Automatic Evidence Selection and Collection for Robust Robotic Perception

Jangwon Lee

The Graduate School

Sungkyunkwan University

Department of Electrical and Computer Engineering

Automatic Evidence Selection and Collection for Robust Robotic Perception

Jangwon Lee

A Masters Thesis Submitted to the Department of Electrical and Computer Engineering and the Graduate School of Sungkyunkwan University

in partial fulfillment of the requirements

for the degree Master of Arts

October 2007

Approved by

Sukhan Lee

This certifies that the master thesis of Jangwon Lee is approved.

Thesis Supervisor: Jee-Hyong Lee

Thesis Committee: Tae-Yong Kuc

Thesis Committee: Sukhan Lee

The Graduate School

Sungkyunkwan University

December 2007

Contents

Figures

Tables

I. Introduction

In recent years, we are able to see various prototype robots which show impressive performance via mass media easily. Some robots seem very useful and they seem able to use immediately, however it is not easy to see that kind of robots in the real situation. Why?

One of the reasons most current personal or domestic service robots stay in research laboratories, instead of being on the market is that they have not enough capability to recognize some objects/humans in real environment. Despite the development of mechanical technology, higher-level technologies, such as perception, learning and reasoning are fall short of general population's expectation.

Especially, object/human recognition problem is the most fundamental and important issue in the robotic research field, because the robot could not provide any services such as manipulation of objects and information delivery without object/human recognition. For this reason, I will limit myself to robotic object perception in this paper. And I will deal with the evidence selection and collection method based on Bayesian theorem for object recognition mainly. Herein the evidence means features such as SIFT (Scale Invariant Feature Transform) [7], line, color, robot motion and so on, for robotic recognition.

Before looking into the proposed approach, we shall briefly see mechanisms of human object recognition.

1 Human Perception: How can we recognize the object?

The research on human perception has been studying in various research fields such as cognitive science, neuroscience, psychology and so forth. And up to the present, in order to understand human perception, such research has been studying actively.

How can we recognize the object/human? This question includes numerously large and complicated problems, so it might not be easy to answer. However it is not the major matter of concern of this article hence we shall see it briefly. For simplifying the problem, let us assume that we are seeing the cup to drink coffee. When we see the cup, our brain receives the cup image from our eyes and then the image are bent by the cornea and the lens and directed toward the retina. In the retina, the photoreceptor neurons collect the light and send signals to a network of neurons that then generate electrical impulses that go to the brain. These signal way thorough several paths to the cerebral cortex and each path not only delivers sensor input but also structuralizes and summarizes the visual information actively.

According to Hubel and Wiesel's research, human's visual system is organized by several layers, higher layer cells which receive input from many simple cells which respond only to specific features such as a vertical line and human can recognize the more complex patterns using these cells. In order words, human perception is not accomplished by the "coffee cup recognition cell," but accomplished by integration of the information which is received by each selected cell [1][2].

The present research on the role of Basal Ganglia (BG) - subcortical nuclei of the vertebrate's brain - also is worthy of attention. The BG are organized into segregated channel which receive inputs from the whole cortex and project to both brainstem nuclei and frontal parts of the cortex via thalamus nuclei. The selection occurs between the channels, a limited number of which being activated at any time. Each channel is connected with one behavior and human select appropriate channel in a given context, for example, the information of the target object in the visual field, the information to be stored in working memory, etc [3][4].

These researches mean that human perception is accomplished by information integration based on suitable cell selection and human takes appropriate behavior using the whole information.

2 Related Approaches

The object recognition has been one of the major problems in computer vision and intensively investigated for several decades. In particular, the object recognition has played an important role for manipulation and SLAM (Simultaneous Localization And Mapping) in robotic field. There have been numerous studies carried out to solve this problem for more robust and fast recognition in real environment. However, in spite of the research of machine vision over last 30 years, human-like visual perception capability in real environment is still not accomplished in a certain level. Many researchers proposed the various 3D object recognition approaches. Among them, the model-based recognition method is the most general one for recognizing the shape and object. It recognizes the objects by matching features extracted from the scene with stored features of the object [8][9].

The problem of proper feature selection from extracted features is another problem. It is a fundamental issue in the recognition field such as face recognition, speech recognition, object recognition and so forth. In order to select significant features, Morita et al. [11] have used multi-objective genetic algorithm. They have searched features using two criteria: minimization of a validity index that measures the quality of clusters and minimization of the number of features. Their results look efficient efficiency for reducing the number of features and clusters with the recognition rates are maintained at the fine level. However, it does not consider automatic selection of valid feature sets. Fan et al. [12] proposed the support vector machines based method for recognition of multi-view faces. This method selects most discriminative features directly without linearly combining the original features. But it contains no mention of the probability of feature integration for recognition. Some research use the approach based on Bayesian theorem to select valid features. Valente and Wellekens [13] proposed the Variational Bayesian learning for model selection criterion. This method has an advantage of fast recognition and robust recognition, but it does not consider mutual information between features. Several researchers [14][15] use a Bayesian classifier to evaluate the recognition rate with different feature sets. However these approaches have the above-mentioned similar problem.

3 Proposed Approach

In this paper, I propose novel evidence selection and collection method based on Bayesian theorem for object recognition and pose estimation in real environment. Herein the evidence means features such as SIFT (Scale Invariant Feature Transform), line, color and so on. The robot begins to make selections and collections of robot's own in current situation by proposed method. First the robot approaches the high possibility region including the target object then look for the evidences from selected region. Second the robot selects proper features in current scene for recognizing the object and pose estimation. The main advantage of this approach is that the robot searches the evidences actively using behaviors such as approaching, turning, and limits the region of recognition so robot can save an unnecessary processing time. In addition, this method is to estimate probability more easily using Bayesian rule and to select an optimal set of evidences automatically for robust robotic perception. Therefore, this strategy is able to help the robot take advantage of evidence selection automatically in real environment. In addition, I expect that this strategy can be applied to other systems which need robust object recognition.

4 Organization of the paper

This paper is organized as follows: Section 0 introduces the overall object recognition framework and role of cognitive perception engine in our recognition framework. Section III describes cognitive perception engine concretely, from the concept till how to select and collect features. And the results of experiment are described in Section IV; Section V concludes by discussing.

II. Overall 3D Object Recognition Framework

Before looking into the methodology for evidence selection and collection, it is necessary to see the whole framework for object recognition. Because the evidence selection and collection is a part of proposed framework for 3D object recognition and its pose estimation. Therefore in this section, I would like to describe the whole framework which is a particle filter based probabilistic fusion framework briefly in the first, second and third section and I will present the role of evidence selection and collection in the proposed framework.

1 Outline of Probabilistic Object Recognition and Pose Estimation

In our previous paper [5], we proposed a probabilistic method based on a sequence of images to recognize an object and to estimate its pose. The probabilistic pose is drawn by particles and is updated by consecutive observations extracted from a sequence of images. The proposed method can recognize not only textured but also texture-less objects because the particle filtering framework of the proposed method can deal with various features such as photometric features (SIFT, color) and geometric features (line, square). Fig. 1 illustrates block diagram of the overall 3D recognition framework.

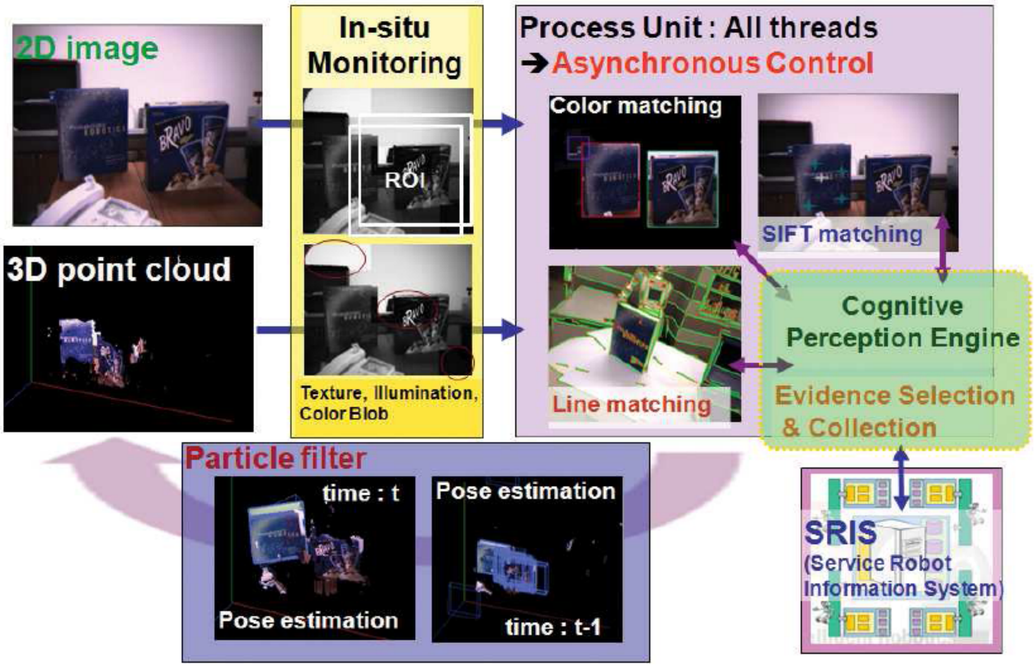


Fig. 1 Overall 3D recognition framework

First of all, the information of circumstance - density of texture, illumination and distance of expected object pose - is calculated from input image and 3D point cloud in In-situ Monitoring. Then the valid features in an input image are selected by the Cognitive Perception Engine (CPE) which perceives an environment automatically using information offered by In-situ Monitoring and keeps the evidences of all objects for their recognition. Valid features for recognizing the object are stored in Service Robot Information System (SRIS) and CPE uses this information as a priori-knowledge. The multiple poses are generated by features extracted from a scene and 3D point cloud. And probabilistic estimation is done by particle filter using measured poses and propagated probability distribution of target object in a sequence of images. See flow chart of the proposed method in Fig. 2.

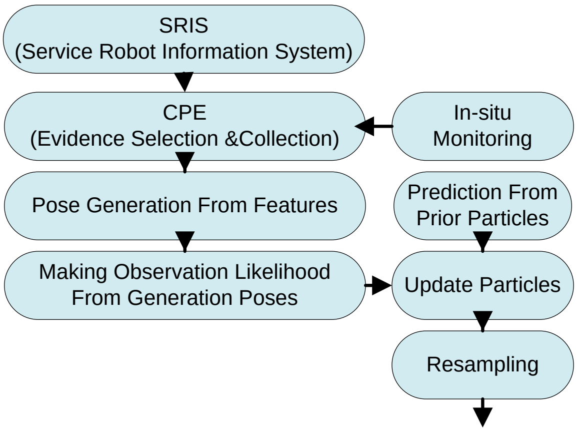


Fig. 2 Flow chart of the 3D Recognition Framework

2 In-Situ Monitoring

The In-situ Monitoring is observing system to check the changes of environment such as illumination, a mount of texture and distance between robot and target object. Fig. 3 shows simple procedure of this monitoring system. If a robot does not know anything about where is the target object, the robot should check all possible area (a). However, if the robot knows approximate locations of the target object, observing candidate area is more efficient (b).

Fig. 3 The In-Situ Monitoring System Window
(a) Observe all possible area and (b) Observe interest region

The detail algorithm of this monitoring system as follows: Initially when this module receive object recognition mission from Task Manger (TM), the system observes whole area and calculates average texture, illumination and distance on the region. At this moment if there are candidate objects, probability distributions of object pose are obtained. In our framework, these probability distributions are described particles. After then this monitoring system makes ROI (Region of Interest) for calculating average texture, illumination and distance to deliver CPE. The ROI is made using vertexes of particles and follow well known equations of perspective projection (1)(2).

$$u = FocalLength \times \frac{x}{z} + CameraCenterX$$
$$v = FocalLength \times \frac{y}{z} + CameraCenterY \tag{1}$$

Amount of texture in ROI is counted pixel which is processed by Canny edge image of current scene. And illumination is calculated using intensity information based on HSI color space in current image. Actually these two values are not the same but relative values about changes of environment. Finally, distance in ROI is calculated using processed image pixel with valid 3D point cloud and average those values. Input, output and summary of this monitoring system algorithm are shown in **Algorithm 1**.

Algorithm 1 In-Situ Monitoring Algorithm

Input:

2D images, 3D point clouds, previous recognition results

1. Make a canny edge image based on 2D images and gets intensity value using HSI color space.
2. Observe whole area and check changes of environment.
3. If previous recognition results are exist, make ROI using the previous particles
   Else return to state 2
4. Calculate average texture, average illumination and average distance.
5. Until recognition mission end, repeat state 1-4.

Output:

average texture, average distance, average illumination

3 Cognitive Perception Engine (CPE)

We assume that the valid features for recognizing each object in a current scene are already defined to the CPE. In Fig. 1 this information could be delivered from SRIS. The evidence selection and collection has been accomplished in the CPE part.

Among N possible features eligible as evidences for object recognition, where N is possibly too large for all the features to be applied, select M (M < N) features optimal in term of efficacy and efficiency for object recognition. Initially when there is no probability distribution on object pose in space, the scene captured by robot camera is used as a whole to determine a set of optimal features based on its texture content, illumination, and distance. After probability distribution of object pose is obtained, robot is supposed to move to the direction where the probability is high, then the decision on the optimal set of evidences is based on the texture, illumination, and distance information of the scene that is weighed by object probability distribution. Detailed strategy for proper evidence selection by using Bayesian rule is described in next section and evidence collection method is written in section III.

4 Particle Filter for Multiple Evidence Fusion

Particle filtering procedure is presented in previous papers [5]. The recognized object pose is estimated by particle filtering in a sequence of images over time in order that we represent the object pose with an arbitrary distribution. We keep a formulation of Motion model and Observation model in [5] which is most important parts in proposed particle filter based framework.

Probability distribution of the object pose is predicted by using probabilistic motion model and previous particles of the object pose. Multiple poses of the object generated from features at current scene without prior particles and similarity weight of each pose are used for making observation model. Then, the observation likelihood can be calculated by using predicted and measured particles considering similarity weight and Mahalanobis distance metric. According to resulting particle's weights, particles are re-sampled to represent probability distribution of the object pose at current scene. These procedures are repeated until the particles are converged to a single pose.

III. Cognitive Perception Engine

The basic concept of "Cognitive Perception Engine (CPE)" is proposed by Sukhan Lee for automatic evidence selection and collection that is a part of the above mentioned 3D object recognition framework [5]. We will see in this section concept of the CPE and how to select and collect the evidences automatically.

1 Conceptual Overview

Before introducing the concept of CPE, I think that it would be a good idea to remind the human perception as I mentioned in the introduction part. Let us consider that a little girl is finding a teddy bear in her room but if there are several similar dolls, in this situation, she may approach the place where she detects several similar dolls (or where girl used to put her favorite dolls), and then looks for another importance evidences such as clean white color, lovely big eyes, soft and warm fur and so on. In this situation, the sufficient evidences may be quickly assembled from sensing cues generated by an asynchronous and concurrent flow of perception cells in her brain. And her brain collects the information as much as possible, and selects the most valid evidences for object recognition. The concept of CPE like this, that is, CPE recommends appropriates behaviors to the direction where the object existence probability is high. These behaviors are called the "Evidence collection." And CPE selects the most valid evidences or evidence set to accomplish dependable perception. We called that this process is "Evidence Selection." Summarizing the above, CPE recommends appropriates behaviors for evidence collection and selects an optimal set of evidences based on environment changes in order to recognize the objects.

2 Bayesian Theorem Based Evidence Selection

We consider in this section the methodology of evidence selection that is a part of proposed framework for 3D object recognition and its pose estimation by using probabilistic method based on Bayesian theorem [5]. But our previous works of proposed method do not conducted the evidence selection for 3D object recognition automatically. It means that we already defined the proper evidence, feature, or feature set, for recognizing target object according to changes of environment. So, in order to select the suitable evidence that is not predefined but adapted to environmental changes, we make use of Bayesian theorem to calculate these probabilities of evidence.

A Bayesian Formulate for Evidence Selection

In general, $p(Object \mid Evidence)$, where *Object* is the target object and *Evidence* means evidence or evidence set for recognizing an object, is not easy to get directly. So, in this paper, I try to calculate using the Bayesian theorem. To assign probability, I consider how much correspondence between the recognized object and its available evidence for recognizing. In probabilistic terms the goal of proposed method is to evaluate evidence which yield the best interpretation of evidence generated by proposed hypothesis in Bayesian sense. To calculate each probability that is used for making candidate of recognition result is accordance basic Bayesian theorem as follow:

$$P(Object \mid Evidence) = \frac{1}{1 + \dfrac{P(Evidence \mid \overline{Object}) \cdot P(\overline{Object})}{P(Evidence \mid Object) \cdot P(Object)}} \tag{2}$$

What is an important part in the derived formula is negative information, $p(Evidences\,|\,\overline{Object})$. Negative information means that the probabilities of each evidence when there is no target object in current scene. Initially, this information is obtained by experimental data of each evidence when there is no target object. And it can be updated by the previous recognition result. For more detail explanation of negative information is described in section D. In addition, I define evidence, *Evidence,* as information that consists of three kinds of factors: amount of texture, the level of illuminations, and variation of distance to object.

B Probability Space Models For Each Evidence

In our framework, we have three evidences - SIFT, Color and Line - that are used as features for recognizing 3D Object and estimating its pose. These evidences have different characters changes of environment. So I made different probability space models for each evidence.

1) SIFT Model

SIFT feature has such kinds of characters: 1) SIFT is sensitive changes of distance. Actually, SIFT matching is 2D matching and this matching algorithm comparatively robust about changes of distance. But in order to get accurate object pose we use 3D point cloud information and trying 3D matching. It depends on the camera lens size, but 3D Point cloud is incorrect relative to the long distance between robot and the target object. Therefore SIFT feature has weak point in distance changes. 2) The more target object has texture; the easier the object can be matched with model. 3) SIFT points extraction works best in around 330 lux illumination condition and the points has decreased from summit. See these kinds of characters in Fig. 4.

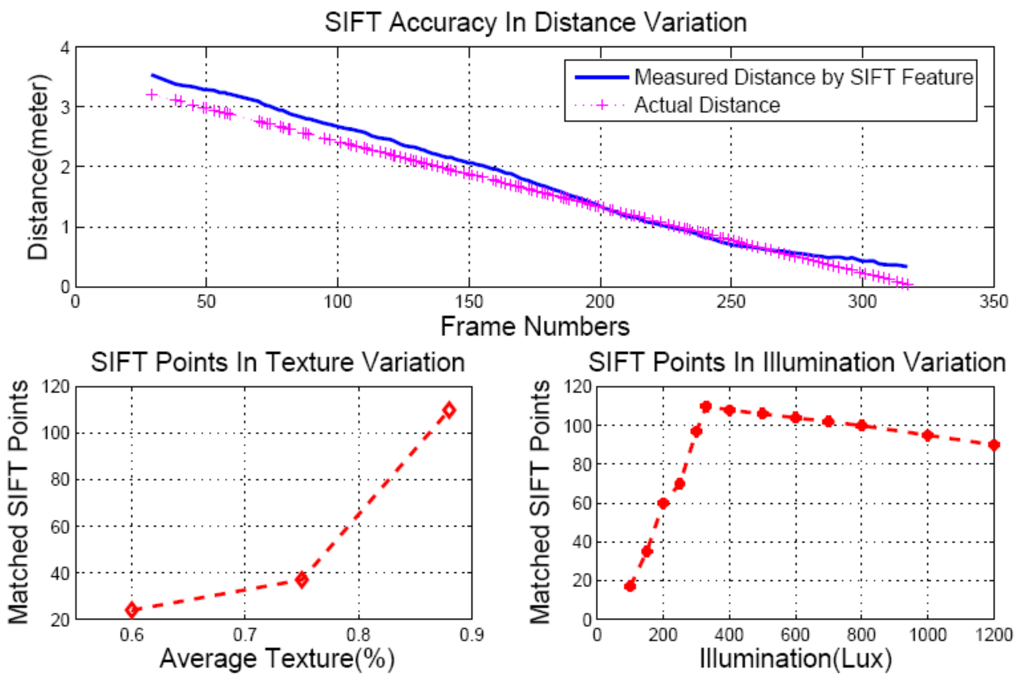


Fig. 4 Characters of SIFT feature based on experiment results

Based on these kinds of characters, I made SIFT probability model using two kinds of functions, Gaussian and linear. In case of the probability space in distance, the function is a Gaussian with 0.6 meter mean value, because I usually make model database for SIFT matching with that distance. In addition, the probability space in texture variation is modeled linearly. This is straightforward because the more texture there are, the better SIFT features are extracted. Illumination case is very similar to the probability space in distance and its mean is 330 lux that is the best condition to extract SIFT feature based on experiments. To obtain entire probability of SIFT, each sub probability is multiplied consecutively because each are independent:

$$P(E|O) = P(E_{\text{distance}}|O) \cdot P(E_{\text{texture}}|O) \cdot P(E_{\text{illumination}}|O) \tag{3}$$

Fig. 5 plots the probability space of SIFT feature in environment changes.

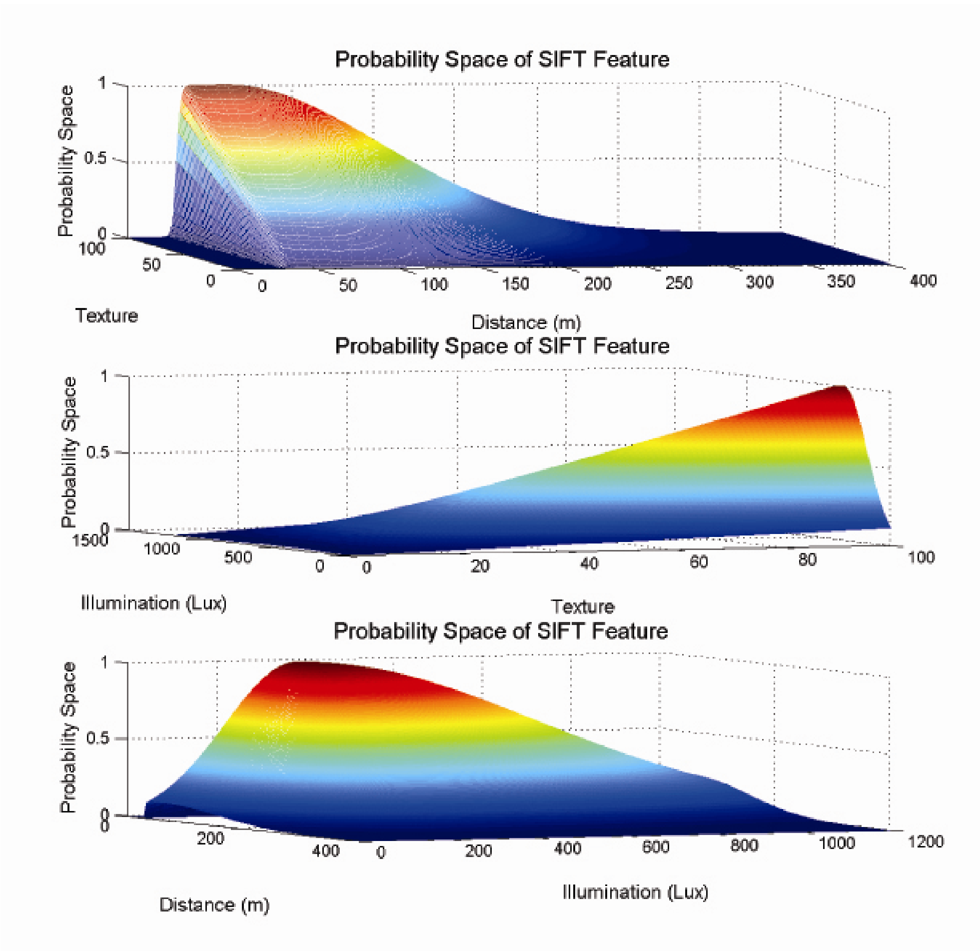


Fig. 5 The Probability Space of SIFT feature

2) Color Model

Color feature is influenced just by changes of distance and illumination, not by texture. If there are no similar color as the target object, color feature is robust about changes of distance. And this feature accuracy shows gaussian distribution in illumination changes. See these kinds of characters in Fig. 6.

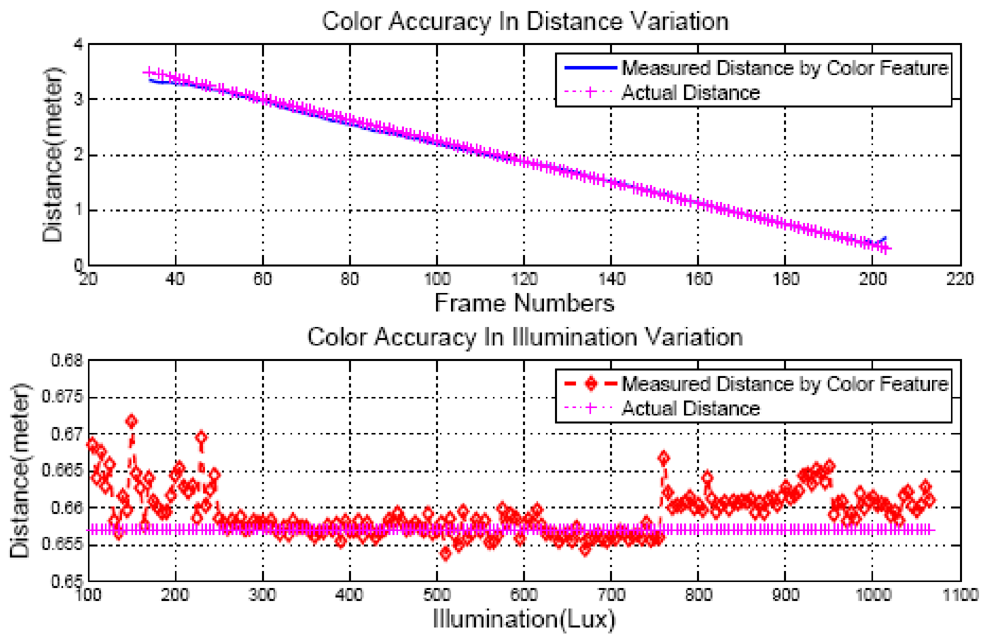


Fig. 6 Characters of Color feature based on experiment results

So I made the probability space of color feature using Gaussian function. The probability space in distance variation is made Gaussian with 1.0 meter mean value. If distance is longer than 1.0 meter, I made the probability values are decreased faster than near case. Because if there are similar color object as the target one, mismatches are often caused by long distance. The probability space in illumination variation is also modeled as a Gaussian function with large variance since illumination is not a dominant factor. In this case, mean value is 330 lux. Fig. 7 shows the calculated result of probability space in Color feature.

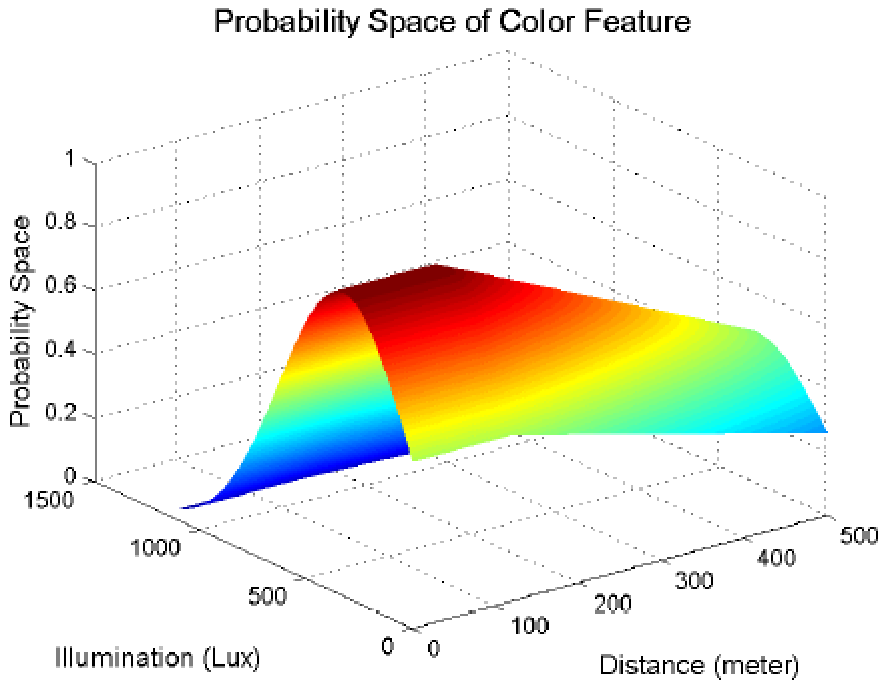


Fig. 7 The Probability Space of Color feature

3) Line Model

Although object identification by using line feature is not so good and may lead to mismatch, it can be widely applied to recognize object because of abundance of line feature. Line feature is affected by three environmental factors: 1) If the distance between robot and the target object is so far or so close, line feature is inaccuracy. Valid distance of line feature is around 1.0 meter. 2) Line feature is not sensitive when amount of texture is more than a certain level. 3) Line feature is more sensitive in dark illumination condition than bright illumination condition. Fig. 8 shows these kinds of characters.

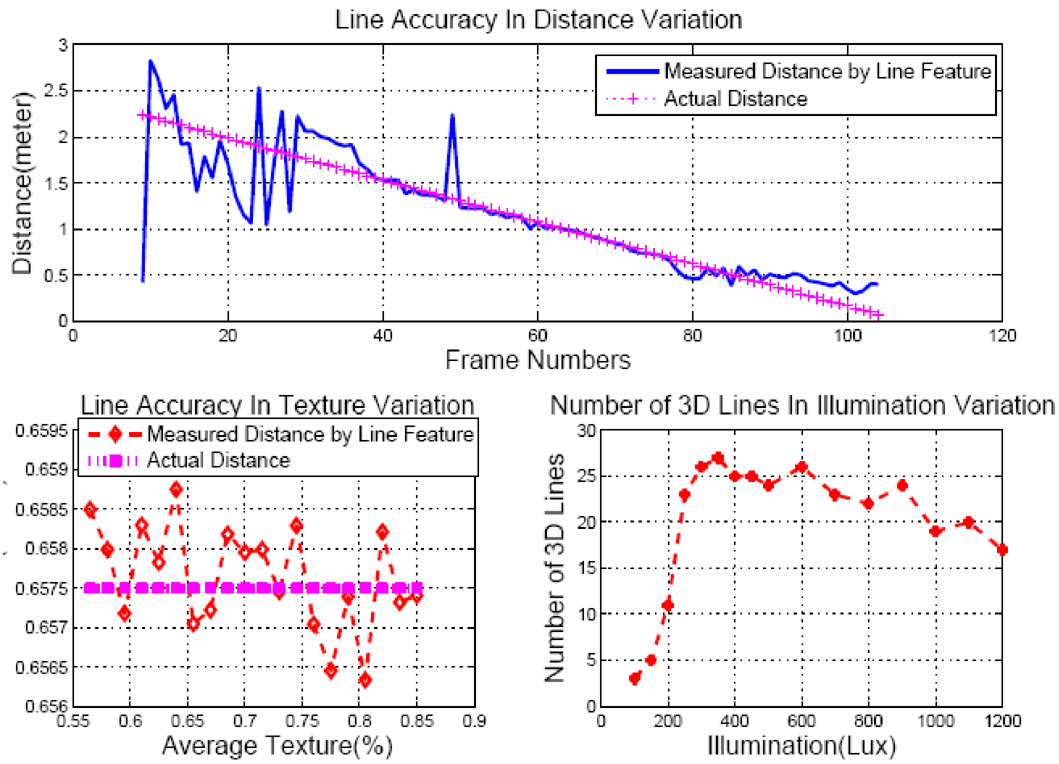


Fig. 8 Characters of Line feature based on experiment results

The probability of line feature can be modeled as Gaussian function. Especially, if amount of texture in space is more than a certain level, line feature has nearly little relationship with texture, so its variance should be large enough. The probability space of line feature is depicted in Fig. 9.

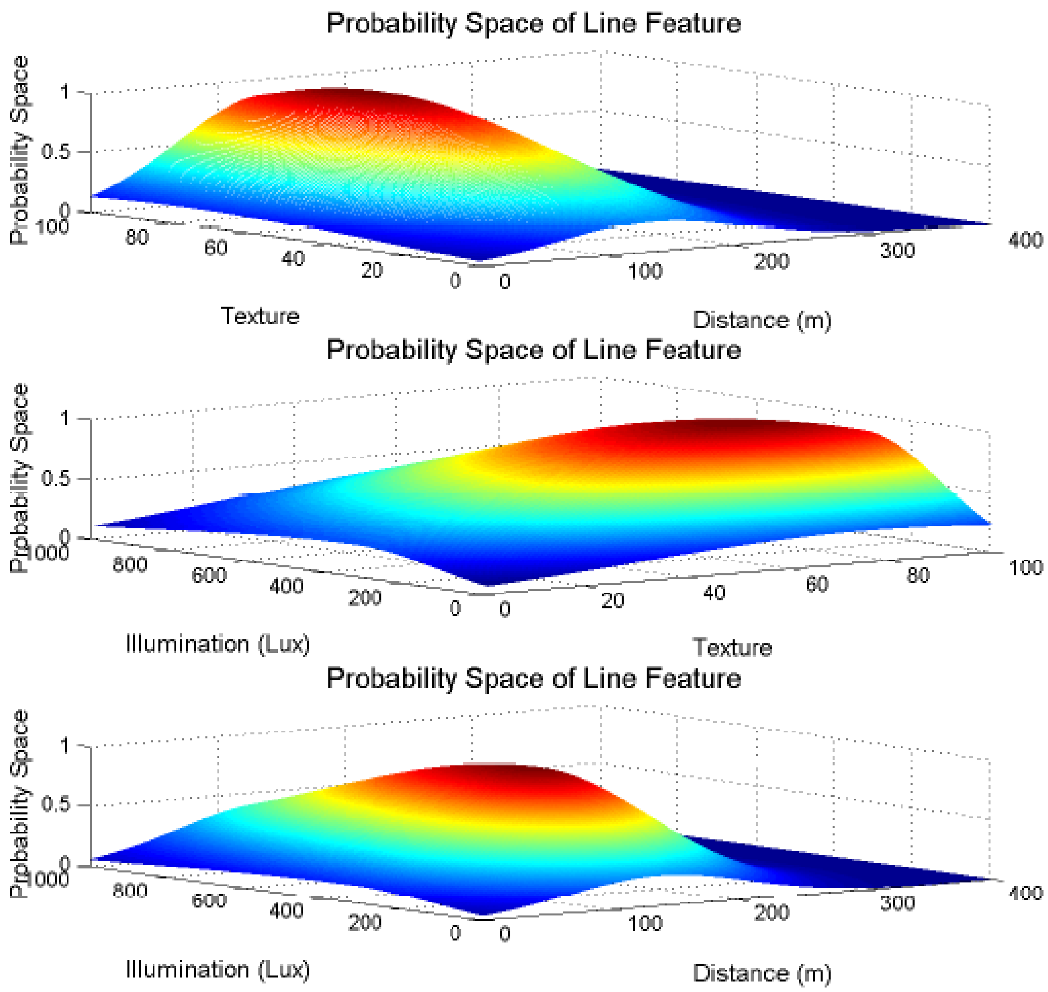


Fig. 9 The Probability Space of Line feature

C Combination of Evidences

As illustrated in Fig.1, we have three different evidences for proposed method of recognition. Those can be combined each other, like Color with SIFT, Color with Line and so on. In this paper, we just only use three kinds of combinations that are Color with SIFT and Color with Line and SIFT with Line. In the section B, we calculated already $P(E_{SIFT} \mid Object)$, $P(E_{Color} \mid Object)$ and $P(E_{Line} \mid Object)$. One of the hypotheses of proposed method is that those evidences are independent events.

So $P(E_{Color+SIFT} \mid Object) = P(E_{Color} \mid Object) \cdot P(E_{SIFT} \mid Object)$ and $P(E_{Color+Line} \mid Object) = P(E_{Color} \mid Object) \cdot P(E_{Line} \mid Object)$ and $P(E_{SIFT+Line} \mid Object) = P(E_{SIFT} \mid Object) \cdot P(E_{Line} \mid Object)$ can be obtained by Bayesian sense.

D Negative Information in Bayesian Term

As stated above negative information means the probability of each evidence when there is no object. It is represented as $P(Evidences \mid \overline{Object})$. In computing what kind of evidence is best for object recognition ($P(Object \mid Evidence)$), the negative information is a sensitive term. But it is difficult to find the probability of evidence when object do not exist. Thus we use previous recognition result of t-1 in order to get the probability. Primarily, the negative information of each feature is obtained using experimental results. We had experiments by changing environment factors such as distance and illumination, and observe probability of

$P(Evidences \mid Object)$ in the situation. And we arranged the initial value of $P(Evidences \mid \overline{Object})$ as lower bound of $P(Evidences \mid Object)$. The reason why we defined the initial value of it that manner is that we think this probability is some kind of error in each evidence. Because we made probability of $P(Evidences \mid Object)$ using environment factors, so the probability of $P(Evidences \mid Object)$ can exist when there is no object. See Fig. 10. The lower bound of each graph is initial value of negative term in our bayesian equation (2).

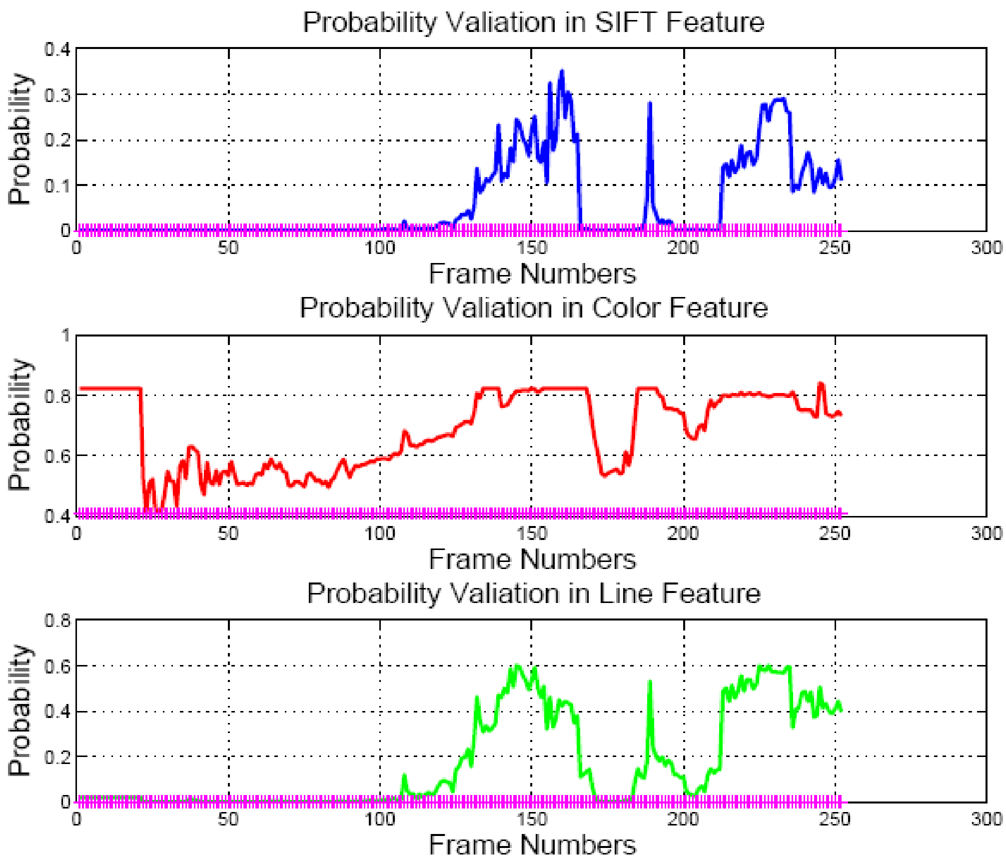


Fig. 10 The Negative Information of Each Feature

Then the negative term is updated by previous recognition result which is the probability at t-1. The update algorithm is that when the best evidence or evidence set is selected, but there is lack of the selected evidence in current scene, the negative term of selected evidence is increasing. For example, when the SIFT evidence is selected as the best evidence in the evidence selection engine, however matched SIFT points are short for recognition, negative term of the SIFT evidence is increasing. However, if the recognition success using the selected evidence the negative term is decreasing until the initial value. Using this feedback system, we can make our selection engine more robustly.

E Utility Function Considering Time Consumption

We got the each probability, $P(Object \mid Evidence)$, of possibility from the above process. But the proper selected evidence has to be considered time consumption. Even if it chooses the best evidence that takes very long time, execution time of recognition, it maybe is a good-for-nothing because of time delay. And our proposed framework also asserts real-time object recognition and its pose estimation in the real environment. So, each probability is refined by utility function that is a kind of weighted normalization function with time consumption.

$$P(i) = \alpha \cdot P(Object \mid Evidence) + (1-\alpha) \cdot (1 - \frac{ExpTime_i}{MaxTime_i}) \tag{4}$$

Where the α means weight for time consumption and ExpTime and MaxTime are real executive time and its maximum time consumption based on experimental results of recognition, respectively. In this paper, I set α is 0.8 and MaxTime of each evidence is that $SIFT = 300ms, Color = 50ms, Line = 800ms$. Final value, $P(i)$, of evidence comes from each probability. Among them, we select the highest one that is used for object recognition and pose estimation.

F Uncertainty Estimation Method and Final Decision

Finally, I calculated entropy of each evidence for final decision using calculated probability from utility function (5).

$$Entropy(i) = -P(i)\log_2 P(i) - (1-P(i))\log_2(1-P(i)) \qquad (5)$$

The reason why we can not use probability directly but use entropy is that using entropy is more efficient for evidence collection with robot behavior. For example, if there are two evidences, SIFT and Line, and probability of SIFT is 0.7 but probability of line is 0.1 then it is more efficient to move to candidate region detecting line feature for removing the candidate region (if there are no target object when robot approached the region) from robot's point of view. So we select evidence which have minimum entropy in current scene for object recognition and pose estimation and evidence collection. See equation (6).

$$Classify(i) = \operatorname{argmin}_{i=1}^{i=n}(Entropy(i)) \qquad (6)$$

3 Evidence Collection with Robot Behavior

G Evidence Collection for Robotic Perception

When the robot does not know the approximate locations of target object if robot can search for finding more information, perception performance of robot can be increased. And in this situation, running the whole recognition framework is inefficient. The reason is that it lays a burden on computing power of robot and if the distance between the robot and target object is very far, some kinds of evidences can not be detected. For example, process of finding the line features spends about 1500ms for processing time and if the distance is very far - over the two meters- SIFT feature points may not be detected. Therefore, I made the method of generating the robot behaviors for finding another evidences and limiting the recognition field.

H Entropy Estimation and Generate Robot Behavior with Considering Space

For active evidence searching, we implemented evidence collection part with considering the two-dimensional interaction space of the robot. Fig. 11 shows the concept of interaction space for active evidence searching. The interaction space is represented by 225(15*15) cells and each cell has 50cm*50cm size. After evidence selection is finished, selected evidence and its entropy is passed for evidence collection engine. And I knew translation values of particles, I can insert calculated entropy in the interaction space. Then, robot approaches and search for evidence to minimize entropy using this interaction space. However, if robot moves using entropy of current scene only, behavior of robot is unstable because information of current scene is changed dynamically in real environment. So, I accumulated entropy values of previous interaction space and integrated it with current information. In this situation, previous information is propagated in accordance with robot motion. Finally, robot select cell which have the minimum entropy and approaches selected cell for finding another evidences or recognizing the target object.

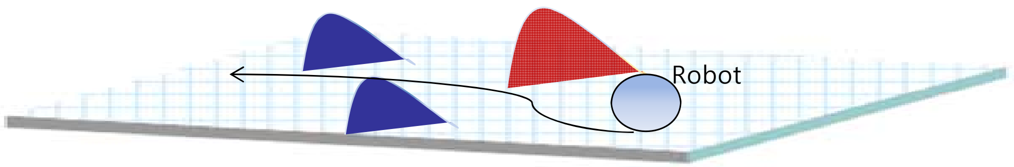


Fig. 11 Evidence Collection with Robot Behavior

IV. Experiment Results

1 Experiment Condition

In order to prove proposed evidence selection method, I made two experimental sets based on scenario. The first experiment set was carried out to demonstrate the proper evidence or evidence set was automatically selected by robot according to distance variation for robust recognition and pose estimation under 380 lux illumination condition. The target object has blue color and a lot of textures in the front side. And, in this experimentation, I put several other objects in the experiment environment. Some object has same color as the target object and some has similar shape of target object. Fig.12 shows these kinds of experimental condition and Graphic User Interface (GUI) of our recognition framework.

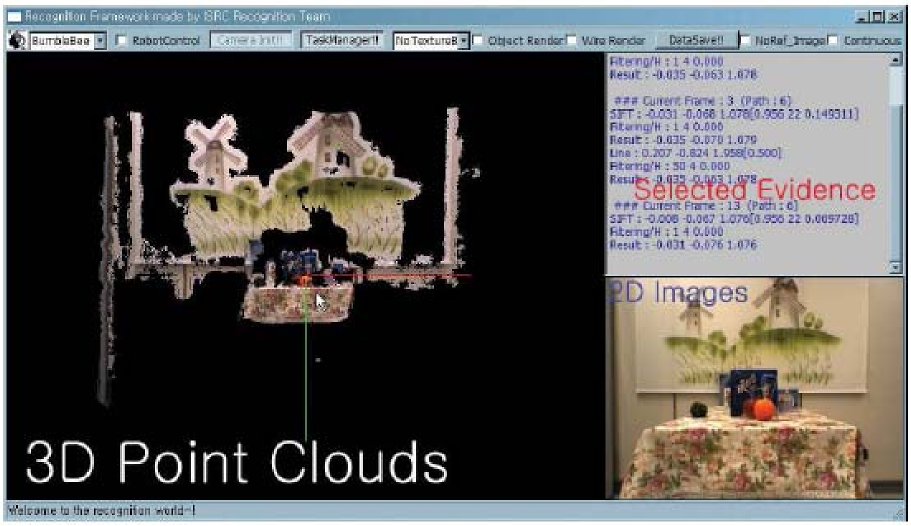
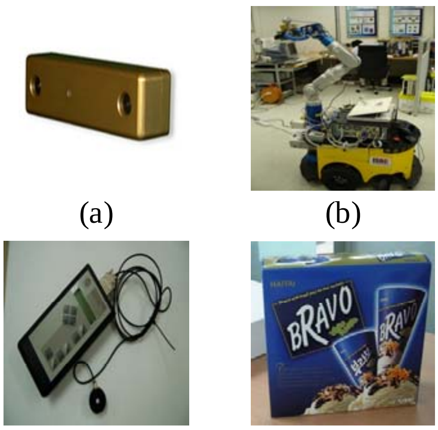


Fig. 12 Experimental Condition and GUI of Our Framework

And second experiment set was carried out to demonstrate evidence selection method in occlusion condition. In this experiment, distance between robot and the target object is fixed as 80cm and illumination is 250 lux.

Fig. 13 Experimental Condition Under Occlusion

All experiments were performed on the PowerBot-AGV platform with the Bumblebee stereo camera. And illumination condition in environment is measured by the illuminator. See Fig. 14.



(a) (b)

(c) (d)

Fig. 14 Experimental equipments and the target object.
(a) Bumblebee Stereo Camera (b) Mobile Robot
(c) Illuminator (d) Target Object

2 Experiment Results

In this paper, I used total six evidences for recognizing object and estimating of its pose: SIFT, Color, Line, Line + Color, SIFT + Color, and SIFT + Line ;however Line + Color, SIFT + Color, and SIFT + Line are combination evidences which are made based on basic features - SIFT, Color and Line. So, I can know characters of all evidences in variation of environment based on above-mentioned properties of basic features presented in section III.

A Evidence Selection under Distance Variation

To demonstrate evidence selection method I observed entropy of each evidence, measured distance using selected evidence and selected evidence. In Fig 15, X-axis of all graphs describes frame number increased time ‘t’. And Y-axis of each graph means measured distance using selected evidence (meter scale), entropy of selected path (0 ~ 1) and selected evidence or evidence set in clockwise rotation once top and the bottom. See top graph in the Fig 15. As I mentioned above previous section, each feature shows different characters according to distance between robot and the target object. For example SIFT feature shows accuracy in short distance around 0.5 meter ~ 1 meter and Line feature is accurate around 1 meter ~ 2 meter, however line feature did not detected when distance is more than 2.5meter. And SIFT feature shows some error in long distance. However, target object is detected using CPE and we can see its accuracy is fairly good in all distance. Two graphs in the bottom indicate evidence selection based on entropy is reliable. We can see color feature is selected in long distance and line combined evidence is selected in middle, and robot choose SIFT combined evidence in short distance.

These results are same as previous knowledge obtained experimental results of each feature in section III

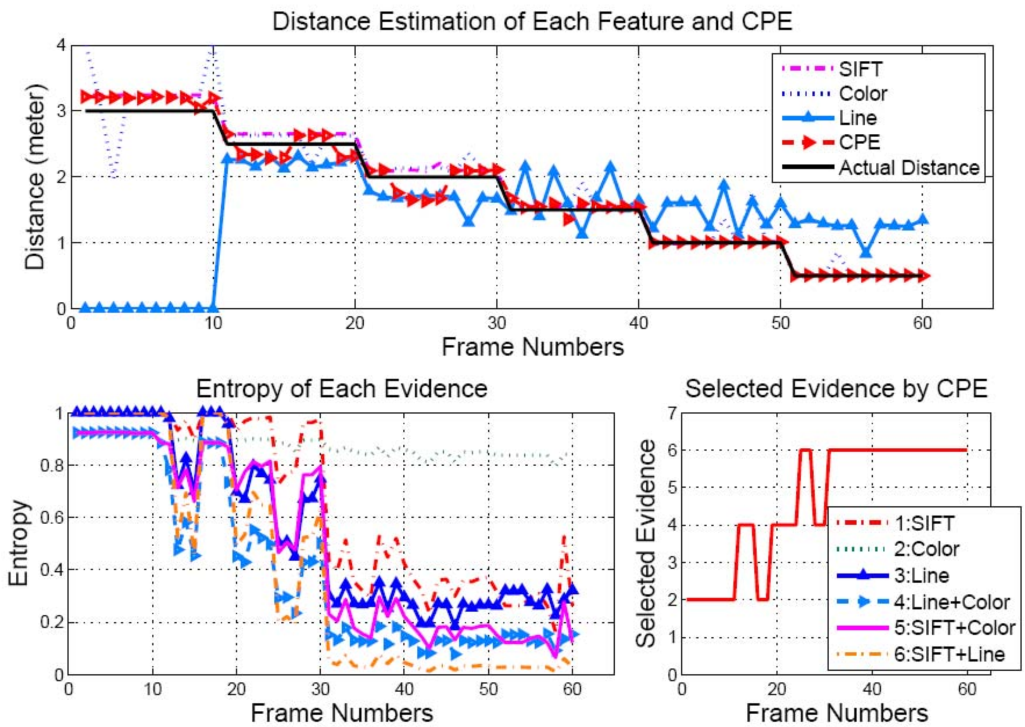


Fig. 15 Experimental results under distance variation.

B Evidence Selection under Occlusion

Sometimes the robot can not recognize object because occlusion and reflection is occurred. In spite of that kinds of troubles are existed, robot may not detect such kinds of problem and selects wrong evidence because I made probability space model using only three environment factors, distance, illumination and a mount of texture. Therefore I batten down the hatches for these errors using negative information in Bayesian equation in section III. Remind above mentioned experimental conditions about second experiment. And see this situation in Fig 13. Firstly, CPE selects most SIFT combined evidence because distance is close and illumination condition and amount of texture is enough. However, in this situation SIFT feature is not extracted, so SIFT matching is fail and negative information of SIFT feature is increasing until matching success. But in this experiment, SIFT matching is fail continuously so negative probability of related SIFT feature is increasing. So, finally, using this feedback system robot can select evidence in this situation. See Fig. 16. After 28 frames robot can select line combined evidence for best evidence.

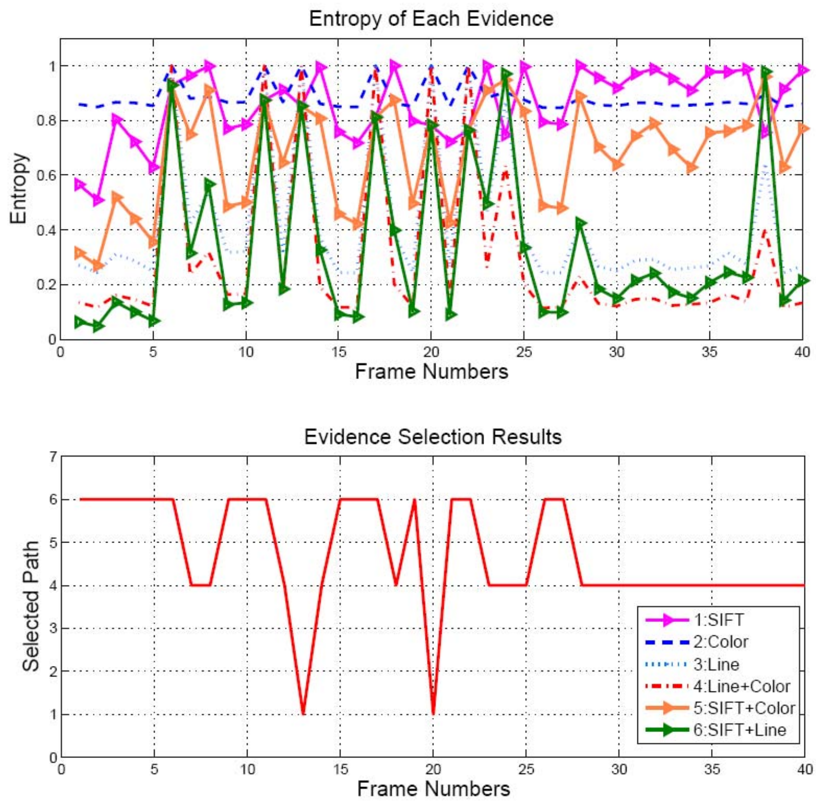


Fig. 16 Experimental results under occlusion

V. Discussion and Conclusion

The main contribution of this paper is to develop and improve the recognition strategy by proper evidence selection and collection using Bayesian rule that can be working robustly in various environmental conditions.

Experimental results demonstrated proper evidence selection in real dynamic environment including changing of illumination, variation of distance, and amount of texture. And proposed approach can select valid evidence considering processing time of each visual processing. Therefore, it can be one of the feasible solutions to deal with trade-off between performance and time consumption in visual processing.

Furthermore, to improve robustness of visual perception not only to select proper evidence but also generation of active behavior that collects evidences should be more studied as a next research step. So, active evidence selection and collection considering robot motion and real-time feedback from environment will be explored as a future work to complete dependable "behavioral perception" paradigm.

References

[1] Carolyn P. Sobel, “The Cognitive Sciences: An Interdisciplinary Approach, Mayfield Publishing,” 2001

[2] Richard MD. Restak, “The Brain,” Bantam books, 1991

[3] Girard, B., Filliat, D., Meyer, J., Berthoz, A. & Guillot, A. (2005) : Integration of navigation and action selection functionalities in a computational model of cortico-basal ganglia-thalamo-cortical loops. Adaptive Behavior, Special Issue on Artificial Rodents. 13 : 115-130

[4] B. Girard, N. Tabareau, A. Berthoz, and J.-J. Slotine, "Selective amplification using a contracting model of the basal ganglia," in NeuroComp France, 2006

[5] Sukhan Lee, Seongsoo Lee, Jeihun Lee, Dongju Moon, Eunyoung Kim and Jeonghyun Seo, “Robust Recognition and Pose Estimation of 3D Objects Based on Evidence Fusion in a Sequence of Images,” IEEE International Conference on Robotics and Automation, Rome, Italy, 2007

[6] Jeihun Lee, Seung-Min Baek, Changhyun Choi, and Sukhan Lee, “A Particle Filter Based Probabilistic Fusion Framework for Simultaneous Recognition and Pose Estimation of 3D Objects in a Sequence of Images,” The 13th International Conference on Advanced Robotics, Jeju, Korea, 2007

[7] D. Lowe. “Object recognition from local scale invariant features,” In Proc. 7th International Conf. Computer Vision (ICCV’99), pp.1150–1157, Kerkyra, Greece, September 1999.

[8] M. F. S. Farias and J. M. de Carvalho, “Multi-view Technique For 3D Polyhedal Object Recognition Using Surface Representation.” Revista Controle & Automacao, pp. 107-117, 1999.

[9] Y. Shirai, “Three-Dimensional Computer Vision,” New York: Springer Verlag.

[10] Trucco and Alessandro Verri, “Introductory Techniques for 3-D Computer Vision,” Prentice Hall, 1998

[11] M. Morita, R. Sabourin, F. Bortolozzi, and C. Y. Suen, "Unsupervised feature selection using multi-objective genetic algorithms for handwritten word recognition," in 7th IEEE International Conference on Document Analysis and Recognition (ICDAR), Edinburgh, Scotland, 2003, pp. 666-670

[12] Z.-G. Fan and B.-L. Lu, "Fast recognition of multi-view faces with feature selection," in 10th IEEE International Conference on Computer Vision (ICCV), Beijing, China, 2005, pp. 76-81 Vol. 1

[13] F. Valente and C. Wellekens, "Variational Bayesian feature selection for Gaussian mixture models," in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Montreal, Quebec, Canada, 2004, pp. I-513-16 vol.1

[14] P. C. Ribeiro, J. Santos-victor, “Human activity recognition from video: modeling, feature selection and classification architecture,” in International Workshop on Human Activity Recognition and Modeling (HAREM), Oxford, UK, 2005

[15] B. Krishnapuram, A. J. Harternink, L. Carin, and M. A. T. Figueiredo, "A Bayesian approach to joint feature selection and classifier design," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, pp. 1105-1111, September 2004

[16] S. Russell and P. Norvig, “Artificial Intelligence-A Modern Approach,” Prentice Hall International, Englewood Cliffs, NJ, USA, 1995

[17] Edwards, P. J., P. J. Edwards, et al, "Minimizing risk using prediction uncertainty in neural network estimation fusion and its application to papermaking Minimizing risk using prediction uncertainty in neural network estimation fusion and its application to papermaking." Neural Networks, IEEE Transactions on **13**(3): 726-731.

[18] Sukumar, S. R., S. R. Sukumar, et al., “Sensor Selection Using Information Complexity for Multi-sensor Mobile Robot Localization,” IEEE International Conference on.Robotics and Automation, 2007

[19] Tyrrell, T, “Computational Mechanisms for Action Selection,” Ph.D Thesis, University of Edinburgh, 1993

[20] Seth, A. K, "The Ecology of Action Selection: Insights from Artificial Life," Philosophical transactions of the Royal Society of London. Series B, Biological sciences (Modelling Natural Action Selection) , 2007

Appendix

1 Hardware Specification

A Mobile Robot - PowerBot AGV

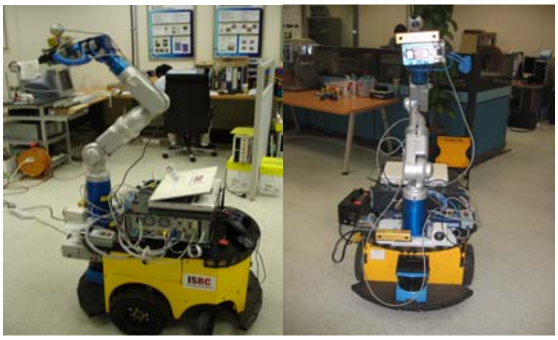


Fig. 17 Mobile Robot - PowerBot AGV

The PowerBot AGV has 85cm*62.5cm*43cm size and 2 drive wheels 2 motors with encoders, 2 real casters, sonar boards and 2500-tick moter encorders.

B Stereo Camera - Bumblebee

Fig. 18 Bumblebee Stereo Camera

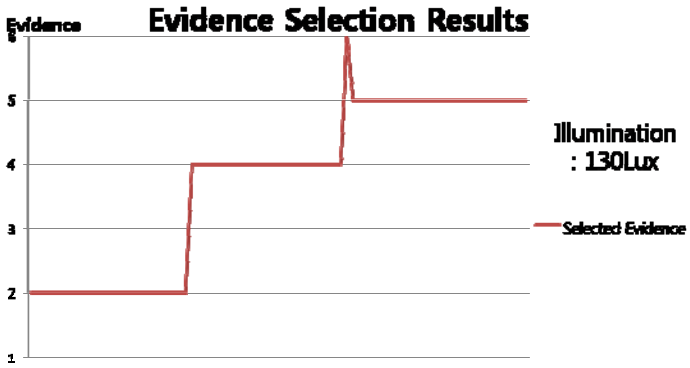
The Bumblebee is approximately 160*40*50mm size with 70° horizontal-field-of-view(HFOV) and 640*480 square pixels at 30Hz/1024*768 square piexels at 15Hz. In this implementation, I used the stereo camera with 640*480 resolutions.

2 Previous Evidence Selection Experimental Results

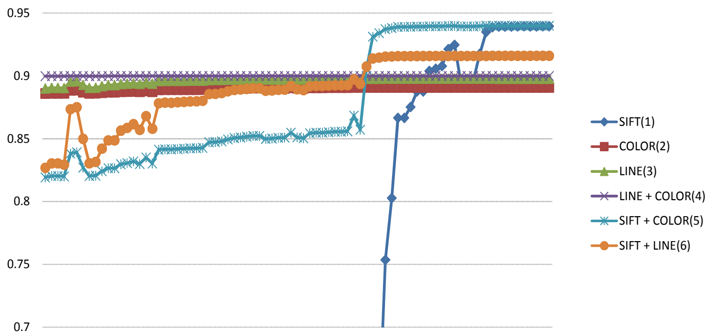
A 130 Lux Illumination Condition

This light level is a little bit dark condition for recognizing the object using SIFT feature, so probability values of SIFT feature is smaller than probability values under 330 lux condition. And detected Line numbers are also smaller than bright environment. However, probability values of color feature are almost same as 330 lux condition, because color feature is not seriously affected by variation of illumination.

Fig. 19 shows experimental results of evidence selection under 130 lux illumination condition. In this graph, X-axis represents a time sequence (steps), Y-axis in Fig. 19 (a) shows selected evidences by robot, and Y-axis in Fig. 19 (b) is probability of each evidence $p(O \mid E)$. In this experimentation, the robot moved to the target object from far distance. While the robot was moving to the object, the In-Situ monitoring was monitoring of environment variation - texture, illumination, and distance -. And then the robot selected optimal evidence set by $p(O \mid E)$. Probability of object given each feature and its variation is represented in Fig. 19 (b).



(a) Evidence selection result in 130 lux environment



(b) The trends of probability of object given each evidence

Fig. 19. The Result of Evidence Selection under 130 lux environment

And Fig. 19 (a) shows selected evidence by robot corresponding to the probability variation $p(O \mid E)$. Note that we additionally used utility function in order to consider the processing time of each evidence. Therefore, although the $p(O \mid E)$ of Line + Color evidence is lager than using Color evidence only, Color evidence is selected when distance between the robot and object is far.

Table I indicates the experimental results of object recognition. 'Mean' and 'Var.' in the table represents average distance and variation separately. 'O' and 'X' indicates amount of texture of the target object. 'O' means object has enough textures, whereas 'X' means object does not have enough textures. And cross marked cells mean that the robot is not able to recognize target object using selected evidence. Finally, gray painted cells selected evidence by the robot to optimize the recognition and pose estimation performance. We can see selected evidence is reliable for recognition by this table.

TABLE I
Experiment Results for Object Recognition under 130 lux Illumination Condition

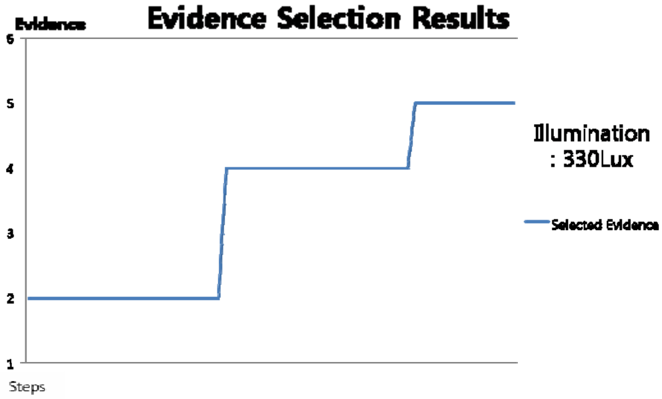
Feature \ Distance		0.5		1.0		1.6	
		Texture					
		O	X	O	X	O	X
SIFT	Mean	0.51	0.48	0.99			
	Var.	0.02	0.03	0.003			
Color	Mean	0.47	0.50	0.93	1.00	1.62	1.55
	Var.	0.05	0.07	0.06	0.14	0.13	0.28
Line	Mean	0.47	0.45	0.94	0.97	1.57	1.59
	Var.	0.05	0.05	0.06	0.03	0.13	0.08
Line +Color	Mean	0.50	0.45	0.97	0.97	1.62	1.65
	Var.	0.01	0.03	0.02	0.02	0.02	0.02
SIFT +Color	Mean	0.51	0.49	1.04	0.97		
	Var.	0.03	0.03	0.07	0.04		
SIFT +Line	Mean	0.51	0.48	1.00			
	Var.	0.02	0.03	0.09			

*Distance Unit = meter

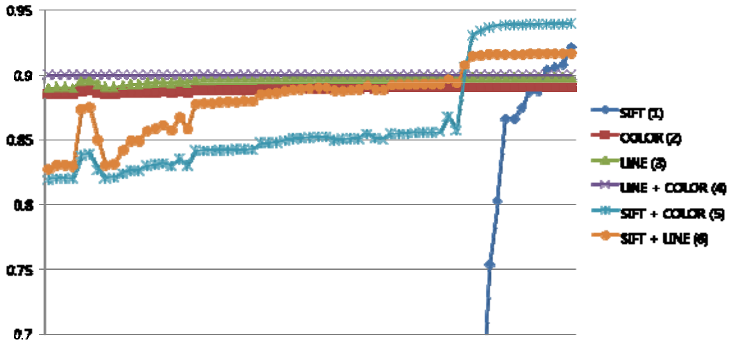
B 330 Lux Illumination Condition

The 330 lux is valid illumination condition to extract the SIFT features about the object. So, in this experiment condition, the number of SIFT features are lager than number of SIFT features under 130 lux environment. And Line features detected easily than dark environment.

Fig. 20 shows experimental results of evidence selection under 130 lux illumination condition. All introductory remarks in this figure are same as previous Fig. 19 and its results look similar to previous results under 130 lux illumination condition. The reason is that although illumination condition affects the probability of each evidence extraction, a distance variation is stronger influence the calculation of the probability of each evidence.



(a) Evidence selection result in 330 lux environment



(b) The trends of probability of object given each evidence

Fig. 20.The Result of Evidence Selection under 330 lux environment

Table II shows the experimental results under bright illumination condition. In this table we can see that evidences including SIFT and Line features are stronger than same evidences under the dark illumination condition. Therefore we can check the results selected by robot which have small variances in estimation of the object pose.

TABLE II
Experiment Results for Object Recognition under 330 lux Illumination Condition

Feature	Distance	0.5		1.0		1.6	
	Texture	O	X	O	X	O	X
SIFT	Mean	0.50	0.48	1.01			
	Var.	0.002	0.02	0.002			
Color	Mean	0.46	0.50	0.99	0.94	1.55	1.53
	Var.	0.03	0.06	0.05	0.23	0.29	0.29
Line	Mean	0.39	0.44	1.44	1.04	1.68	1.59
	Var.	0.07	0.05	0.27	0.03	0.07	0.09
Line +Color	Mean	0.47	0.45	0.94	0.98	1.61	1.63
	Var.	0.05	0.04	0.06	0.03	0.02	0.02
SIFT +Color	Mean	0.51	0.49	1.02	0.97		
	Var.	0.005	0.03	0.004	0.04		
SIFT +Line	Mean	0.519	0.49	1.07		1.66	
	Var.	0.017	0.03	0.18		0.20	

*Distance Unit = meter

Abstract

로봇의 강인한 인식을 위한 능동적인 증거 수집 및 선택 방법

이 장 원

전자전기컴퓨터공학과

성균관대학교

본 논문에서는 로봇의 강인한 인식을 위하여 제안된 로봇 인지 엔진 (Cognitive Perception Engine; CPE) 을 이용한 능동적인 증거의 수집 및 선택 방법에 관한 연구를 기술한다. 여기서, 증거라 함은 물체 인식의 목적을 이루기 위한 특징점 및 관련 행동에 관한 모든 과정을 의미하며, 로봇은 로봇 인지 엔진을 이용하여 인식 확률을 높이기 위한 증거를 능동적으로 수집하며 선택하게 된다. 본 연구에서는 로봇의 일관성 있고 합리적인 판단을 위하여 베이지안 이론(Bayesian Theorem)을 기본으로 하는 확률적 판단 방법을 제안하며, 이와 같은 연구결과는 로봇뿐만이 아닌 지능형 자동차, 차세대 보안 시스템 등 광범위한 지능형 시스템에 적용이 가능하리라고 생각한다. 마지막으로 실제 환경에서의 물체 인식 실험을 통하여 로봇 인지 엔진의 효율성 및 합리성을 보이고자 한다.

Keywords: 3D recognition, Robust Perception, Cognitive Perception Engine, Feature Selection, Evidence Selection and Collection