Dynamic Object Recognition Using Precise Location Detection and ANN for Robot Manipulator

Kyekyung Kim  
Intelligent Robotics Research Division  
ETRI, Korea  
kyekyung@etri.re.kr

Jaemin Cho  
Department of Computer Software  
UST, Korea  
Johns4378@etri.re.kr

Jihyeong Pyo  
Department of Computer Software  
UST, Korea  
pyohj@etri.re.kr

Sangseung Kang  
Intelligent Robotics Research Division  
ETRI, Korea  
kss@etri.re.kr

Jinho Kim  
Department of Electronics  
Kyungil Univ., Korea  
ho@kiu.ac.kr

Abstract—This paper presents vision based dynamic object recognition system for robot manipulation tasks by increasing the needs of automation machine vision. Object recognition or localization technology is used for pick and place task using robot. The dynamic object recognition system detects landmark features using neural network and provides grasping points of randomly located object, bin-picking object, visual servoing object to robot. The characteristic of dynamic object is free of posture, location, shape, distance, stacked form and is not restricted in illumination condition. This paper uses neural network based feature extractor and object classifier to recognize dynamic object. Dynamic object recognition system goes through image processing less impact to illumination effect, landmark feature extraction according to an object, coupled NN based object detection and recognition. We have evaluated performance of dynamic object recognition by testing detection of location, estimation of posture, distance to object and by identifying object type. And the other performance has evaluated by processing pick and place task using robot manipulator.

Keywords—Dynamic object recognition; Object detection; Robot manipulator; Artificial neural network

I. INTRODUCTION

Technical developments in robotics and vision based object recognition have focused on the manufacturing automation in production site. Recently, automatic assembling or packing process of electronic component using robot and object recognition technologies have interested intensively because of improvement in productivity, cost reduction, labor substitution. Most automation of manufacturing process have applied to well-arranged component. If the component is not located at the right position, it makes process error when automatic control robot has used for manipulation without vision. The need of automation process using vision based object recognition has increased gradually for flexible field adaptiveness. Vision based object recognition for flexible adaptation is related to solutions to the problem of diverse environment condition according to light, all sorts of different shapes and sizes, different kinds of traits in manufacturing industry.

The different recognition methods have applied to a great variety of objects in production line because the objects have different characteristics from the others that are seen in shape, material, size or the purpose of recognition, etc. And also, the object is influenced by changes in the ambient lighting conditions has a different characteristics. Object recognition in a real world application field has considered as a nontrivial task due to various types of objects. Therefore, a process improvement in each step of object recognition is always looked for ways to apply technology to streamline the project or improve flexibility.

However, the challenge problem of object recognition in a real application field is to provide high accuracy and robust performance in each step of object location detection, feature vector extraction and object recognition even a poor quality object image has captured. It seems to have not intra characteristics but inter characteristics even same object because the input image captured at a real application field is frequently distorted by illumination effect and it is transformed geometrically in accordance with the posture of object. And also it has different kinds of features according to an object type. Some features extracted from the distorted object image has geometrical transformation or partial information loss that leads to low object detection rate, low object recognition result, inflexibility to apply object recognition technology. Especially, accuracy of position detection, orientation and distance estimation is essential for robot manipulator. The distorted object image and various kinds of object shape and materials adversely affect to very precise object recognition results. For reasons mentioned above, object pick and place task using robot is considered the most challenging work in real world application field.

In this paper, dynamic object recognition for flexible field adaptation has proposed that includes object image distorted by various lighting condition and various kinds of object image by a different object characteristics according to shape, surface material, position or orientation. Investigation of image quality and application of combined image processing for selective image to enhance image characteristics go through in image preprocess step. Neural network based feature detector has used to detect landmark features, which are to detect an object region from complex background or randomly piled objects. NN based feature descriptor has used to detect object location that is robust to transformed features because neural network
has good generalization ability for a partially different features or missing features. The estimation of object orientation to be picked or placed and distance estimation are needed to implement completely for pick and place task using robot.

The performance evaluation has processed by detecting object location and recognizing various kinds of objects such as visual servoing object, bin-picking object, many kinds of mixed object. Object picking and placing task using robot manipulator has processed to evaluate object recognition performance.

II. RELATED WORK

Object recognition [1-3] using local image descriptors has applied to robotics applications [4]. SIFT [2] has a characteristic of invariant to image scale and rotation, but it requires precise segmentation of object. Several object segmentation algorithms [5-6] have proposed to extract object region from the background has proposed, but these are not effective in complex cluttered environments due to ambiguity between object and background image regions. Another object segmentation method for solving ambiguity of object region from background has processed using background subtraction and optical flow [7-8]. However, there are challenge problems to separate object features from a cluttered background, object features for robot gripper. Object motion, stereo vision and feature tracking has used to effectively extract object features from background for robot manipulation [9].

Object detection approaches are implemented by matching feature descriptor and by comprising recognition result of object region [10]. Object recognition approaches have been developed using global or local descriptions of the objects [11]. Local feature descriptor is robust to occlusions, changes in the viewpoint such as rotation and scale change, and real-time processing for robotics applications [11].

Object detection and recognition under unrestricted real world environment is very important for robot applications, but it is hard to get a very precise result for pick and place tasks for robot manipulator.

III. DYNAMIC OBJECT RECOGNITION

The purpose of dynamic object recognition is to pick an object from a position and to place the object to another position using robot manipulator. Dynamic objects includes randomly located multiple types of objects, visual servoing object, randomly piled bin-picking object, etc. The dynamic object recognition system consists of image acquisition, image processing for enhancement of poor quality image, the detection of object location using feature descriptor, object recognition, posture and distance estimation for picking recognition target object.

Worktable for dynamic object recognition is composed of several cameras and lighting which are positioned to adapt for the purpose each object recognition, and mini-conveyor has installed to recognize moving object and to perform pick and place task using robot manipulator. And also, the degree of precision of object location or distance has determined using laser sensor. Worktable configuration for dynamic object recognition is shown in Fig. 1.

![Fig. 1. Worktable configuration for dynamic object recognition.](image)

A. Image Aquisition

Image acquisition step includes selection of vision sensors such as camera or lighting and specification the location of the vision sensors. The input image of a dynamic object of several types of object is acquired by high resolution camera, which is fixed at a distance within 1m. Bin-picking object images are captured by stereo camera and visual servoing image is acquired at random position as shown in Fig. 1. All kinds of objects are positioned randomly that means free of location, rotation, distance even shape or material of object.

B. Image Enhancement

The detection of object location, the estimation of orientation and distance of pick and place object in actual environment show lower accuracy than restricted circumstance like laboratory. Ambient lighting effects and surface material of an object are related to accuracy of object recognition. Object is affected by aforementioned reasons has deformed characteristics differ from original object and loss of object information. Therefore, it is need to preserve as much as possible features from a distorted image in image processing step.

One method is to capture several images with different exposure shift value and combined features are extracted from those images. The other method is to determine image quality with acquisition image and then several image processes are combined according to image quality [12].

In previous method for object recognition [13], adaptive binarization method for local region of an object image and the Difference of Gaussian (DoG) filter have applied concurrently to extract edge information of the object. The material of an object affects negatively to low recognition performance because surface reflection of an object has caused information loss or distortion. Combination method of both image preprocessing has applied to strengthen weak points of poor quality of object image. However, the combined image preprocessing method has negative impact on normal quality object image that rather damages the characteristics of an object by emphasizing excessively. Selective sharpening method [12] has used to apply image preprocessing adaptively.

DoG filter [13] has used to extract effectively edge information even an object image has low in contrast of pixel intensity by illumination effects, as in Eq. (1). DoG filter
detects edges by differencing between Gaussian images at specific theta and calculates zero crossing values.

\[
\text{DoG}(x,y) = \frac{e^{-(x^2+y^2)/2\sigma_1^2} - e^{-(x^2+y^2)/2\sigma_2^2}}{2\pi\sigma_1\sigma_2}
\]  (1)

Adaptive local binarization [13] has used to preserve features of an object even an object image has an irregular contrast of pixel intensity by illumination or shadow, as in Eq. (2).

\[
t(x,y) = m(x,y) \left\{ 1 + k \left(1 - \frac{s(x,y)}{R} \right) \right\}
\]  (2)

where \(m(x,y)\) is mean and \(s(x,y)\) is standard deviation of the intensities of pixel \((x,y)\) in a \(w \times w\) window. \(R\) is the maximum standard deviation and \(k\) is a value in \([0.2, 0.5]\).

Local adaptive binarization method has computed threshold for each pixel in local window. The threshold has decided by analyzing the intensity of pixels in local window. A threshold \(t(x,y)\) of center pixel \((x,y)\) in a \(w \times w\) window is appeared in Eq. (2).

Binarization by accumulating edge and blob analysis have processed to detect object region, and then sharpness of edge has estimated by calculating contrast of pixel intensity [12]. Sharpness of edge has estimated by pixel intensity value of left and right adjacent pixels around edge pixels, as in Eq. (3).

\[
\Delta \text{DoM}_x = \left[ I_n(x+2,y) - I_n(x,y) \right] - \left[ I_n(x,y) - I_n(x-2,y) \right]
\]  (3)

\[
\Delta \text{DoM}_y = \left[ I_n(y+2,x) - I_n(y,x) \right] - \left[ I_n(y,x) - I_n(y-2,x) \right]
\]  (4)

where \(\Delta \text{DoM}_x\) is horizontal edge sharpness and \(I_n(x,y)\) is pixel intensity of median filtering image. \(\Delta \text{DoM}_y\) of vertical edge sharpness is estimated in the same way as Eq. (3). Horizontal sharpness of \(I(x,y)\), \(S_x(x,y)\), has calculated as in Eq. (4).

\[
S_x(x,y) = \frac{\sum_{x-\omega \leq k \leq x+\omega} \left| \Delta \text{DoM}_x(k,y) \right|}{\sum_{x-\omega \leq k \leq x+\omega} \left| \Delta \text{DoM}_x(k,y) \right|}
\]  (5)

\[
S_y(x,y) = \frac{\sum_{y-\omega \leq l \leq y+\omega} \left| \Delta \text{DoM}_y(l,k) \right|}{\sum_{y-\omega \leq l \leq y+\omega} \left| \Delta \text{DoM}_y(l,k) \right|}
\]  (6)

where \(\omega\) is the size of window.

Edge sharpness of object region \(S\) has calculated using horizontal and vertical edge sharpness of \(I(x,y)\), as in Eq. (5).

\[
S = \sum_{0 \leq n \leq N} \frac{\text{SharpPixel}_n}{N}
\]  (7)

\[
\text{SharpPixel}_n = \begin{cases} 1, & \text{if } \ Th \leq \sqrt{S_x^2 + S_y^2} \\ 0, & \text{Otherwise} \end{cases}
\]  (8)

where \(N\) is the number of edge pixels.

Selective sharpness of object region has processed using selective unsharp masking [14] on the result of sharpness estimation. The unsharp masking is to enhance object region image by correcting illumination because it may amplify background noise from illumination and it brings a negative results of crushed features extraction even heterogeneous features are combined to complement distorted object feature. For a linear unsharp masking is expressed as in Eq. (6).

\[
I_s(x,y) = (\beta + 1) \cdot I(x,y) - \beta \cdot G(x,y) \otimes I(x,y)
\]  (9)

where \(I(x,y)\) is input image, \(G(x,y)\) is a Gaussian lowpass kernel, \(\beta\) is a real number controlling the number of sharpening, and \(\otimes\) is a 2D convolution.

C. Object Detection Based Landmark Feature Extraction

Object is consisted several local features such as character, digit, symbol, graphic font character or bar code as shown in Fig. 2. These several local features are used to segment object region and landmark feature among several local features are selected as a representative feature according to each object. Several neural networks have used to extract landmark feature and local features of each object. Feature vector has extracted from statistical and structure characteristics of segmented object region. The feature extractor has implemented using three layer perceptron trained by modified backpropagation algorithm [15].
D. Object Recognition Using detailed type classifier

For discriminating target object, feature vectors have generated from edge and feature blobs of smoothed binary image as shown in Fig. 3. Several kinds of features are extracted, including statistical and geometrical features from a segmented sub-image of object region. The classifier for recognizing segmented object region from background image has implemented using three layer perceptron, which is learned by modified backpropagation algorithm [15]. To increase recognition performance, coupled artificial neural network has used to discriminate various kinds of object features.

![Coupled artificial neural network for dynamic object recognition](image)

Fig. 3. Coupled artificial neural network for dynamic object recognition.

E. Object Pose Estimation

For estimating pose of object, lines from edge of object and pose has detected by Hough transform. The pose of unformed object has estimated by maximal axis detection [16].

IV. EXPERIMENT RESULT

We have experimented to evaluate performance of dynamic object recognition such as object region detection and object recognition of several kinds of objects for pick and place task with robot manipulator. Working table for recognizing objects has consisted of 1800x870x200(mm), LED ring light of 200x200x25(mm) / 21W, vision sensor of Toshiba teli DU657M as shown in Fig. 4 (a). Vision sensors have mounted at different height or different distance within 1m distance for visual servoing, bin picking or recognition of several types of objects. Different shape or material objects are used for dynamic object recognition as shown in Fig. 4(b).

![Experimental environment for object recognition](image)

Fig. 4. Experimental environment for object recognition

To experiment this as like real world environment, database for training and testing object images have built under various lighting condition, different location and distance during 6 months as shown in Fig. 4. Recognition target objects are randomly located in a worktable or piled in a bin and input images of 2248x2050 grey 8bit BMP have captured. Several kinds of objects are detected using combined image processing with selective sharpness masking that is used for precise pick and place task for grasping with robot as shown in Fig. 5.

![Object detection results for robot manipulation](image)

Fig. 5. Object detection results for robot manipulation
The performance of object recognition for 5 kinds of objects has estimated as shown in Table 1 and 2. Object sample images of 3,600 are used for training and testing object. Recognition errors are mainly disappeared due to over-segmentation or under-segmentation of object due to surface reflection by illumination condition.

**TABLE 1. RECOGNITION ERROR RATE OF TRAINING OBJECT SAMPLES**

<table>
<thead>
<tr>
<th>Object sample</th>
<th>No. of Training sample</th>
<th>Detailed type classifier</th>
<th>Error rate(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA001</td>
<td>18</td>
<td>360</td>
<td>0</td>
</tr>
<tr>
<td>SA002</td>
<td>18</td>
<td>360</td>
<td>0</td>
</tr>
<tr>
<td>SA002</td>
<td>18</td>
<td>360</td>
<td>0</td>
</tr>
<tr>
<td>SA002</td>
<td>18</td>
<td>360</td>
<td>0</td>
</tr>
<tr>
<td>SA002</td>
<td>18</td>
<td>360</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>90</td>
<td>1,800</td>
<td>0</td>
</tr>
</tbody>
</table>

**TABLE 2. RECOGNITION ERROR RATE OF TESTING OBJECT SAMPLES**

<table>
<thead>
<tr>
<th>Object sample</th>
<th>No. of Testing samples</th>
<th>No. of Recognition samples</th>
<th>Error rate(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA001</td>
<td>360</td>
<td>352</td>
<td>2.2</td>
</tr>
<tr>
<td>SA002</td>
<td>360</td>
<td>350</td>
<td>2.8</td>
</tr>
<tr>
<td>SA002</td>
<td>360</td>
<td>351</td>
<td>2.5</td>
</tr>
<tr>
<td>SA002</td>
<td>360</td>
<td>353</td>
<td>1.9</td>
</tr>
<tr>
<td>SA002</td>
<td>360</td>
<td>354</td>
<td>1.7</td>
</tr>
<tr>
<td>Total</td>
<td>1,800</td>
<td>1,760</td>
<td>2.2</td>
</tr>
</tbody>
</table>

**V. CONCLUSIONS**

This paper has proposed about dynamic object recognition for flexible field adaptation that includes different lighting condition and various kinds of objects. Selective image method and combined image processing have applied to enhance image quality. Artificial neural network based feature detection and object recognition have used that is robust to transformed object image and several kinds of object recognition because neural network has good generalization ability for a partially different features or missing features. The estimation of object orientation to be picked or placed and distance estimation are implemented for pick and place task using robot.

The performance evaluation has processed by detecting object location and recognizing various kinds of objects such as visual servoing object, bin-picking object, many kinds of mixed object. Object picking and placing task using robot manipulator has processed to evaluate object recognition performance. The recognition error rate for training object samples of 0% and recognition error rate for testing object samples of 2.2% have obtained.

**ACKNOWLEDGMENT**

This work was supported by the R & D program of MOTIE & KEIT [10063413, Development Next Generation Manufacturing Robot for Smart Factory and Work-Space Sharing with Human]

**REFERENCES**