Small Object Discovery and Recognition using Actively Guided Robot

Sudhandhu Mittal
BITS-Pilani
India
sudhanshu0301@gmail.com

Siva Karthik M
Robotics Research Centre
IIIT-H, India
sivakarthik.m@research.iiit.ac.in

Suryansh Kumar
Robotics Research Centre
IIIT-H, India
suryansh@research.iiit.ac.in

Madhava Krishna
Robotics Research Centre
IIIT-H, India
mkrishna@iiit.ac.in

***Abstract*—In the field of active perception, object search is a widely studied problem. To search for an object in large rooms, it would be expensive to explore and check each object's similarity with the object of interest. The expense could uncontrollably bloat as the number of objects to be searched increases. If the objects are of the order of a 2-5cm, they appear very small, making it difficult for the present algorithms to recognize them. A general human strategy in such cases is to sparsely identify, from far away (4-6m), if the object of interest is present in the scene. Subsequently, each of the possible objects is analysed from closer proximity to recognize, for further manipulation. In this work, we present a similar framework. We reduce search-space, by identifying existential probability of a small object from a distance followed by a closer 3-D analysis of its point cloud to accurately recognize it. This is achieved by 2-D modelling of the objects using Gaussian Mixture Models followed by recognizing objects using efficient RGB-Depth based algorithm.**

I. Introduction

Object search in indoor environment is a widely studied problem with a fair amount of advancement. Alongside, there are a lot of challenges posed which are not addressed currently. A robot might have to search for a particular object in large unknown environments, where the objects may lay scattered on floor. We try to address a similar case, where the environment is unknown, spans over $10m$ and objects as small as $2-5cm$ lie on the floor. In this work, we wish to accomplish active object search using early probabilistic inferences based on sparse images and object viewpoint selection for robust object recognition.

The present object search paradigms cater to the aspect where the objects may be close to the camera, large in size and are generally lying on tables in an environment, which is small. Since a dominant part of the image captures the object, both 2D and 3D quality features points can be extracted. On the contrary, for a mobile robot working in a large indoor environment, the robot-object distance can vary significantly, often it is too large to infer reliable information using traditional approaches. Also, an RGB-D camera like Microsoft Kinect, fails to render quality 3-D information for distances more than $4m$. While spatial topological relations prove useful in large spaces with excessive partitions [2], they might not be fully utilizable in a case where search for small objects needs to be done in large rooms. In a large room, it proves to be expensive to visit all spaces, checking each object.

We try to build a framework resembling what humans generally do in such scenarios. When a particular object is to be searched for in a large room, where many small objects are lying on the floor, an initial inference about each object's similarity to the object of interest is ascertained. Of the many objects present in the large scene, we try to move towards the object with maximum true belief. Further, at last we do a final check from the best viewpoint for confirmation before performing our task with the object. We adopt a similar

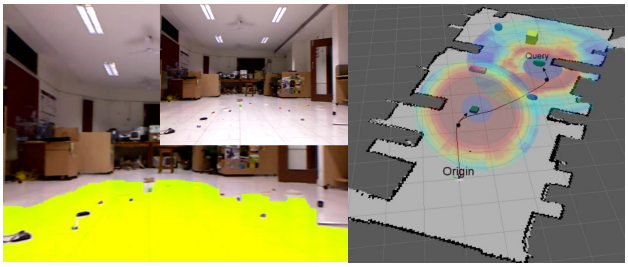

Fig. 1. Very small objects on the ground are segmented(Left). An optimal path to reach the queried object despite many objects to search for(Right).

search sequence to search for a queried object O_q through our proposed pipeline that consists of three essential components. Firstly, for a given scene we detect and segment out the objects that are lying on the floor to acquire separate segmented objects(Figure 1). This is performed using a process described in [8], using which the floor and non-floor parts are accurately distinguished. Secondly we find the existential probability of the detected object O' being a particular object O_q from far away, from where, images with resolution as low as 20×20 pixels are available. This is achieved through individual Gaussian Mixture Models(GMM) learned for all objects(Sec. III-B). If the existential probability of the detected object being the object of interest poses strong belief, then the robot is guided towards a viewpoint for that object, which gives the maximum probability of recognition. This constitutes the third phase(Sec. III-C). A distance and view angle is computed based on an object recognition profile. This profile characterizes the performance of object recognition over all camera viewpoints with respect to the object.

We present several experiments that characterize the advantage of the proposed method and verifies its efficacy in section IV. We show that the method described in [8] is an apt and necessary precursor to detect objects of 2-5cm height with high fidelity. We then analyze the performance of GMMs in terms of their effectiveness in signifying the existence of the object from far away, which helps reducing the search space. We depict

viewpoint based recognition probability profiles obtained from the RGB-D Visual Bag-of-Words(BOW) [13] model and its utility in effective recognition of the object. Section IV also discusses in detail, the criticality of the GMMs and Visual-BOW modules together for the guided search(Figure 1).

II. Related Work

The problem of object search has been studied in the past, in various related contexts like environment summarization, object oriented exploration, spatial semantic modelling, etc.

This problem dates back to 1976, when Garvey [3] proposed an indirect object search method showcasing the need to limit the search space. Subsequently, Bajcsy [7] introduced the term active perception to the community. In the recent past, works like [1], [6] argue about strong correlation between 3D structure of the surrounding environment and object placement, showing that organization is highly expressible in terms of spatial topological relations. Other works like [5] study the case where a robot simultaneously explores and searches for objects. [4] provides another solution for search and localization of objects using a monocular camera with zooming capabilities to overcome the limitations of low resolution images of distant small objects.

[2] gives an intensive strategy based on the probabilistic model, POMDP, making use of uncertain semantics between the object and its location, for prioritizing the search effort to promising locations in a partially known environment. In this pioneering work, a probabilistic semantic mapping framework is proposed, defining joint distribution between each object category and room category. Hence, at a higher level of abstraction, we would be able to discover a plausible location of the object O_q.

In our work, we try to bridge the voids encountered in scenarios where semantic relations start to weaken. For instance, when robot enters a particular room it may find a marker pen or a water bottle, unbiased towards any location. Since such objects do not posses any semantic relationships with the environment or among themselves, they have to be searched all over.

III. System Overview

The motivation behind the system is to reduce the number of objects to be looked closely while searching for an object in a large room. We decide to approach a selective few, based on prior interpretation of objects in the scene from far away. The flow diagram(Figure 2) presents the main idea of our framework for the object search problem. Both RGB and Depth data from the Microsoft Kinect sensor are exploited, maximizing its limited capabilities to our advantage. The sequence of operations listed below play a vital role in frictionless execution of our framework.

Let $\mathcal{O} = \{O_i\}_1^N$ be the universal set of N objects that are possible to exist in the environment. A robot is given a task to search for a particular queried object O_q. We wish to find the object O_q in minimum time with maximum accuracy. This can be achieved by visiting only a few of them, which we believe are close to being O_q. There are three modules which contribute to achieve the goal of object search. The object

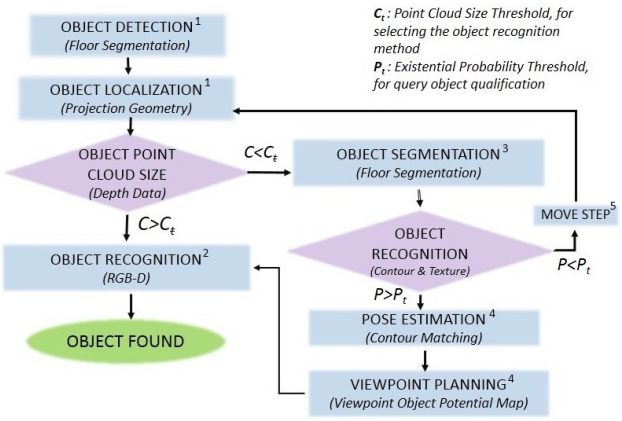

Fig. 2. Object search system overview

detection and localization module(Sec. III-A) is responsible for detecting and segmenting the objects on the floor. This relies on the method described in [8] where small objects on the floor are detected through a Graph Cut on an MRF formulated using the superpixels of the image as nodes(Figure 3). After all the objects are extracted, a belief of each object being similar to O_q is assigned through a probability, P. This is done using a set of GMMs G learned over feature vectors(III-B) generated for each of the objects in $\mathcal{O}$. Now that we have a weak inference about the objects in the scene, we go towards the objects showing high belief. This is followed by recognizing the objects using BOW model from the best viewpoint, derived from the object profile.

The flow of the algorithm(Figure 2) is as follows:

1) : The robot enters a large room with objects lying scattered on the floor. The scene is divided into small clusters of pixels called superpixels(Figure 3). The small objects on the floor are generally accommodated in a couple of superpixels, whose boundaries are generally aligned with those of objects. Each of these superpixels would act as the nodes of an MRF. A graph cut over the MRF would give a clear labelling for each pixel, if it is a floor or non-floor region. This module can extract objects as far as 4-$6m$ on a floor using a single image from monocular kinect camera of resolution 640×480(Figure 1). The k extracted objects are in set $\mathcal{F} = \{f\}_i^k \subset \mathcal{O}$. Due to unavailability or high noise in depth data, the object fails to be localized using a Kinect depth sensor at distances greater than $\sim 4m$. We localize each of the objects in $\mathcal{F}$ using projective geometry. A detailed description is given in section III-A.

2) : For each object in $\mathcal{F}$, a check over available point cloud size (C_i) is performed. If $C_i \geq C_t$ (Figure 2) we have a clear rendering of the object and recognition can be performed over the RGB-D point cloud data using an SVM classifier learned over RGB-D based Bag of Words descriptors of all objects $\mathcal{O}$(III-C). The recognition is performed by viewing it in an appropriate pose which is analysed based on a VOP(Viewpoint Object Potential) map(Figure 4) described in section III-C.

3) : If $C_i \leq C_t$, we do not have clear RGB-D information about the object. Hence from far away, for each of the objects in $\mathcal{F}$, we assign a probability(P_i), of F_i being similar to O_q. This is obtained from a GMM modelled for each of the objects

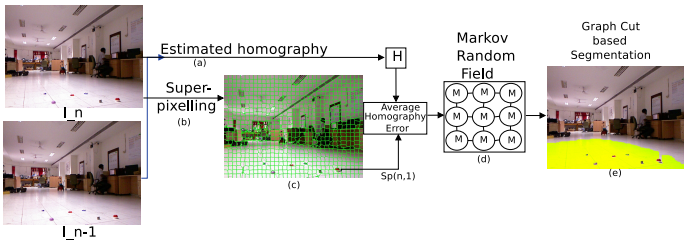

Fig. 3. Flow chart shows different stages of object segmentation pipeline.

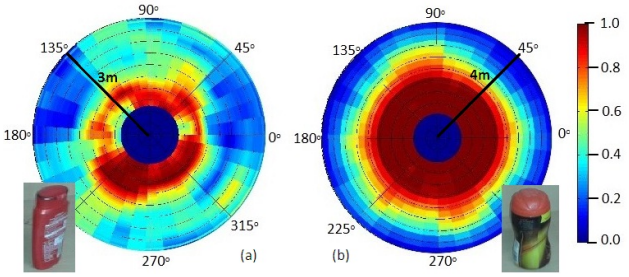

Fig. 4. VOP map depicts the accuracy of recognizing an object from different viewpoints. (a)Shows the profile for an object with a slim sideline and a wide body. (b)Shows the profile for a symmetric body.

in $\mathcal{O}$ using the object contour and texture(III-B). Hence, we have an early inference of F_i being O_q even before going close to it. Even when we have an object segmented into a few pixels, we would be able to make an inference P_i about it in a probabilistic manner. Hence, of all the objects in $\mathcal{F}$, some objects would be completely ruled out because they have a very low probability of similarity. And therefore we would reduce the number of objects that we need to visit.

4) : If $P_i \geq P_t$, it means there is a substantial belief in the object, and its pose is thus estimated as described in III-B. For all objects with $P_i \geq P_t$, the robot reaches each object to recognize it from a best viewpoint, planned from VOP map. Further, it visits all such objects using distance based greedy approach.

5) : If for all objects in $\mathcal{F}$, $P_i \leq P_t$, then the algorithm iterates after moving a finite step towards the objects, to gather more information about them.

Classifying objects merely based on their GMMs would be erroneous since there would be objects with similar contours and texture in the scene(Figure 5). Therefore, the objects need closer inspection. The accuracy of the object being recognized greatly depends on view angle and distance(Figure 4). The pose estimated from the best match in step 3 transforms the potential(VOP) map of the object to find the best viewpoint instead of moving aimlessly towards the object. VOP map is an intrinsic object property which describes the probability of recognizing that object from all possible distance and angle in 2D occupancy grid space. All the decisive operations involved in the framework are as explained below.

A. Object Detection and Localization

1) Small Object Detection using Floor Segmentation: In a typical setting we are discussing, the first task is to classify floor and non-floor regions so that the objects are extracted out. One of the main hindrances to identifying the objects lying on the floor is their small size(2-5cm). Through our previous work [8], we are able to segment the floor area even when there are several low lying objects present on it(Figure 3). A brief discussion of the adoption of [8] in our context is presented below.

The image superpixelling adopted in [8] helps in clustering a given scene along the contours and edges leading to numerous superpixels by capturing the local features of the constituent elements of an image. Since the objects in our case are small, it is observed that the whole of the object gets clustered into 1-3 superpixels, whose boundaries are aligned with those of the objects. Simultaneously, the consolidated homography error of all the tracked pixels in a particular superpixel can be calculated. Thus, we can isolate areas which differ in overall homography errors despite the individual homography errors for pixels being very low. Further, a Markov Random Field is formulated using the homography error of the superpixels, with the superpixels as its nodes. The MRF helps capture the neighbourhood information of a node as well, apart from its own homography error. The MRF is posed such that its minimum energy configuration corresponds to the target segmented image. Each of the superpixel is thus given a label of being a floor or non-floor and hence the segmentation happens. The energy function of MRF is given by

$$\psi(x,\xi) = \sum_i \psi_i(x_i,\xi_H) + \sum_{(i,j)\in\mathcal{N}} \psi_{ij}(x_i,x_j,\xi_H) \quad (1)$$

where $\psi_i(\cdot)$ is the unary term associated with i^{th} super-pixel and $\psi_{ij}(\cdot,\cdot)$ is the smoothness term defined over neighbourhood system $\mathcal{N}$. Here $x = \{x_1, x_2, ..., x_n\}$ is the set of random variables corresponding to superpixels of image. Each of these random variables, the super pixels takes a label $x_i \in \{0, 1\}$ based on whether it is a floor or object.

The unary term of a superpixel would be,

$$\psi_i(x_i,\xi_H) = (\xi_H^2)\bar{x_i} + (\xi_H^2)x_i \quad (2)$$

where ξ_H is the consolidated homography error associated with each of the superpixel using KLT feature tracker.

The smoothness term is defined using Pott's model as follows,

$$\psi_{ij}(x,\xi_H) = \lambda_2 \sum_{(i,j)\in\mathcal{N}} (\xi_{Hi} - \xi_{Hj})^2, if\ x_i \neq x_j \quad (3)$$

where λ_2 determines the degree of smoothness. Post formulation of MRF, the problem now is to find the global minima of the energy function. This is defined as,

$$x^* = argmin_x \psi(x,\xi) \quad (4)$$

The global minima of this energy function is computed using a graph cut formulation. A weighted graph $G = (V, E)$ is constructed using the vertices as the superpixels connected to the neighbours as the adjacent superpixels. The weights of the edges are defined using the unary and the smoothness terms defined above. The min-cut of this graph is computed, which corresponds to the global minima of the energy function. Once the minimum energy configuration is found, the labels of the superpixels would explain if they are floor or non floor regions. And hence, we segment out the superpixels that have a label of non floor, thus extracting out the object.

Fig. 5. The contours of objects segmented from a typical scene.

2) Object Localization: Extraction of depth from single image proves to be challenging task, thus making object localization tougher even after object is detected. By using perspective projection geometry of a pinhole camera, for localizing objects in 2D space, we estimate the distance of the object from the camera. Given height H of the Kinect camera of focal length f and with normal of the floor known, the 2D coordinates of the object can be obtained [9].

B. Probabilistic recognition and pose retrieval of small objects

As discussed earlier, we are searching for specific small objects lying on the floor, in spacious indoor settings. Unlike larger objects, where local features can be successfully extracted, we consider smaller objects where, local features if found are insignificant and sparse(Figure 5). This limitation proves to be tricky to handle in the case of object recognition from far away. Thus we have to bank on generative models based on other possible features of the objects.

A general human approach in such cases, is to look at the rough shape of the object and infer about it based on the colour composition. In [18], it is shown that contours play a major role in recognition of objects. We build a feature descriptor for the whole small object, based on which separate GMMs are estimated for each of the object. GMMs have been used in the past to build localized object models [17]. These models would help us determine the existential probability of a specific object in the scene and hence its probable pose as well.

The process of building the model and its application is as follows. From $\mathcal{O}$, each object O_i would be pictured at equally spaced orientations differing by θ, from $0-2\pi$ which gives T such images.

Such images are captured at 5 different distances in the range 5-$8m$ between the camera and the object. Hence for each object O_i, we have $5T$ images. Each of the image is subjected to contour detection to extract its outer shape(Figure 5).

Further, these contours are described using unique Rotation and Translation invariant Hu image Moments [14] which describe an image in a 7 dimensional space. Hence for a binary image which contains a contour, we obtain a 7 dimensional vector. This is followed by extracting the RGB histogram of the object. Each of the channels Red, Green and Blue are assigned 25 bins each to form a 75 dimensional histogram. The 7 Hu Moments and the 75 Dimensional histogram of Red, Green and Blue are concatenated to form an 82 dimensional feature vector F.

This leads to a formation of $5T$ feature vectors of 82 dimensions each for a given object and hence a matrix of $5T \times 82$ where each row is a vector corresponding to an orientation θ of the object at a certain distance of camera. All such vectors of N objects are stacked to form a tall $5TN \times 82$ matrix. Let F_{θ_i,d_i} represent a vector that is formed for an object with orientation θ_i and distance d_i.

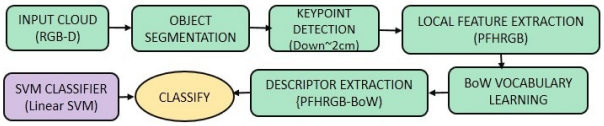

Fig. 6. Object recognition pipeline for RGB-D data.

This matrix is further transformed into a new subspace using Principal Component Analysis [15] where the dominating 7 components which contribute to the variance are identified. Hence, we obtain a new matrix of size $5TNX7$ where each row corresponds to the F_{θ_i,d_i} projected to the new 7 dimensional space. Let the new vectors in the transformed space be denoted by V_l^n, where $n \in [1, N]$, $l \in (1, 5T)$. For each of the N objects, a Gaussian Mixture Model(GMM) [16] G_i is built using the vectors corresponding only to that object, V_l^i, $l \in [1, 5T]$. A generic GMM is given by

$$p(x) = \sum_{i=1}^{N} w_i.g(x|\mu_i, \Sigma_i) \tag{5}$$

where, w_i is the weight, μ_i is the mean and Σ_i is the covariance of the i^{th} Gaussian. Since we are modelling data of 7 dimensions, the Gaussians would be 7 variate and hence the means would 7 dimensional as well. The weights W, means μ and covariance matrices Σ of GMM of an object would be estimated by the standard Expectation-Maximization [16] algorithm. After modelling a GMM for each of the objects, we obtain $\mathcal{G} = \{G\}_1^N$. To check, if an object O' is similar to object O_q, its image is used to extract the 82 dimensional feature vector and transform it into the 7 dimensional Principal Component space resulting in a vector V'. The GMM G_q will give the likelihood of V' corresponding to object O_q. Now that we have reduced the search space, a comparison between V', V_l^q vectors corresponding to O_q is made. Since all the vectors V_l^n are spaced at equal intervals of θ, if V' is the closest to V_l^q, its orientation in that view could probably be that of V_l^q.

C. Robust Viewpoint Planning based Object Recognition

1) Object Recognition Method: In this part, we evaluate object recognition efficiency with RGB-D data using BoW model based on two local feature descriptors, namely PFHRGB and SHOTCOLOR available in Point Cloud Library [12]. Kinect 3-D data lacks quality features for small and simple objects like cup, battery, marker, etc. Thus, we adopt the BoW model, which defines the object in terms of feature occurrence statistics.

PFHRGB: It is the colored version of Point Feature Histogram(PFH) [10], encoded from RGB-D information. PFHRGB is binned into a 250-bin histogram(125 Depth+125 RGB). PFH encodes neighbourhood's geometrical properties by generalizing both mean curvature and surface normals.

Fig. 7. (a) Our Small object dataset. (b) Kinect RGB-D Washington dataset(20 objects)

SHOTCOLOR: This descriptor is based on the Signature of Histograms of Orientations(SHOT) [11] descriptor. SHOTCOLOR is binned into 1344 binned histogram(352 Depth+ 992 RGB). The SHOT descriptor obtains a repeatable local reference frame using eigenvalue decomposition around an input point. Additional signature of 9 values encapture local reference frame.

In the pipeline(Figure 6), we use a simple depth based object segmentation technique. The segmentation process involves numerous sub-tasks namely, dominant plane extraction using RANSAC, denoising of the extracted plane and euclidean clustering to extract the dense cloud of the object kept on that plane.

The Object Recognition results based on BoW model claims the novelty of an acceptable object recognition method in terms of time and accuracy. In the results(Table I), we also show performance over various keypoint selection methods on our dataset. (Figure 7(a)).

TABLE I. RECOGNITION ACCURACY WITH VARIOUS KEYPOINTS ON OUR DATASET

Keypoints (with PFHRGB_BoW)	Accuracy(%)	Avg Time (s)	Avg no. of keypoints	Avg no. of total points
SIFT_3D	78.23	0.247	49	1328
Harris_3D	71.4	0.106	42	1328
Subsample-2cm	86.5	0.142	43	1328
Subsample-1cm	91.43	0.143	136	1328

2) Viewpoint Object Potential Map: Further, this part of the work shows that object Recognition results for a particular object vary when viewed in different poses from different distances. This trend is observed due to change in feature quantity and quality of the object cloud captured from different views. Taking a simple example, when objects viewed from narrow side may be easily confused with other objects due to lack of identity features in comparison to other distinct sides.

Accuracy of various simple objects with varied shapes like objects with only single vertical axis of symmetry(Figure 4(b)), objects with only two planes of symmetry (Figure 4(a)), etc. were analysed for various distances and view angles. Thus, Viewpoint Object Potential Map is obtained.

IV. RESULTS

A. Object detection and segmentation

In figure 8, it can be seen, by using a monocular camera, objects of height 2-5cm are classified as non floor, which are then segmented out.

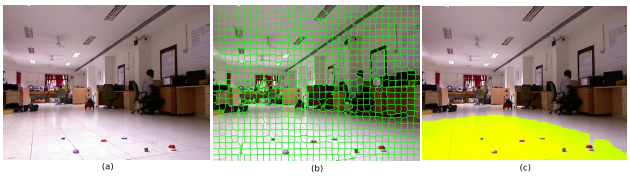

Fig. 8. (a) Typical Scene (b) Superpixeled image. (c) Segmented image.

B. Probabilistic recognition using GMMs

Figure 9 shows the confusion matrix between 13 different objects when classified using GMMs. An object F_i belongs to a class S if the GMM corresponding to class S gives the highest probability for F_i among all the GMMs. A confusion can be seen where the actual object class is 6 and predicted class is 2(Figure 9). This is because of similarity in texture and the outer shape of those objects. If 2 is bigger in size than 6, they would appear similar when 2 is farther than 6 from the camera. Object 12 shows minimal confusion with 2, 3, 4, 6. Object 12 may be similar in shape compared to other objects, but its texture is clearly a differentiating feature. Also, a confusion exists between 11 and 12 due to the texture they share. The significance of GMM-Module in the pipeline is that, even when the objects occupy a few pixels and appear confusing, a probabilistically favourable decision about them can be made. Also, when the objects are far and small, an early weak decision about a certain object's existence proves to be significantly advantageous.

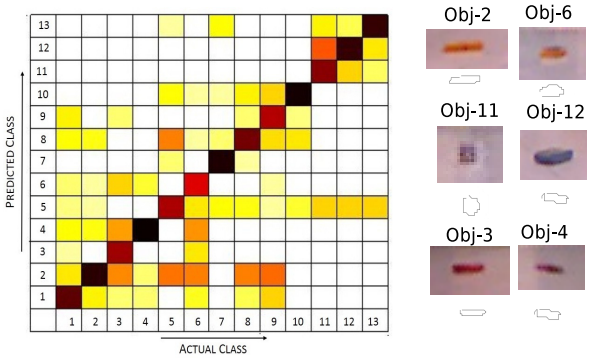

Fig. 9. (a) Confusion matrix for object recognition by GMM-Module

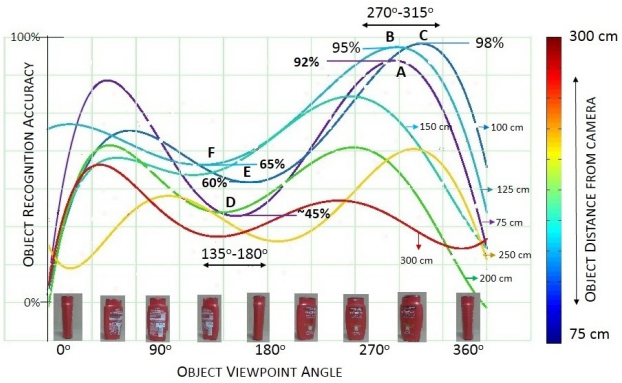

Fig. 10. Recognition accuracy analysis for a particular object (Figure 4(a)).

C. Analysis of viewpoint based recognition

Visual-BOW model based on PFHRGB-BOW descriptors performs better than SHOTCOLOR BOW model, evaluated over Washington dataset(Table III) and our dataset(Table II). Along with PFHRGB-BOW model, subsampled keypoint selection displays better performance over its counterparts(SIFT-3D & HARRIS-3D) on our dataset(Table I).

Even after successful object detection, robot may land up not recognizing the objects even from closer proximity due to weak viewpoint selection. From the experimentation over a particular object(Figure 10), the objects recognition accuracy distributed over various viewpoints is found to be 87.6% for range 75-125 cm, but by using the proposed viewpoint planning method, accuracy level for recognition may boast up to 98% in same settings. Figure 10 shows accuracy as high as 92-98 % from view angle lying between 270°-315°, whereas from view angle 135°-180°, the accuracy levels may plunge as low as 45%-60% for the same object. Therefore, if such object maps(Figure 4) are known prior, the recognition performance of the object categorization algorithms can be enhanced for mobile robots applications.

D. Discussion on Results.

Here, we explain various scenarios where our pipeline comes into play. Thus we focus on the utility of each module.

In Figure 11(a), there are several objects on the floor. There is no prior input from the GMM-Module about the objects. Hence, to search for an object O_q, the robot has to go close to each of the objects aimlessly to recognize them . It covers a lot more distance than it has to cover optimally, to reach O_q. This proves to be expensive.

In Figure 11(b) The utility of GMM-Module(III-B) can be observed here. A prior idea of the objects is acquired before the robot moves towards them. As indicated by GMMs, one of the objects has a high probability of being O_q. Thus, robot moves towards the best viewpoint, obtained form VOP map to recognize it.

In Figure 11(c), the robot examines more than one objects. This happens when the GMM-Module confuses between objects which may appear similar. Here, for two objects, the belief of them being O_q, $P_q \geq P_t$(Figure 2) , due to which, a close examination of each of them should be done. In such a case as well, since we are able to ascertain the beliefs on objects, the search space is thus reduced.

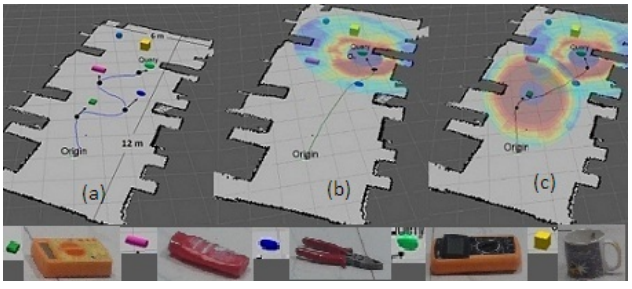

Fig. 11. (a) Object search by greedy approach (b) Path when one object among others has high existential probability. (c) Path when two objects have high existential probabilities.

TABLE II. PERFORMANCE OF RGB-D BASED BOW MODEL, ON OUR SMALL OBJECT DATASET, FIG. 7(A)

Descriptor (Subsample-2cm)	Accuracy(%)	Avg time (s)	Avg no. of Keypoints	Avg no. of total points
PFHRGB_BoW	86.5	0.142	43	1328
SHOTCOLOR_BoW	84.1	0.180	43	1328

TABLE III. PERFORMANCE OF RGB-D BASED BOW MODEL, ON WASHINGTON DATASET (20 OBJECTS), FIG. 7(B)

Descriptor (Subsample-2cm)	Accuracy(%)	Avg time (s)	Avg no. of Keypoints	Avg no. of total points
PFHRGB_BoW	93.75	0.791	78	6302
SHOTCOLOR_BoW	93.2	1.047	78	6302

REFERENCES

[1] A. Aydemir, and P. Jensfelt, *Exploiting and modeling local 3D structure for predicting object locations*, in IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012.

[2] A. Aydemir, A. Pronobis, M. Göbelbecker, and P. Jensfelt, *Active Visual Object Search in Unknown Environments Using Uncertain Semantics*, in IEEE Transactions on Robotics, vol. 29, no. 4, pp. 986-1002, 2013.

[3] T. Garvey, *Perceptual strategies for purposive vision*, AI Center, SRI International, Menlo Park, CA, USA, Tech. Rep. 117, Sep 1976.

[4] K Sjöö, G L Dorian, P Chandana and P Jensfelt and D Kragic *Object Search and Localization for an Indoor Mobile Robot*, Journal of Computing and Information Technology, vol. 17, no. 1, pp. 67-80, 2009.

[5] Hiroaki Masuzawa and Jun Miura, *Observation planning for efficient environment information summarization*, IEEE/RSJ International Conference on Intelligent Robots and Systems, 2009.

[6] K. Sjöö, A. Aydemir, and P. Jensfelt, , *Topological spatial relations for active visual search*, Robotics and Autonomous Systems, vol. 60, no. 9, pp. 1093-1107, 2012.

[7] R. Bajcsy, *Active perception*, Proc. IEEE, vol. 76, no. 8, pp. 9661005, August 1988.

[8] S. Kumar, M. S. Karthik and K. M. Krishna, *Markov Random Field based Small Obstacle Discovery over Images*, in IEEE/RSJ International Conference on Robotics and Automation, 2014.

[9] G. Stein, O. Mano, and A. Shashua, *Vision-based ACC with a Single Camera: Bounds on Range and Range Rate Accuracy.* IEEE Intelligent Vehicles Symposium, 2003.

[10] R. Rusu, N. Blodow, Z. Marton, and M. Beetz, *Aligning point cloud views using persistent feature histograms*, in IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008.

[11] F. Tombari, S. Salti, L. Di Stefano, *Unique signatures of histograms for local surface description*, European Conference on Computer Vision, 2010.

[12] R. Rusu, and S. Cousins, *3D is here: Point Cloud Library (PCL)*, in IEEE/RSJ International Conference on Robotics and Automation, 2011.

[13] G. Csurka, C. Bray, C. Dance, and L. Fan, *Visual categorization with bags of keypoints*, Workshop on Statistical Learning in Computer Vision, ECCV , 1–22, 2004.

[14] M.-K. Hu, *Visual pattern recognition by moment invariants*, IRE Transactions on Information Theory, vol. 8 no. 2, 179-187, 1962.

[15] I. Jolliffe, *Principal Component Analysis*, Springer Verlag, 1986.

[16] A. Jeff Bilmes, *A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models*, Technical Report TR-97-021, ICSI, 1997.

[17] A. Hegerath, T. Deselaers, & H. Ney, *Patch-based Object Recognition Using Discriminatively Trained Gaussian Mixtures*, British Machine Vision Conference, 2006.

[18] J. Shotton, A. Blake, and R. Cipolla, *Multiscale Categorical Object Recognition Using Contour Fragments.*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 7, pp. 1270-1281, 2008.