Object Recognition Using Image Warping in an Intelligent Space

Doojin Kim, Jung H. Oh, and Beom H. Lee
Department of Electrical & Computer Engineering, Seoul National University (ASRI), Seoul, Republic of Korea
Email: {djkim00, bulley85, bhlee}@snu.ac.kr

Abstract—We propose a method to improve object recognition performance of a robot with an intelligent space system (iSpace). Existing SIFT-based object recognition is powerful, but it has still some limitations. Among limitations, we focus on degraded performance according to the viewpoint change. In order to improve the performance, we use appearance estimation of the object in the point of robot’s view using image warping based on the robot pose and object pose. The proposed method is evaluated by experiments.

Index Terms—image warping, intelligent space, SIFT-based object recognition

I. INTRODUCTION

Object recognition is the most fundamental problem for robots to interact with the world, but still there is no obvious solution to the problem. The study to attract the attention among early recognition techniques was face recognition using the PCA technique [1] from a variety of sample data. However, this technique can recognize faces only if faces are in the center of the images. After SIFT [2] has been developed, the object recognition techniques based on the point features have become the mainstream. SIFT extracts the robust points in the scale space, and generates the descriptors by converting image gradients around the points into vectors. Therefore SIFT points are robust to various changes. SIFT significantly has improved recognition performance, but several limitations still exist. One of them is its slow processing time. Therefore, the fast version of SIFT were developed like SURF [3], PCA-SIFT [4] which complement the processing time by approximation. Recently, as the reduction technique of the processing time with GPU acceleration become common, SiftGPU [5] has been developed. Although the performance of SIFT became faster by the above the processing time reduction techniques, it is not enough simultaneously to recognize various objects, because the number of the feature points increases faster than the number of objects. In order to solve that problem, BoF [6], which performs object recognition using a dictionary of descriptors, was proposed. Researchers of object recognition expand their research field to object classification due to BoF. Apart from that, HOG [7] descriptor for the model image was proposed. It was used to recognize pedestrians using SVM with various sample data. There are various trials in order to use the HOG for other objects. However, among various recognition techniques, SIFT-based object recognition is feasible for object recognition of a robot. Therefore, we employ the SIFT-based object recognition technique, and use SiftGPU as our SIFT point extractor because of its fast processing time.

We deal with object recognition problem using the robot in an intelligent space (iSpace). The iSpace is the complex sensor system built in the task space of the robot. The system estimates a variety of information from the task space, and aids the robot by sharing the information with that. If an object exists in the space with the iSpace, object can be recognized by the sensor of the iSpace or learned by the human operator to the iSpace. Nevertheless, the robot needs to recognize the object in order to interact with that. In this situation, we focus on overcoming the difficulty due to the viewpoint change among various factors to make object recognition failed. Our method estimates the appearance of the object at the viewpoint of robots and transforms the model image using robot pose and object pose transferred from the iSpace. Then the transformed model image is used by new model for SIFT-based object recognition. Our method improved the recognition performance and it was evaluated by experiments.

II. INTELLIGENT SPACE

We constructed the iSpace with vision sensors to observe 8m × 5m space. Fig. 1 shows the concept of the iSpace. We use GeviCam GP-3780 camera. This camera captures the color image in 1/3 lower bit rate than the normal RGB camera due to its Bayer color filter. We needs to combine many cameras into one network. Thus we use the lower bit rate camera. The image capturing speed of the camera is 30 FPS and the resolution of the captured image through the camera is 1032 × 779. The captured image requires an additional conversion process using Bayer pattern weights. However, the required processing time is trivial. The camera’s lens has 68° horizontal field of view (FOV) and 54° vertical FOV. Wide FOV means covering wide area, but also much image distortion. Thus, we adopt a little wider FOV lens than normal lens generally used in the webcam. The camera’s interface is the gigabit Ethernet. The captured images are transferred to the control server and processed.
in real time. Handoff between cameras would not be discussed because it is out of scope here.

\[ e_x \text{ and } e_y \text{ means the orthogonal distance between the edgel from the model and the matched edge in x-direction and y-direction respectively. } N_{\text{edgel}} \text{ means the total number of edgels in the landmark model.} \]

### A. Robot Detection

Detection of robot using its appearance in the image from the camera of the iSpace is desired, but it requires much computational cost and has low detection rate. For high detection rate, we attached a geometrical landmark on the robot body to represent the robot. We use the edge model extracted from this landmark. The edge model is matched to find robot in the input image. The center pixel position of the matched region is the robot position in the input image and the rotation of the matched region is the robot rotation \( \theta \). Fig. 2 shows an example of the landmark and the edge model of that.

The success of the robot detection is determined by the score function of the landmark. The score function [8] is as follows:

\[ S_i = R_i \cdot (1 - w_{\text{fit_error}} \cdot E_{i,\text{norm}}). \]  

\( S_i \) is the score of the i-th landmark. \( R_i \) is the model coverage of the i-th landmark. \( w_{\text{fit_error}} \) is the weighting value of the fitting error. \( E_{i,\text{norm}} \) is the normalized fitting error of the i-th landmark. We set \( w_{\text{fit_error}} \) as 0.25 which is the same as [4].

The model coverage \( R_i \) is the value which indicates visibility of the edge model. This is calculated by the edge length and the equation is as follows:

\[ R_i = \frac{L_{i,\text{found}}}{L_{i,\text{total}}}. \]  

\( L_{i,\text{found}} \) is the contour length of found edge for i-th landmark. \( L_{i,\text{total}} \) is the contour length of edge model for i-th landmark. The fitting error \( E_i \) is as follows:

\[ E_i = \frac{\sum_{\text{all edgels}} (e_x^2 + e_y^2)}{N_{\text{edgel}}}. \]

### B. Robot Localization

The positions of cameras in world coordinates are stored beforehand in the control server. The relation between world coordinates and pixel coordinates and image undistortion was conducted by geometrical camera calibration [9]. We assumed the robot moves on even terrain because the task space is the indoor space. Each camera’s the optical axis is perpendicular to the bottom surface of the task space. Thus, the robot position in the world coordinates does not have \( z \) coordinate. Instead, robot rotation \( \theta_w \) is added. Equation (4) represents the robot position \( W \) in the world coordinates and the robot position \( P_i \) in the pixel coordinates of the i-th camera.

\[ P_i = [u \ v \ \theta_1]^T, \quad W = [x \ y \ \theta_w]^T, \quad (4) \]

\[ \hat{P}_i = K^{-1} \hat{P}_i W, \quad P = \hat{P}_i TW, \quad (5) \]

\( K \) is the transformation matrix from the world coordinates to the pixel coordinates of the i-th camera. The product of \( K \) and \( \hat{P}_i M_i \) is \( \hat{P}_i \) which is invertible.

### III. OBJECT RECOGNITION USING IMAGE WARPING

In [10], SIFT descriptor was most robust to various condition changes. However, if the difference of observation positions is large, SIFT matching would be failed. Fig. 3 shows the relation between observation position change and object recognition. We define the angle between the position where the model image was observed and the position where the robot is observing the object as viewing angle. We use the robot position and the object position transferred from the iSpace in order to overcome the recognition failure at a large viewing angle. At first, we calculate the viewing angle of the robot using the information from the iSpace. In order to estimate the model image at the robot viewpoint, we use the affine model [11] Eq. (6):

\[ A = \lambda R_1(\psi) T_1 R_2(\phi) \]  

where

\[ R_1(\psi) = \begin{bmatrix} \cos(\psi) & -\sin(\psi) & 0 \\ \sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad T_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \]

\[ R_2(\phi) = \begin{bmatrix} \cos(\phi) & -\sin(\phi) & 0 \\ \sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \]
\( \lambda \) is zooming factor, and \( \psi \) is spin angle of the camera. \( \theta \) and \( \phi \) are viewpoint angles. \( \psi \) become 0°, and \( \phi \) become -90° because the robot moves on the even terrain. Then, \( \theta \) is the viewing angle. Therefore, Eq. (6) become as follows:

\[
A = \lambda R_z(0)T_R, (-90°) = \lambda IT_R = \begin{bmatrix}
0 & 1 & 0 \\
-1 & 0 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]  

(7)

We estimate the object’s image at the viewing angle of the robot using Eq. (7). The Eq. (7) is used as the warping function. After the estimation, the estimated image is used as new model image for object recognition. Fig. 4 represents the estimated model image at viewing angle 50°.

IV. EXPERIMENTAL RESULTS

Using the pose information transferred from the iSpace, the robot calculates the current viewing angle. As shown in Fig. 3, we set viewing angle is 0° at the front of object. We performed recognition test at viewing angle 50°, 60°, 70°, and 80°. The distance between the object and robot is constantly set as 1.5m. We define the object’s area in the input image as valid region. We checked how much matched features exists in the valid region. The matched features in the valid region are called valid features. Fig. 5 shows the pronto-parallel images of objects and Objn means the n-th object. Fig. 6 shows ratios of valid features in matched features. ds means that the original model image is down-sampled as 1 octave. warp means that model estimation was performed. As shown in the graphs in Fig. 6, the number of the valid features with the model estimation are bigger than without the model estimation at the both the original and down-sampled model image.
We also tested the object contour generation through the homography estimation using SIFT and RANSAC. The object contour could be generated even if the outliers exist in the matched features. However, in order to generate the correct object contour, the number of inliers has to be enough (at least 50%). Table I is the result of the object contour generation. S means success, and F means failure. The model image size of obj1, obj2, and obj3 are 463 × 640, 450 × 640, and 450 × 640 respectively. As shown in Table I, the object contour generations with model image estimation are more successful than without the model image estimation at the wide viewing angle 60° and 70°.

<table>
<thead>
<tr>
<th>Model type</th>
<th>50°</th>
<th>60°</th>
<th>70°</th>
<th>80°</th>
</tr>
</thead>
<tbody>
<tr>
<td>obj1 warp</td>
<td>S</td>
<td>S</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>obj1 ds</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>F</td>
</tr>
<tr>
<td>obj2 warp</td>
<td>S</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>obj2 ds</td>
<td>S</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>obj3 warp</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>obj3 ds</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>F</td>
</tr>
</tbody>
</table>

V. CONCLUSION

In this paper, we proposed the model image estimation method in order to overcome the limitation of viewing angle in existing SIFT-based object recognition. This method used the information about the task space from iSpace. As shown in experimental results, we make the SIFT-based object recognition more robust to the viewing angle by using the estimated model image. However, we did not prove that our method is effective for the scale change. Thus, we will expand our method to recognition performance improvement for the scale change at future research.

ACKNOWLEDGMENT

This work was supported in part by the National Research Foundation of Korea(NRF) grant funded by the Korean government (No. 2013R1A2A1A05005547), in part by the Brain Korea 21 Project, in part by ASRI, and in part by the Industrial Foundation Technology Development Program of MOTIE/KEIT [Development of CIRT(Consortium for Intelligent Robot Technologies)]

REFERENCES


Doojin Kim received his B.S in Information & Communication Engineering from Inha University, Incheon, Korea, and M.S. degrees in Electronics Engineering from Seoul National University, Seoul, Korea. He is currently a Ph. D candidate in Department of Electrical and Computer Engineering at Seoul National University, Seoul, Korea. His research interests include image recognition, intelligent space, multi-camera tracking.

Jung H. Oh received his B.S in Electrical Engineering from Seoul National University, Seoul, Korea in 2012. He is currently a M.S. candidate in Department of Electrical and Computer Engineering at Seoul National University, Seoul, Korea. His research interests include simultaneous localization and mapping, multi-agent system coordination, fault tolerance system and intelligent space.

B.H. Lee received his B.S and M.S degrees in Electronics Engineering from Seoul National University, Seoul, Korea in 1978 and 1980, respectively, and his Ph.D. degree in Computer, Information, and Control Engineering from the University of Michigan, Ann Arbor, in 1985. From 1985 to 1987, he was with the School of Electrical Engineering at Purdue University, West Lafayette, IN, as an Assistant Professor. He joined Seoul National University in 1987, where he is currently a Professor at the School of Electrical Engineering and Computer Sciences. Since 2004, he has been a Fellow of the Robotics and Automation Society.