td>nut</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>bolt</td>
<td>3</td>
<td>bolt</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>nut</td>
<td>4</td>
<td>nut</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>gear</td>
<td>1</td>
<td>gear</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>shaft</td>
<td>2</td>
<td>shaft</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>gear</td>
<td>1</td>
<td>gear</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>shaft</td>
<td>2</td>
<td>shaft</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>nut</td>
<td>4</td>
<td>nut</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>gear</td>
<td>1</td>
<td>gear</td>
<td>4</td>
</tr>
<tr>
<td>14</td>
<td>gear</td>
<td>1</td>
<td>gear</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>gear</td>
<td>1</td>
<td>gear</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>gear</td>
<td>1</td>
<td>gear</td>
<td>4</td>
</tr>
<tr>
<td>17</td>
<td>shaft</td>
<td>2</td>
<td>shaft</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>gear</td>
<td>1</td>
<td>gear</td>
<td>4</td>
</tr>
<tr>
<td>19</td>
<td>nut</td>
<td>4</td>
<td>nut</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>shaft</td>
<td>2</td>
<td>shaft</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>shaft</td>
<td>2</td>
<td>shaft</td>
<td>3</td>
</tr>
</tbody>
</table>
5. REFERENCES


