JAREE-Journal on Advanced Research in Electrical Engineering
Volume 4, Number 1, April 2020

Object Extraction Using Probabilistic Maps of Color, Depth, and Near-Infrared Information

Muhammad Attamimi, Kelvin Liusiani, Astria Nur Irfansyah, Hendra Kusuma, and Djoko Purwanto

***Abstract*—Object extraction is one of the important and challenging tasks in the computer vision and/or robotics fields. This task is to extract the object from the scene using any possible cues. The scenario discussed in this study was the object extraction which considering the Space of Interest (SOI), i.e., the three dimensional area where the object probably existed. To complete such task, the object extraction method based on the probabilistic maps of multiple cues was proposed. Thanks to the Kinect V2 sensor, multiple cues such as color, depth, and near-infrared information can be acquired simultaneously. The SOI was modeled by a simple probabilistic model by considering the geometry of the possible objects and the reachability of the system acquired from depth information. To model the color and near-infrared information, a Gaussian mixture models (GMM) was used. All of the models were combined to generate the probabilistic maps that were used to extract the object from the scene. To validate the proposed object extraction, several experiments were conducted to investigate the best combination of the cues used in this study. The best result was achieved from the model with combination of depth, near-infrared, and color information in HSV color space; the results of recall, precision, and F1-score were 83.58%, 88.40%, and 84.96% respectively.**

***Keywords*—object extraction, probabilistic maps, color information, depth information, near-infrared information.**

I. Introduction

Recently, the study of intelligent systems and/or machines including intelligent robots has developed rapidly. One of the important abilities of intelligent robot is the ability to recognize their environments. There were many approaches that have been made to realize such ability, one of them known as and developed under the field of visual recognition. The problem discussed in such field includes object detection, object recognition, and so on as proposed in [1], [2], [3], [4], [5], [6], [7], which were related to one another. For example, to complete the object recognition task of a given scene, the system needs to know where the target object existed in the scene. Such a task is known as object detection which outputs the location of the target object. Object detection is the first step in realizing the object recognition in a scene. One of the approaches to detect the object is by applying object extraction.

Generally, the object extraction is a task to localize and extract the object from a given scene. This task is difficult because we need to know which parts become a non-object and which parts become an object. In the object extraction of an image, we need to interpret each pixel of becoming an objects pixel or not. The result of such interpretation can be a category of the objects pixel and non-objects pixel. There were many algorithms existed to solve such categorization problem. In this study, a probabilistic approach is adopted to solve the problem.

Another problem to be considered in object extraction task is the use of the cues. Many previous studies have dealt with object extraction task by exploting the color information [1], [2]. However, several studies have been shown that in many visual recognition tasks the use of multiple cues was better than the use of single cue such as color information [3], [4], [5], [6], [7]. Therefore, the use of multiple cues is considered in this study. Thanks to the Kinect V2 sensor [8], multiple cues consisted of color information, depth information, and near-infrared information can be acquired simultaneously.

In addition, there are many scenarios in object extraction. In this study, the scenario of object extraction which considering the Space of Interest (SOI) is introduced. The SOI is the three dimensional area (space) where the object probably existed. Here, the introduced SOI is modeled by a simple probabilistic model by considering the geometry of the possible objects and the reachability of the system estimating from the depth information captured from a Kinect V2 sensor. This scenario is particularly more effective if the robot is involved. For example, the robot has physical embodiments which can express its reachability of the object in a space. We can use such information to extract the object inside the SOI. Of course the object acquisition can also be done effectively by implementing SOI. This study considers such scenario by assuming the Kinect V2 sensor as the robots eyes and the area on the table top is the possible SOI of such assumed robot.

The use of depth information is more powerful when combined with color information and near-infrared information captured simultaneously from the Kinect V2 sensor. The appearance of the objects is represented by the color information. In this paper, two color spaces, that are Red, Green, and Blue (RGB) color space, and Hue, Saturation, and Value (HSV) color space are used and compared. According to [4], [9], near-infrared can be used as a material information. Therefore, the combination of color information and near-infrared information is important as the cues used for object extraction. A Gaussian mixture models (GMM) is used to model the color information and near-infrared information. Thanks to probabilistic models, the probabilistic maps of color, depth, and near-infrared information can be generated. The probabilistic maps are integrated by considering their independencies. The pixels that have probability more than a predetermined value are considered to be the objects pixels. We can extract the object by smoothing the objects pixels.

Muhammad Attamimi, Kelvin Liusiani, Astria Nur Irfansyah, Hendra Kusuma, and Djoko Purwanto are with the Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia e-mail: attamimi@ee.its.ac.id.

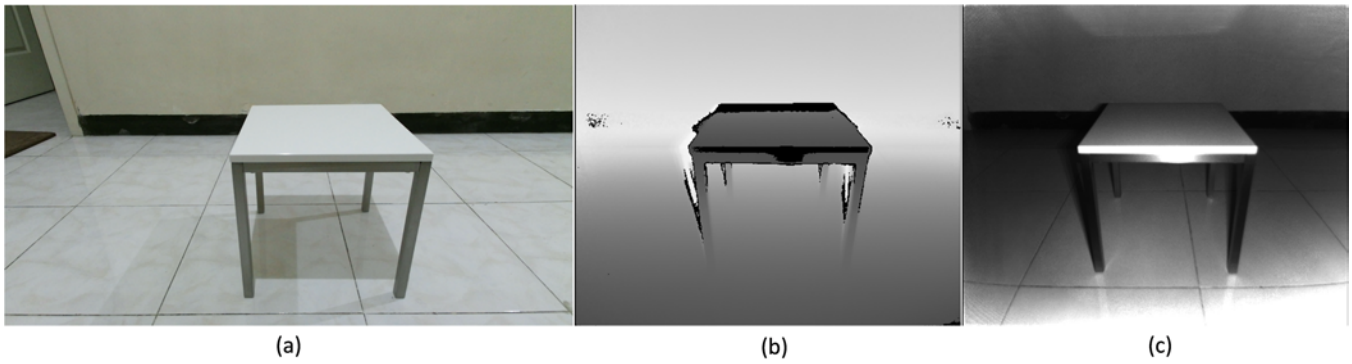

Fig. 1. Examples of captured data from a Kinect V2 sensor: (a) color information, (b) depth information, and (c) near-infrared information.

There are several works related to this study, i.e., object extraction ([4], [10], [11]), object detection ([1], [2], [3], [4], [5], [12], [13], [14]), object recognition ([4], [6], [7]), the use of near-infrared information ([4], [9], [15]), and the use of object extraction in data acquisition system ([16], [17]).

There are several data acquisition systems that are done manually and some automatically. In [16], color and depth information were captured using the first version of Kinect camera (Kinect V1) [18]. However, due to the limitation of the sensor, near-infrared information was not able to be captured simultaneously. To this end, we use the second version of Kinect sensor (Kinect V2) [8] in this study. Of course the approach used to extract the objects was also different with the proposed method.

The remainder of this paper is organized as follows. An overview of the proposed method followed by the details of each process is presented in section 2. Next section discusses the experimental setting and the results. Finally, section 4 concludes this study.

II. Proposed Method

In this study, a Kinect V2 sensor is used because simultaneous data that consists of color information, depth information, and near-infrared information, can be acquired from the sensor. The information acquired from the sensor is needed to build the object extraction system. Examples of the data captured from Kinect V2 shown in Figure 1. One can see from the figure, the color information is acquired in the form of color image in 1920 x 1080 resolution (see Figure 1 (a)), whereas both of the depth information and near-infrared information are required in the form of one channel of 16-bit image in 512 x 424 resolution.

Since the Kinect V2 sensor consists of two different sensors that are CCD camera and time-of-flight camera, there are two problems need to be considered. The first problem is the differences in the image resolution. Instead of compensating the depth information, we map the color information to the depth information to realize a simple and essential data (color, depth, and near-infrared) integration. We can use Kinect For Windows SDK [20] to calculate the corresponding point between the color space and depth space. The second problem

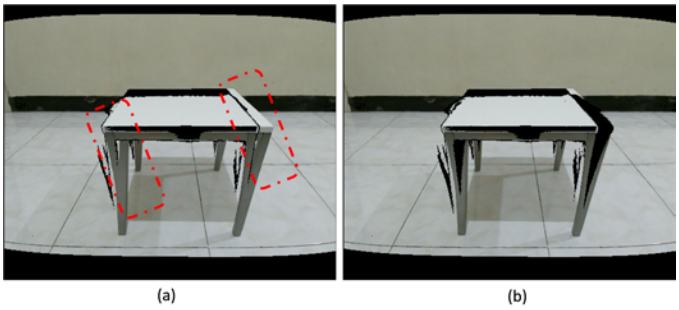

Fig. 2. Color mapping results: (a) without occlusion region removal (regions consisted of pseudo-color showed in red rectangles), and (b) with occlusion removal.

is the occluded regions occurred. Such regions cause a psedo-colors (see Figure 2 (a)) which lead to false recognition [21]. In [21], the occluded regions are detected and compensated because two CCD cameras are utilized. However, if we use one Kinect V2 sensor only, there is no way to compensate the occluded regions. Yet, we can adopt the algorithm in [21] to remove the occluded regions. The result of mapped color image is depicted in Figure 2 (b).

The architecture of proposed object extraction system is illustrated in Figure 3. One can see from the figure that the input data of the proposed system is the images captured from the Kinect V2 sensor which consists of color, depth, and nir-infrared image. The input images are then processed. First, the captured depth information is processed to create a three dimensional points in the camera coordinate system using the Kinect For Windows SDK [20]. The results are then transformed into world coordinate system by applying transformation matrix to each three dimensional points. The transformation matrix is calculated by considering the angle that are pan and tilt of the camera which can be measured easily; and the height of the camera from the ground. It should be noted that another two axes except the height are considered similar to the ones in the camera coordinate. After all the points transformed into world coordinate, a SOI filtering can be applied to output the probabilistic map of SOI.

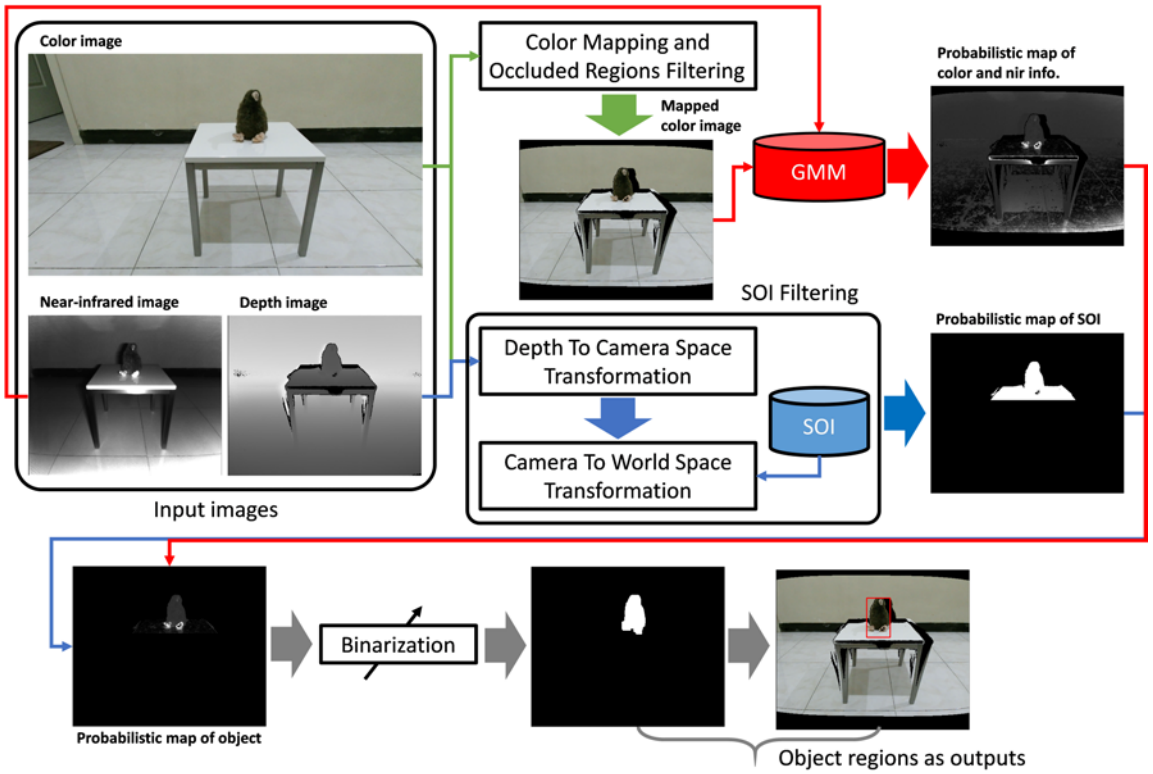

Fig. 3. Overview of proposed object extraction.

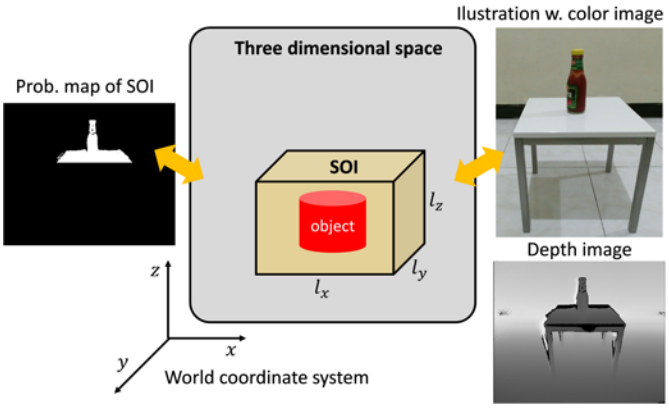

Fig. 4. The illustration of SOI used in this study.

The SOI illustrated in Figure 4, is three dimensional space where each possible location is a candidate for the object to exist in the scene. Considering the geometry of the objects and possible space in the world coordinate system (i.e., the block space with size $l_x \times l_y \times l_z$), the SOI can be modeled. Here, the object existed in the household environment is considered. Combining with the prior knowledge that an object is laid on the table top (see Figure 4), we can model the SOI using a simple probabilistic map. Given a depth image, the transformation into a world coordinate system can be done by calculating the corresponding position of the Kinect V2 sensor to a predetermined origin point. In this study, the origin point is chosen coaxis with the camera coordinate system except the height of the camera. By providing the pan, tilt, and height of the camera, the transformation matrix can be calculated. The matrix is then applied to the whole points captured from the sensor. Here, the SOI is modeled by assuming that each point $w = (x, y, z)$ in the world coordinate system.

Next, to create a mapped color image, color mapping and occluded regions filtering are needed. To map the color information into depth space, a Kinect For Windows SDK [20] is used. To deal with the problem of occlusion, the algorithm proposed in [21] is implemented on the proposed object extraction. The mapped color image along with the near-infrared is then inputted to the GMM for learning the model and for estimating the probabilistic map of color and near-infrared information. In the learning phase, the table top pixels are extracted by considering a simple geometry on three dimensional points that have been transformed into world coordinate system. The extracted pixels than inputted to the GMM that initialized by a k-means algorithm [22]. Once the model is learned, we can use it to calculate the likelihood of an input vector.

Once the probabilistic maps of captured images are generated, the probabilistic map of the object can be calculated. Finally, after a binarization process, the object region can be

estimated. It should be noted that a rectangle of the object can also be calculated by finding the location of the pixels in the object region which locate in the most left, top, right, and bottom.

III. Experiments

In this study, several experiments regarding the use of cue (i.e., depth information, color and near-infrared information, the variation on choosing the color space) were conducted. The details of experimental settings and the results are described as the following.

A. Experimental Settings

In this paper, the experiment was carried out in the living room shown in Figure 5 (a) using 12 objects depicted in Figure 5 (b). Each object consisted of eight data frames which total of 96 data frames was collected. Each frame consisted of the color image in resolution of 1920 x 1080; depth image and near-infrared image, both of them in a resolution of 512 x 424. The collected data then was processed.

For comparison, two color spaces, i.e. RGB color space and HSV color space were used. The learning phase of SOI and the GMM of each color space were conducted. Therefore, there were three probabilistic models calculated of each given data frame. In this study, we compared the use of cues in the form of: 1) depth information only, 2) depth information combined with color in RGB color space and near-infrared information, and 3) depth information combined with color in HSV color space and near-infrared information. It should be noted that the binarization process of all probabilistic maps of the object was the same.

The evaluation metrics used in this study were recall, precision, and F1-score. To calculate the metrics, the ground truth of each objects region was necessary. We labelled manually the correct objects rectangle in mapped color image and calculated the metrics.

TABLE I. Recall, precision, and F1-score of each used cues.

The cues used	Recall [%]	Precision [%]	F1-score [%]
Depth only	**96.35 ± 4.07**	21.11 ± 8.06	33.90 ± 9.95
Depth + GMM-RGB	94.48 ± 7.27	50.99 ± 16.97	64.36 ± 13.29
Depth + GMM-HSV	83.58 ± 14.65	**88.40 ± 8.88**	**84.96 ± 9.75**

B. Experimental Results

First, the evaluation metrics were calculated for each combination of cues, and the results were listed in Table 1. One can see from the table that by introducing the SOI, the recall of the object was high. This indicated that the extraction of the object can be done in high rate of recall. However, the precision was low if the depth information only was used. Table 1 illustrated that the use of multiple cues made the proposed method more precise. Among the models used in this study, the model with the combination of SOI and a GMM modeled using the HSV color information and near-infrared information was achieved the best result compared to the others. This result indicated that the use of HSV color space can handle the illuminating changes and shadows occurred on the object region.

Next, the evaluation metrics of each object using the best model was depicted in Figure 6. One can see from the figure that object #11 achieved the worst result. This result was caused by many missing pixels due to the shape and the size of the objects. However, if we used The PASCAL Visual Object Challenge (VOC) evaluation metric [23] which considers the correct extraction when more than half intersection between the rectangle outputted from the ground truth and the one outputted from the proposed system achieved, the rate of object detection was 100%. Finally, some examples of the object extraction of scenes is illustrated in Figure 7. It can be seen that the use of HSV color space provided a better extraction output compared to the other ones.

Fig. 5. Experimental settings: (a) a living room with the table to put the object and a Kinect V2 sensor, and (b) household objects used in this study.

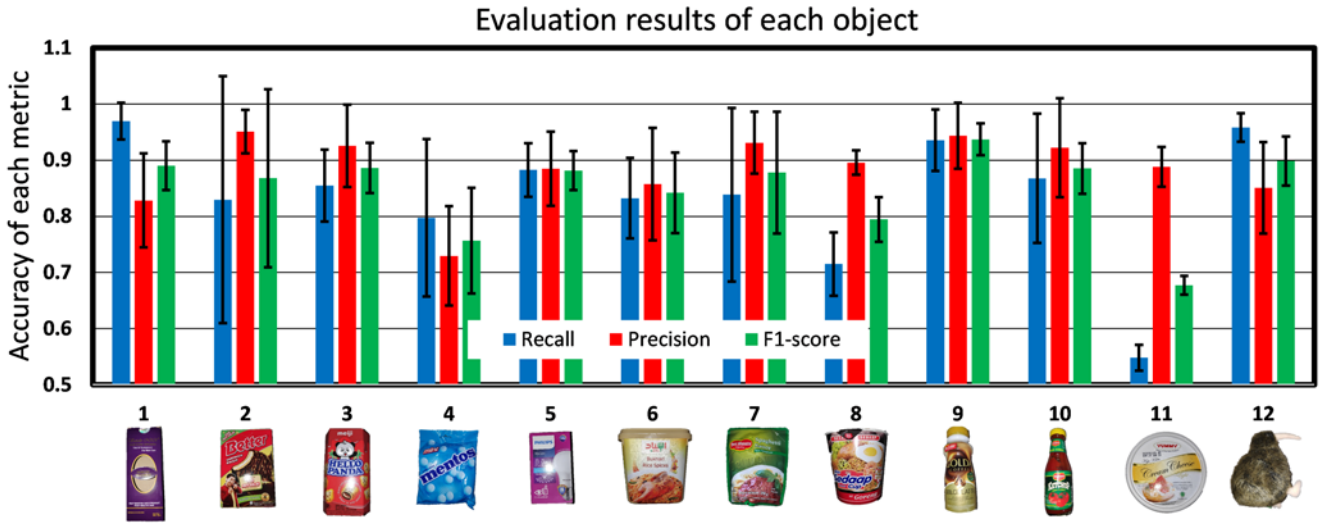

Fig. 6. The evaluation results of object extraction using the combination of SOI and a GMM of color information in HSV color space and near-infrared information.

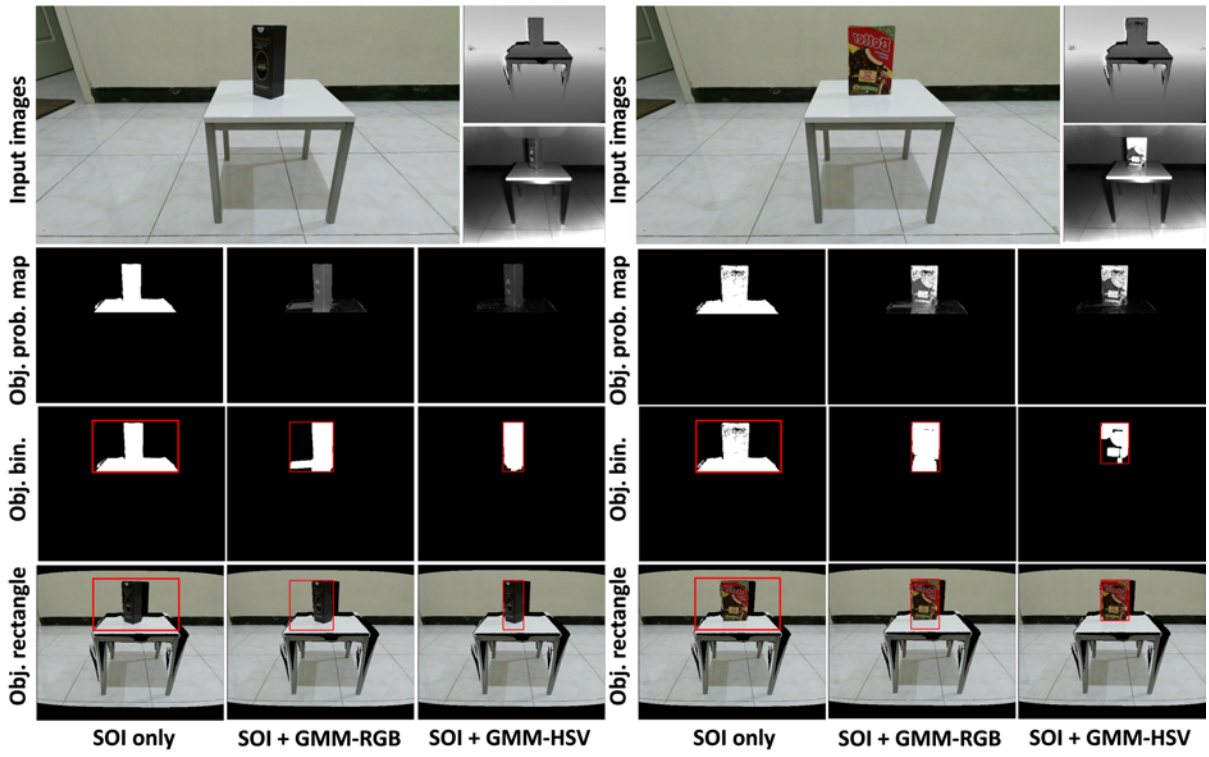

Fig. 7. Examples of object extraction results of given input images using: 1) SOI only, 2) SOI and a GMM of color information in RGB color space and near-infrared information, and 3) SOI and a GMM of color information in HSV color space and near-infrared information.

IV. Conclusion

In this paper, the object extraction using multiple cues consisted of color, depth, and near-infrared information has been proposed. Multiple cues were captured from a Kinect V2 sensor in the form of images. The captured images were processed to output the corresponded images. The proposed method was based on probabilistic models which were mod-

eled using corresponded images. Here, the scenario of Space of Interest (SOI) was introduced and the probility of SOI was modeled in a simple probabilistic model. The color information was also exploited and compared by using two color spaces, i.e., RGB color space and HSV color space. Along with near-infrared information, a GMM was used to model the information. Each probabilistic models were then use to generate probabilistic maps which were used to calculate an object probabilistic map. Using the probabilistic map of the object, we can binarize the image to estimate the object region. Several experiments have been conducted and the best result was achieved from the model with combination of depth, near-infrared, and color information in HSV color space. The results of recall, precision, and F1-score were 83.58%, 88.40%, and 84.96% respectively.

V. Future Work

In the future, first, we are planning to test with various types of data including the data that have similar shapes but different patterns. Next, we are planning to implement the system for an autonomous data acquisition system which adopts the SOI scenario. We also plan to compare the model which uses a hierarchical Bayesian framework such as a Bayesian GMM or the non-parametric models such as infinite GMM [24].

References

[1] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. IEEE Conference on Computer Vision and Pattern Recognition. Columbus. 2014;580587.

[2] Okuma K, Taleghani A, de Freitas N, Little JJ, Lowe DG. A Boosted Particle Filter: Multitarget Detection and Tracking. 2004;2839.

[3] Obata M, Nishida T, Miyagawa H, Ohkawa F. Target tracking and posture estimation of 3d objects by using particle filter and parametric eigenspace method. Proc of Comput Vis, & Pattern Recognit. 2007;6772.

[4] Muhammad Attamimi, Takaya Araki, Tomoaki Nakamura, and Takayuki Nagai. Visual Recognition System for Cleaning Tasks by Humanoid Robots. International Journal of Advanced Robotic Systems: Humanoid, 2013;114.

[5] Attamimi, M., Nagai, T., Purwanto, D. Particle filter with integrated multiple features for object detection and tracking. Telkomnika (Telecommunication Computing Electronics and Control).

[6] Muhammad Attamimi, Akira Mizutani, Tomoaki Nakamura, Takayuki Nagai, Kotaro Funakoshi, Mikio Nakano. Real-time 3D visual sensor for robust object recognition. IROS Taipei. 2010;45604565.

[7] Muhammad Attamimi, Tomoaki Nakamura, and Takayuki Nagai. Hierarchical Multilevel Object Recognition Using Markov Model. Proceedings of the 21st International Conference on Pattern Recognition. 2012.

[8] Kinect for Xbox One: https://en.wikipedia.org/wiki/Kinect.

[9] Salamati, N, Fredembach, C, Ssstrunk, S. Material Classification Using colour and NIR Images. IST 17th Colour Imaging Conf. 2009;216222.

[10] Mokhtar M. Hasan, Pramod K. Mishra. Superior Skin Color Model using Multiple of Gaussian Mixture Model. British Journal of Science. 2012; 6(1): 114.

[11] Stephen J. Krotosky and Mohan M. Trivedi. A Comparison of Color and Infrared Stereo Approaches to Pedestrian Detection. IEEE Intelligent Vehicles Symposium, 2007;8186.

[12] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst. 2015;9199.

[13] Dong W, Wang Y, Jing W, Peng T. An Improved Gaussian Mixture Model Method for Moving Object Detection. TELKOMNIKA (Telecommunication Comput Electron Control). 2016;14(3A):115.

[14] Sumardi, Taufiqurrahman M, Riyadi MA. Street mark detection using raspberry pi for self-driving system. Telkomnika (Telecommunication Comput Electron Control). 2018;16(2):62934.

[15] N. Salamati and S. Ssstrunk. Material-Based Object Segmentation Using Near-Infrared Information.

[16] Lai K, Bo L, Ren X, Fox D. A Large-Scale Hierarchical Multi-view RGB-D Object Dataset. Int. Conf. on ICRA. 2011;18171824.

[17] Kai Welke, Jan Issac, David Schiebener, Tamim Asfour, Rdiger Dillmann. Autonomous acquisition of visual multi-view object representations for object recognition on a humanoid robot. Int. Conf. on ICRA. 2010;20122019.

[18] Kinect for Xbox 360: https://en.wikipedia.org/wiki/Kinect.

[19] Alex Krizhevsky. ImageNet Classification with Deep Convolutional Neural Networks. NIPS. 2012.

[20] Mansib Rahman. Beginning Microsoft Kinect for Windows SDK 2.0: Motion and Depth Sensing for Natural User Interfaces. Apress Berkely, CA, USA 2017.

[21] Muhammad Attamimi and Takayuki Nagai. A Visual Sensor for Domestic Service Robots. Journal on Advanced Research in Electrical Engineering. 2018; 2(1):3136.

[22] Conrad Sanderson, Ryan Curtin. An Open Source C++ Implementation of Multi-Threaded Gaussian Mixture Models, k-Means and Expectation Maximisation. In Proc. of International Conference on Signal Processing and Communication Systems. 2017.

[23] Everingham M, Gool L, Williams C. K, Winn J, and Zisserman A. The Pascal Visual Object Classes (VOC) Challenge. Int. J. of Computer Vision. 2010; 88(2): 303338.

[24] Rasmussen C. E. The Infinite Gaussian Mixture Model. In Advances in Neural Information Processing Systems. 2000;554560.