OBJECT DETECTION FOR SERVICE ROBOT USING RANGE AND COLOR FEATURES OF AN IMAGE

Dipankar Das

Department of Information and Communication Engineering, University of Rajshahi, Rajshahi-6205, Bangladesh

ABSTRACT

In real-world applications, service robots need to locate and identify objects in a scene. A range sensor provides a robust estimate of depth information, which is useful to accurately locate objects in a scene. On the other hand, color information is an important property for object recognition task. The objective of this paper is to detect and localize multiple objects within an image using both range and color features. The proposed method uses 3D shape features to generate promising hypotheses within range images and verifies these hypotheses by using features obtained from both range and color images.

KEYWORDS

Object Detection, Range Image, Generative Learning, Discriminative Learning

1. INTRODUCTION

The field of robot vision is developing rapidly as robots become more capable of operating with people in natural human environments. For robots to be accepted in the home and in offices, they need to identify specific or a general category of objects requested by the user. For this purpose, most of the existing object recognition methods use either color image information [1] or range image information [2]. Although the color information is useful for generic object recognition purposes, the depth information is very useful for robots to accurately localize objects. In this paper we propose a detection and localization technique for robot vision, which integrates both range and color information of images to detect and localize multiple objects simultaneously. Our approach consists of first generating a set of hypotheses for each object using a generative model pLSA [10] with bag of visual words (BOVW) [15] representing 3D shape of each range image. To generate more reliable hypotheses within a range image, we use a singular value decomposition filter on the normal enhanced range images. Then, the discriminative part of our system verifies each hypothesis using an SVM classifier with merging feature that combines both 3D shape and color appearance of the object. The color feature is extracted from the corresponding hypothesis location in the color images.

Most of the previous 3D object recognition methods compare unknown range surface with the models in the database and recognize as the one with the smallest metric distance [3, 4, 14]. These methods require a global model, which needs an extra surface modeling process from multiple range images. In this paper, we propose an alternative solution using both generative and discriminative learning, which measures the similarity between images using local and global feature sets.

DOI : 10.5121/ijcsea.2014.4302

2. RELATED WORKS

Object recognition with local and global feature characteristics has been an active research topic in computer vision community [5]. Recent studies show that local surface descriptors are useful tools for object recognition when occlusion happens [6, 7]. In [6], Li and Guskov proposed a method to find salient keypoints on local surface of 3D objects and similarity has measured using the pyramid kernel function. In [2], Chen and Bhanu have introduced an integrated local surface descriptor for surface representation and 3D object recognition. However, previous methods have been applied for only some 3D free-form objects with a great deal of surface variation in an image without any background clutter. On the other hand, we propose a practical and reliable detection and localization approach that can be operated on simple as well as on well defined 3D objects in a real world background. Moreover, in our approach, we use local keypoints of range images to generate hypotheses of objects and a combination of range and color features to verify each hypothesis. In object extraction, some segmentation approaches have been proposed that utilize range and/or color features of objects [8, 9]. The edge-region-based range image segmentation approach proposed in [9] was limited to only three simple objects and is not applicable for multiple complex objects within an image. On the other hand, in order to extract the desired target object for robots Shibuya *et al*. [8] have proposed a method that uses both color and range images. They use color-distance (CD) matting approach to segment the target object from the scene. The method requires explicit user interaction to indicate the desired target object through a touch sensitive panel. However, our approach automatically detects and localizes all target objects in a scene using both generative and discriminative classifiers without any user interaction.

In object extraction from range images, some segmentation based approaches have been proposed that utilize edge features of objects. The edge-region-based range image segmentation approaches proposed by the authors of [9,16], were limited to only three simple objects and are not applicable to multiple complex objects within an image. A model based edge detection in range images of piled box-like objects using modified scan line approximation technique has been proposed by Katsoulas and Weber [17]. However, all of the previous methods use global edges of 3D range images for objects extraction and obtain good results on some regular objects. Their reliance on global edge properties makes them vulnerable to clutter and occlusion. In this paper, we propose a new method to detect and localize multiple objects in a range image using local features, which can tolerate partial occlusion and cluttered background.

3. OVERVIEW OF THE PROPOSED APPROACH

Our proposed approach for multiple object detection and localization is shown in Figure 1. In the training stage, all the labeled training datasets containing multiple objects per image are presented to the system. In the generative part, the pLSA model [10] is learned for multiple topics using the bag-of-visual-words (BOVW) detected on the 3D range images. At the same time the SVM classifier is learned using both color feature (3D color histogram) and range image's features (BOVW). During the testing phase, when a new test image is given, the system generates a set of promising hypotheses with a bag of visual words using the pLSA model. Then we extract the 3D color histogram from hypothesis locations and combine it with the BOVW. These merging features and their corresponding locations are verified using the multi-class SVM classifier to detect and localize multiple objects within an image.

4. IMAGE DATA COLLECTION

We have constructed the sensor system for obtaining color and range images as shown in Figure 2(a). The acquisition devices consist of a Logicool USB camera for acquiring color images and a Swiss laser ranger (SR- 4000) for obtaining range images. In our system, we have created a common interface for acquiring both range and color images. Images were collected in complex real-world backgrounds that contain multiple objects (Figure 2(b)-(c)). All range images were acquired using the Swiss ranger, SR-4000, at a resolution 176(*horizontal*) × 144(*vertical*) in an office environments.

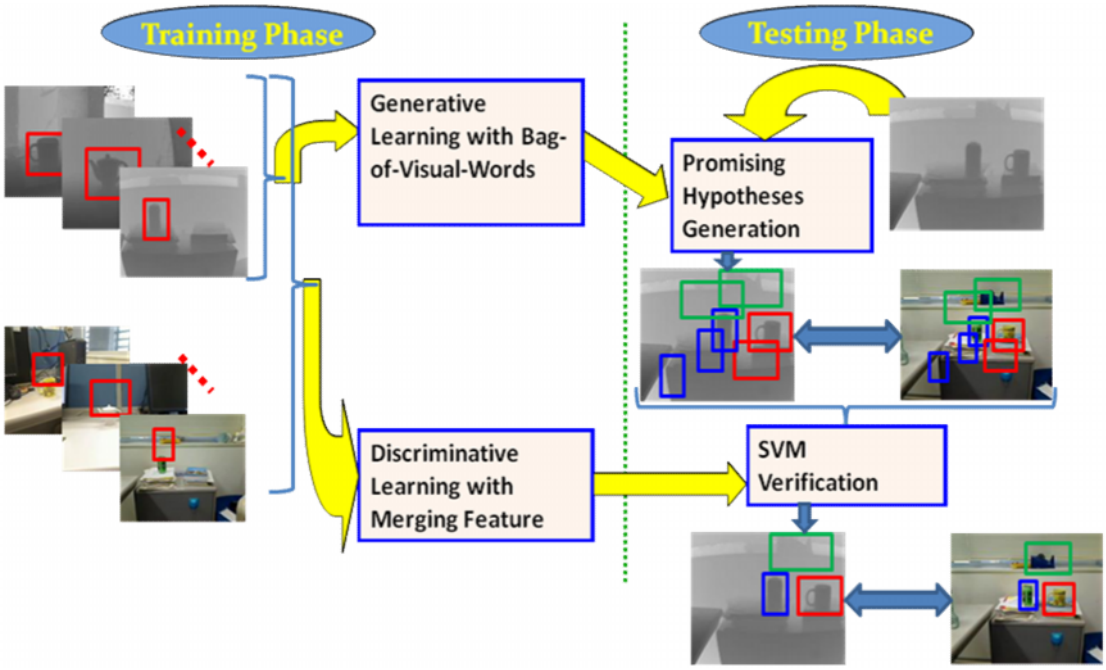


Figure 1: Abstract view of the proposed approach

(a) (b) (c)

Figure 2: Experimental setup and example images: (a) image acquisition devices, (b) distance image, (c) color image.

The resolution of the Logicool color camera was set to the same value as in the lager ranger. All training images were annotated using our annotation tools. We determine the transformed hypotheses locations on color images using the coordinate transformation technique as described in section 5.

5. RANGE IMAGE PRE-PROCESSING

In the images of MESA Swiss ranger, SR-4000, the saturation noise occurred when the signal amplitude or ambient light is too great as shown in Figure 3(a). In this case, the most significant bit (MSB) of the measured pixel is flagged as 'saturated'. This can be avoided by reducing the integration time, increasing the distance or changing the angle of faces of reflective objects or if necessary by shielding the scene from ambient light. However, in an image of a scene that contains multiple objects, it is difficult to completely eliminate saturation noise. In the preprocessing stage, our filtering method removes the saturation noise from the image (as shown in Figure 3(b). Since the saturation noise sometimes occurs contiguously, the conventional averaging or median filter is not appropriate for this purpose. In our filtering approach, instead of searching over all the image regions, we only concentrate on the surroundings of the flagged positions due to saturation noise. If (x, y) be a flagged position due to saturation noise, then our weighted averaging filter of size $m \times n$ is given by:

$$g(x,y) = \frac{\sum_{i=-s}^{s}\sum_{j=-t}^{t} w(i,j) * f(x+i, y+j)}{\sum_{i=-s}^{s}\sum_{j=-t}^{t} w(i,j)}, \qquad \textbf{(1)}$$

where,

$$w(i,j) = \begin{cases} 0 & if \;\; (i,j) - th \;\; pixel \;\; is \;\; flagged \;\; pixel \\ 1 & otherwise \end{cases},$$

$m = 2s + 1$, $n = 2t + 1$, and f is the pixel value at (x, y).

(a) (b)

Figure 3: Image corrupted by saturation noise and result of filtering. (a) Range image with saturation noise, and (b) filtered image.

6. 3D FEATURE EXTRACTION AND GENERATIVE LEARNING

In order to learn the generative model, pLSA, we first compute a co-occurrence table, where each of the range images is represented by a collection of visual words, provided from a visual vocabulary. This visual vocabulary is obtained by vector quantizing keypoint descriptors computed from the training images using the k-means clustering algorithm. To extract visual words, we first determine keypoints on the range image that are insensitive to changes in depth, viewpoint, and rotation. These keypoints are detected on the corner points and uniformly sampled object's edges taking the edge strength as the weight of the sample. For this purpose, we construct 3D edges from range images and use singular value decomposition technique to eliminate noisy edges from the edge map.

6.1. 3D Edge Map Construction

Here we summarize the technique to build the 3D edge map from range images. The edges of 3D images are classified in two main categories [9, 13], representing depth value discontinuities, and surface normal vector discontinuities. The edges corresponding to pixels where depth discontinuity occurs are called jump edges or boundary edges. The second type of edges where surface normal vector discontinuity occurs are called fold edges, or crease edges or roof edges. In this research, edges of 3D range images are formed by combining both jump and fold edges as shown in Figure 4.

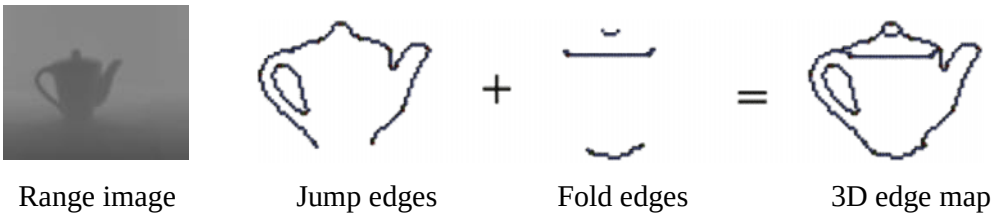


Figure 4: 3D edge map construction from jump and fold edges.

Jump edges are detected using depth discontinuity of the images as shown in Figure 5(b). Suppose N_x, N_y, and N_z be the surface normals for the surface defined by X, Y, and Z coordinates of the 3D range image. These surface normals can be used to enhance a range image in particular the edges. An image represented by the normal component is referred to as the normal enhanced image. For example, Figure 5(c) shows the normal enhanced image represented by the normal component N_y. Fold edges are detected using normal enhanced images with surface normal discontinuity. However, the normal-based approach is sensitive to noise in the range data and may corrupt the object surface while enhancing the edges. Thus, we use a singular value decomposition filter (SVDF) to generate more reliable noise free fold edges.

Figure 5: 3D edge map for the object teapot. (a) An example range image represented by the depth values, (b) detected jump edges, (c) image represented by the normal component N_y, (d) N_y edge map, (e) N_y edge map after applying SVDF on (c), and (f) combined 3D edge map (b + e).

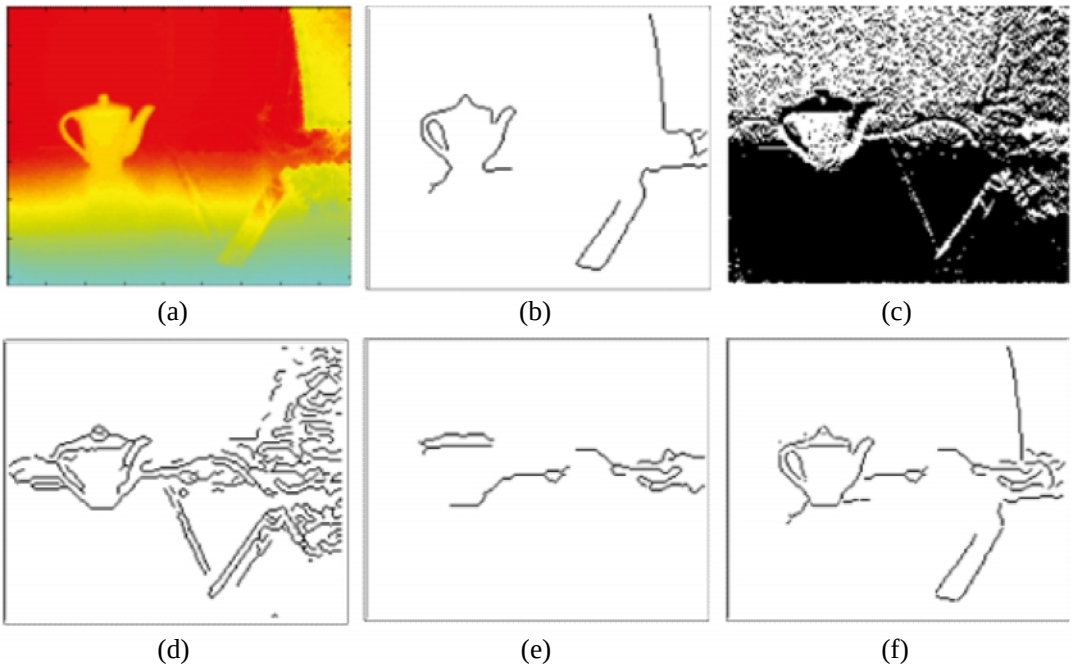


Figure 6: 3D edge map for the object teapot. (a) An example range image represented by the depth values, (b) detected jump edges, (c) image represented by the normal component N_y, (d) N_y edge map, (e) N_y edge map after applying SVDF on (c), and (f) combined 3D edge map (b + e).

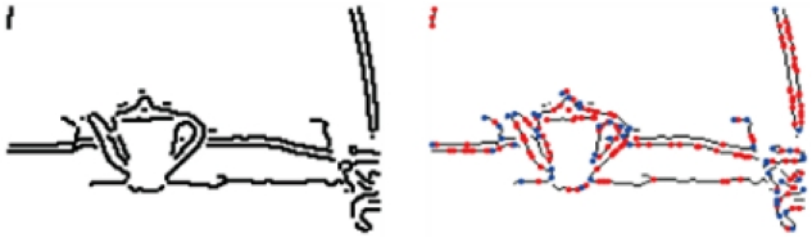
6.2. Singular Value Decomposition Filter (SVDF)

When we use a normal enhanced image to determine the edges, the edge map includes a great deal of false edges as shown in Figure 5(d). In order to eliminate this effect, we use SVDF on the normal enhanced images. To construct SVDF, we first create a normal enhanced image from the range data and obtain the singular values $S \in \{s_1, s_2, \cdots\cdots, s_k\}$. Since the resolution of the range sensor is 176 (*horizontal*) × 144 (*vertical*), the matrix of normal enhanced image is of size 144 (*row*)×176 (*column*) and has full rank due to noise. Thus we have 144 singular values giving k = 144 in our experiment. The singular value decomposition technique produces singular values in decreasing order: $s_1 > s_2 > \cdots\cdots > s_k$. Then we normalize s_i by s_i / s_1. The filtered image is formed by taking the first n singular values such that $s_n \geq \delta$. It can be shown that $\delta = 0.1$ produces approximately noise free fold edges for our experiments (Figure 5(e)).

6.3. Keypoint Detection and Description

For 3D edge map distinct corner points are detected using the corner detection algorithm and all of the corner points are selected as a set of keypoints (n_1). The second set of keypoints (n_2) is obtained by sampling the 3D edge map. We use weighted uniform edge sampling technique to determine another set of the keypoints (n_2). Our total number of keypoints is $n = n_1 + n_2$. Figure 7 shows detected keypoints on the image, where blue and red points indicate the corner and edge sampling keypoints, respectively. Each of the generated keypoint is described by: (i) the normal components, and (ii) the orientation of edge map within the patch. The normal components N_x, N_y, and N_z of the 3D surface normal are defined by X, Y , and Z coordinates of the range image. The descriptor is a $N \times N$ square patch around each keypoint pixel for all of the three normal

components and edge orientations. Thus, our descriptors represent the local 3D surface patch and edge orientations around each of the keypoint. It is row reordered to form a vector in $4N^2$ dimensional feature space. The patch size tested for our experiment is $N = 5$ and 7.



3D edge map Detected keypoints

Figure 7: Detected edges and keypoints on a range image.

6.4. Learning the pLSA Model

To learn generative model pLSA, we construct a codebook of feature types by clustering a set of training keypoints according to their descriptors. For clustering, we use *k*-mean clustering algorithm with codebook size of 300. The codebook is necessary to generate a BOVW histogram for each training or testing image. Then the pLSA model is learned for a number of topics given to the system using BOVW histograms constructed from the training dataset. The model associates each observation of a visual word, *w*, within an object, *o*, with a topic variable $z \in Z = \{z_1, z_2, \cdots\cdots, z_k\}$. Here, our goal is to determine $P(w|z)$ and $P(z|o)$ by using the maximum likelihood principle. The model is fitted for the entire training object without knowledge of labels of bounding boxes. Then the topics are assigned based on the object specific topic probability, $P(z_k \mid o_j)$ under each category.

7. COLOR FEATURE EXTRACTION AND DISCRIMINATIVE LEARNING

In our learning approach, along with pLSA, a multiclass SVM classifier is also learned using both 3D shape and color features. The 3D shape is represented by the histogram of visual words extracted from the range image. Although shape representation is a good measure of objects similarity for some objects, shape features are not sufficient enough to distinguish among all types of objects. In this case, color appearance is a better feature to find the similarity between them.

7.1. Color Feature Extraction

In order to determine color feature from color image, we compute 3D HSV global color histograms from all training images collected from different environments in varying lighting conditions. In the HSV color space, the quantization of hue requires the most attention. The hue circle consists of the primary colors red, green and blue separated by 120°. A circular quantization by 20° steps sufficiently separates the hues such that the three primaries and yellow, magenta, and cyan are represented each with three sub-divisions. The saturation and value are each quantized to three levels yielding greater perceptual tolerance along these dimensions. Thus *H* is quantized to 18 levels, and *S* and *V* are quantized to 3 levels. The quantized HSV space has $18 \times 3 \times 3 = 162$ histogram bins. The final histogram is normalized to sum to unity in order to fit appropriately in our SVM kernel.

7.2. SVM Learning

We use the SVM to learn the discriminative classifier using merging feature. The combination of both shape and color appearance features for an image *I*, are merged as:

$$H(I) = \alpha H_S(I_R) + \beta H_A(I_C), \quad (2)$$

where both α and β are weights for the shape histogram, $H_S(I_R)$ and color appearance histogram, $H_A(I_C)$, respectively. The multi-class SVM classifier is learned using the above merging feature giving the higher weight to the more discriminative feature. We use the LIBSVM [11] package for our experiments in a multi-class mode with the *rbf* exponential kernel.

8. HYPOTHESIS GENERATION AND VERIFICATION

In recognition stage, promising hypotheses are generated within the range image for probable object's locations. For this purpose, visual words are extracted from the range image using the same technique as described in subsection 4.3. Each visual word is classified under the topic with the highest topic specific probability $P(w_i \mid z_k)$. Then promising hypotheses are generated for each of the object using the hypothesis generation algorithm of Das *et al.* [12]. In the verification step, merging features are extracted from the regions of the image bounded by the windows of promising hypotheses. The shape feature is obtained from the generated hypothesis area on the range image. However, in order to obtain the color feature the corresponding hypothesis location on the color image is detected using the following coordinate transformation

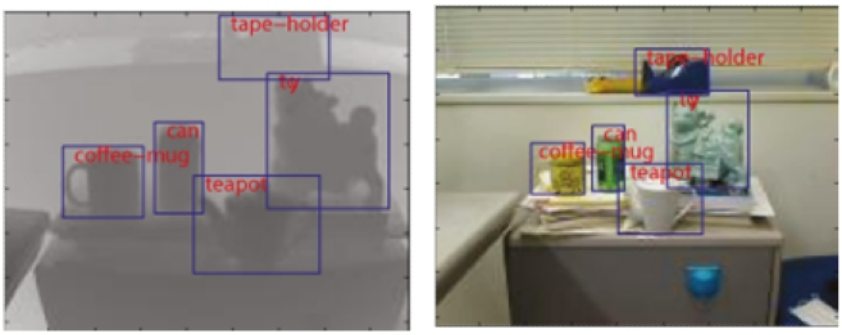
algorithm:

1. Using the ground truth bounding boxes of both range and color images of the training data, calculate the *X* and *Y* coordinates of the range image for which *X* and *Y* coordinates of the corresponding color image are equal and denote these coordinate by X_{CR} and Y_{CR}, respectively.
2. From both range and color images, determine the overlapping starting coordinates of the color image with the range image and denote these coordinates by X_{SC} and Y_{SC}.
3. Calculate the hypothesis coordinates location on the color image (X_{HC}, Y_{HC}) using the hypothesis coordinate of the range image (X_{HR}, Y_{HR}) as:

if $X_{HR} \leq X_{CR}$ **then**
$X_{HC} \leftarrow X_{HR} + X_{SC} - (X_{SC}/X_{CR}) * X_{HR}$
else
$X_{HC} \leftarrow X_{HR} - (X_{SC}/X_{CR}) * (X_{HR} - X_{CR})$
end if

4. Similarly calculate the Y_{HC} coordinate for hypothesis location of the color image.

The above algorithm determines the hypothesis windows on the color image using the detected hypothesis windows on the range image as shown in Figure 8. Then, the merging features are extracted from both range and color images and fed into the multi-class SVM classifier in the recognition mode. Only the hypotheses for which a positive confidence measurement is returned are kept for each object. Objects with the highest confidence level are detected as the correct objects. The confidence level is measured using the probabilistic output of the SVM classifier.



(a) (b)

Figure 8: (a) Detected regions on a range image, and (b) corresponding regions on a color image.

9. EXPERIMENTAL RESULTS

The proposed technique has been used to carry out the detection and localization of multiple objects within the range image. Since there is no standard database suitable for color-distance feature based object recognition, we evaluated our system using our own database.

9.1. Database

Our color-distance database consists of images of ten specific objects grouped into two datasets. The dataset-1 comprises of 251 images of five objects, namely, coffee-jar, spray-cleaner, toy-1, video-camera, and stapler. Among 251 images 150 images (single object per image) are used for training the model and the rest 101 images of 312 objects are used for testing the system. The dataset-2 contains 230 images of another five objects, namely, coffee-mug, tape-holder, toy-2, teapot and can. Among 230 images 150 images are again used for training the system and the remaining 80 images are used for testing purposes.

9.2. Classification Results

We performed experiments to detect and localize multiple objects against complex real-world background. In the training period, both the pLSA and SVM models are fitted for all training objects. During recognition stage, given an unlabeled image of multiple objects, our main objective is to automatically detect and localize all objects within the image. In our approach object presence detection means determining if one or more objects are present in an image and localization means finding the locations of objects in that image. The localization performance is measured by comparing the detected window area with the ground truth object window. Based on the object presence detection and localization, an object is counted as a true positive object if the detected object boundary overlaps by 50% or more with the ground truth bounding box for that object. Otherwise, the detected object is counted as false positive. In the experimental evaluation, the detection and localization rate (DLR) is defined as:

$$DLR = \frac{\#of\,true\ positive\,objects}{\#of\ annotated\ objects} \quad \textbf{(3)}$$

$$FPR = \frac{\#of\ false\ positive\,objects}{\#of\ annotated\ objects} \quad \textbf{(4)}$$

Table 1. Experimental results on dataset-1.

Objects	**Detection rate for merging feature without SVDF**		**Detection rate for merging feature with SVDF**	
	DLR	**FPR**	**DLR**	**FPR**
Coffee jar	0.72	0.36	0.77	0.44
Spray cleaner	0.88	0.13	0.90	0.15
Toy-1	0.95	0.00	0.98	0.03
Video camera	0.69	0.31	0.79	0.40
Stapler	0.65	0.26	0.70	0.31
Average Rate	***0.78***	***0.21***	***0.83***	***0.27***

Table 2. Experimental results on dataset-2.

Objects	**Detection rate for merging feature without SVDF**		**Detection rate for merging feature with SVDF**	
	DLR	**FPR**	**DLR**	**FPR**
Coffee mug	0.74	0.05	0.79	0.13
Tape holder	0.74	0.26	0.79	0.23
Toy-2	0.97	0.62	0.97	0.70
Teapot	0.67	0.43	0.80	0.35
Can	0.84	0.14	0.96	0.12
Average Rate	***0.79***	***0.30***	***0.87***	***0.30***

Tables 1~2 show the detection and localization rates on our datasets for the ten objects. In the experiments, the effect of Singular Value Decomposition Filter (SVDF) on the normal enhanced image is also evaluated in terms recognition rate. It can be shown that the proposed method is able to produce the average recognition accuracy of 78% and 79% for dataset-1 and dataset-2, respectively. However, the method generates more accurate hypothesis on the image using the SVDF. As a consequence, final recognition rate has significantly increased. On dataset-1 and dataset-2 the final recognition accuracy are 83% and 87%, respectively.

The detection and localization results on datasets are shown in **Figure 9**. The blue rectangular bounding box indicates the location of each object within an image. The first column of the **Figure 9** shows the detected target objects with their locations on the depth images and the second column illustrates the corresponding object's locations on the color images. The locations on the color images are detected by our algorithm as discussed in section 5.

10. CONCLUSIONS

Our system has shown the ability to accurately detect and localize multiple objects using both range and color information. For this purpose, in the experiment, an optical camera and a laser ranger are integrated. Our method is useful for service robots, because it can use 3D information to know the exact position and pose of the target objects. Our local descriptors on edges generate accurate hypotheses, which are then verified using more discriminative features with chi-square merging kernel. Since we use scale invariant local properties, it is robust for partial occlusion, background clutter, and significant depth changes. The detection rate and computational efficiency suggest that our technique is suitable for real time use.

Current range sensors have problems with transparent and reflecting materials. Although our approach can remove saturation noise due to reflection properties, our system does not work well for transparent objects. In future, we will explore the possibility of detecting transparent objects by integrating both range and color sensors, and using a combination of more robust features from both sensors.

REFERENCES

[1] Z. Stefan and M. Manuela (2006), "Detection and Localization of Multiple Objects", In *Proc. Humanoids*, Genoa, Italy.

[2] H. Chen and B. Bhanu (2007), "3D Free-Form Object Recognition in Range Images using Local Surface Patches", *Pattern Recognition Letters*, Vol. 28, No. 10, pp. 1252–1262.

[3] A.S. Mian, M. Bennamoun and R.A. Owens (2006), "Three-Dimensional Model-Based Object Recognition and Segmentation in Cluttered Scenes", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 28, No. 10, pp. 1584-1601.

[4] A.E. Johnson and M. Hebert (1999), "Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 21, No. 5, pp. 433-449.

[5] R.J. Campbell and P.J. Flynn (2001), "A Survey of Free- Form Object Representation and Recognition Techniques", *Computer Vision and Image Understanding*, Vol. 81, No. 2, pp. 166-210.

[6] X. Li and I. Guskov (2007), "3D Object Recognition from Range Images using Pyramid Matching", In *Proc. ICCV*, Rio de Janeiro, Brazil, pp. 1-6.

[7] A. Frome, D. Huber, R. Kolluri, T. B¨ulow and J. Malik (2004), "Recognizing Objects in Range Data using Regional Point Descriptors", In *Proc. ECCV (3)*, Prague, Czech Republic, pp. 224-237.

[8] N. Shibuya, Y. Shimohata, T. Harada, and Y. Kuniyoshi (2008), "Smart Extraction of Desired Object from Color-Distance Image with User's Tiny Scribble", In *Proc. IROS*, Nice, France, pp. 2846–2853.

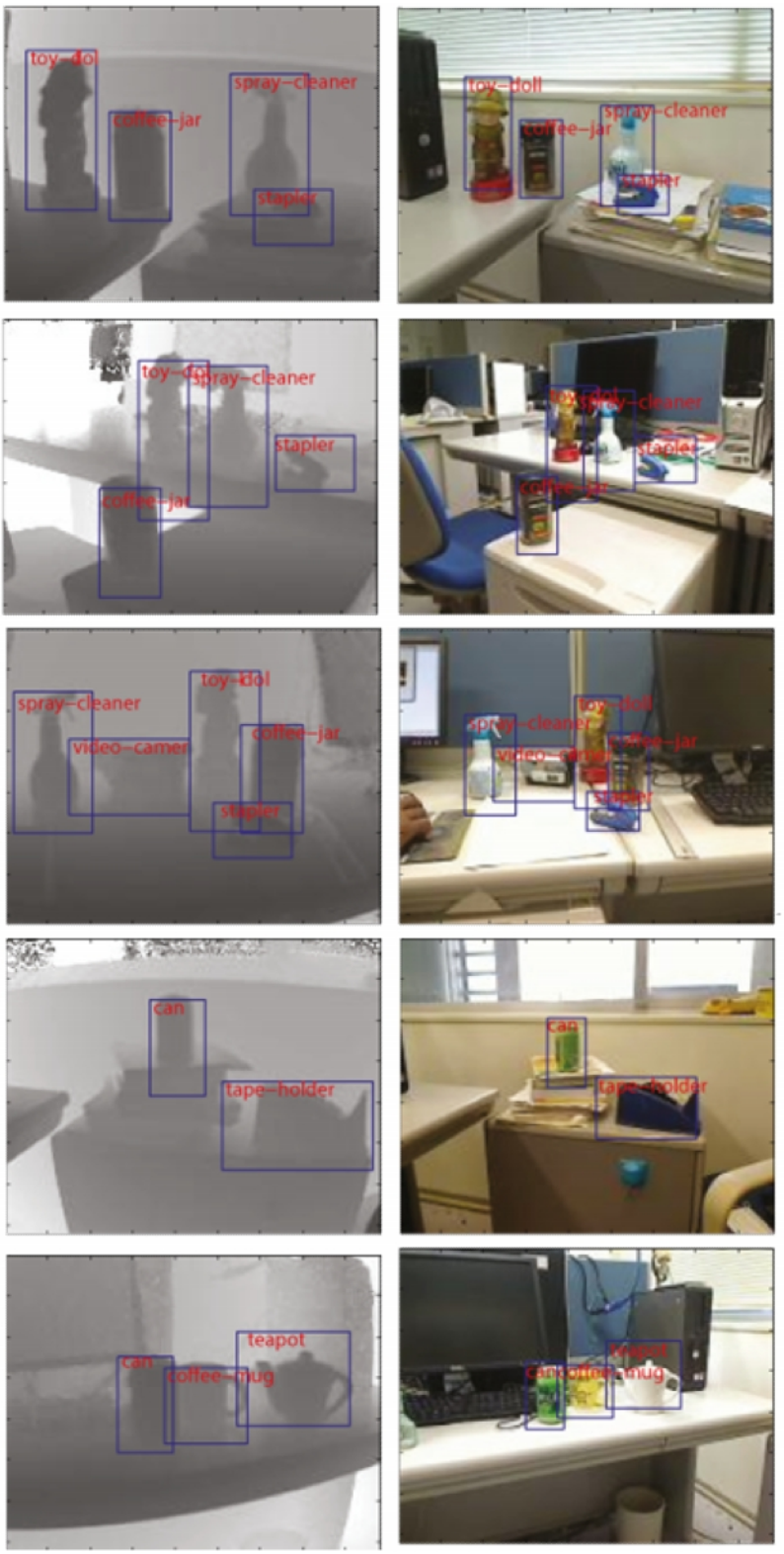


Figure 9: Detection and localization results on range and color images.

[9] M.A.Wani and B.G. Batchelor (1994), "Edge-Region-Based Segmentation of Range Images", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 16, No. 3, pp. 314-319.
[10] T. Hofmann (2001), "Unsupervised Learning by Probabilistic Latent Semantic Analysis", *Machine Learning*, Vol. 42, No. 1/2, pp. 177–196.
[11] C.C. Chang and C.J. Lin (2008), "Libsvm: A Library for Support Vector Machines", http://www.csie.ntu.edu.tw/cjlin/libsvm/.
[12] D. Das, A. Mansur, Y. Kobayashi, and Y. Kuno (2008), "An Integrated Method for Multiple Object Detection and Localization", In *Proc. ISVC*, Nevada, USA, pp. 133-144.
[13] T.-J. Fan, G. G. Medioni, and R. Nevatia (1989), "Recognizing 3-d objects using surface descriptions", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 11, No.11, pp. 1140–1157.
[14] E. R. van Dop and P. P. L. Regtien (1996), "Object recognition from range images using superquadric representations", In *IAPR Workshop on Machine Vision Applications*, pp. 267–270, Tokyo, Japan.
[15] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray (2004), "Visual categorization with bags of keypoints", In *European Conf. on Computer Vision, Workshop on Statistical Learning in Computer Vision*, Prague, Czech Republic.
[16] A. K. Sood and E. Al-Hujazi (1990), An integrated approach to segmentation of range images of industrial parts. In *IAPR Workshop on Machine Vision Applications*, pages 27–30, Kokubunji, Tokyo, Japan.
[17] D. Katsoulas and A.Werber (2004), Edge detection in range images of piled box-like objects. In *International Conference on Pattern Recognition*, pages 80–84, Cambridge, England, UK, IEEE Computer Society.

Authors

Dipankar Das received his B.Sc. and M.Sc. degree in Computer Science and Technology from the University of Rajshahi, Rajshahi, Bangladesh in 1996 and 1997, respectively. He also received his PhD degree in Computer Vision from Saitama University Japan in 2010. He was a Postdoctoral fellow in Robot Vision from October 2011 to March 2014 at the same university. He is currently working as an ass ociate professor of the Department of Information and Communication Engineering, University of Rajshahi. His research interests include Object Recognition and Human Computer Interaction.