TELKOMNIKA, Vol.16, No.6, December 2018, pp.3008~3015
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v16i6.9466

Particle Filter with Integrated Multiple Features for Object Detection and Tracking

Muhammad Attamimi*[1], Takayuki Nagai[2], Djoko Purwanto[3]
[1,3]Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2]Department of Mechanical Engineering and Intelligent Systems,
The University of eElectro-Communications, Tokyo, Japan
*Corresponding author, e-mail: attamimi@ee.its.ac.id

Abstract

Considering objects in the environments (or scenes), object detection is the first task needed to be accomplished to recognize those objects. There are two problems needed to be considered in object detection. First, a single feature based object detection is difficult regarding types of the objects and scenes. For example, object detection that is based on color information will fail in the dark place. The second problem is the object's pose in the scene that is arbitrary in general. This paper aims to tackle such problems for enabling the object detection and tracking of various types of objects in the various scenes. This study proposes a method for object detection and tracking by using a particle filter and multiple features consisting of color, texture, and depth information that are integrated by adaptive weights. To validate the proposed method, the experiments have been conducted. The results revealed that the proposed method outperformed the previous method, which is based only on color information.

Keywords*: object detection, object tracking, multiple features, features integration, particle filter*

1. Introduction

Recognizing objects in the scene needs several steps. The first step is detecting those objects. One of the challenges in object detection is to detect all the target objects without fail. This means that the detection system needs to adapt to different kinds of scenes especially when the illuminating conditions change drastically such as in the dark places. It also comes with the diversity of the objects in which the system needs to deal with. These issues motivate us to propose object detection that able to adapt to the illuminating changes. For solving the object detection problems, it is necessary to refer the issues. First, the detection using single feature is difficult, because there is such situation when the detection will definitely fail using that feature. For example, in the dark scene, feature such as color does not give useful information. There is also another situation, i.e. the detection that involved various types of objects; such detection will be difficult even if in the normal illuminating condition. For instance, it is straightforward using color information when dealing with colorful objects. However, such system will fail when detecting the object without texture, such as white dishes, cup, and so forth. Therefore, the use of multiple features is crucial to handle the feature weakness in particular cases to build a stronger object detection.

Next, the pose of the target objects in the scene that is arbitrary. In general, there are several terms needs to be considered when dealing with the pose, such as rotation, translation, and the depth (or the position of target object in the real world; this position is responsible for the projection size of the target object in the two-dimensional scene). If all possible poses are considered, the computational cost will high. Hence, the inference of likely pose is needed for object detection system. This study proposed an object detection based on multiple features and particle filter. Features such as color, texture, and depth information are used in this study to solve the first problem. The second problem can be solved by approximating the target object with a combination of several parts such as rectangle part of the target image. Particle filter completes the object detection in solving the second problem due to its ability to infer the existence of the object in the scene. To determine the candidate's regions of target objects, a probability map is generated.

Received April 2, 2018; Revised October 2, 2018; Accepted October 24, 2018

Ultimately, the regions with high probabilities are selected as target's regions. There are several works related to object detection and/or object tracking using particle filter [1–4]. Feature hierarchies was studied in [5], whereas Convolutional Neural Networks (CNN) was modified and used in [6, 7]. Some works on detection of moving objects were done in [8, 9]. In robotics area, the study of visual recognition for cleaning task that includes detection of objects on table top was done in [10]. Almost of these studies used single feature that is color information whereas our method based on multiple features; which consist of not only color information but also texture and depth information; that are adaptively integrated. It is also considered using depth information while the previous method does not.

The remainder of this study is organized as follows. An overview of the proposed method is presented in section 2. A discussion on the extraction of multiple features used in this study will be given in section 3. Details of proposed object detection and tracking using particle filter are provided in section 4. The experimental setting and results are discussed in section 5. Finally, section 6. concludes this study.

2. Overview of the Proposed Method

In this study, a 3D visual sensor [11] is used. This sensor consists of one time-of-flight (TOF) and two CCD cameras, which is able to acquire color information and point clouds including depth information in real time by calibrating the TOF and the two CCD cameras. Therefore, color, texture, and depth information, which are captured by the sensor, are used to realize the object detection and tracking. Multiple features can be the key solution in dealing with object detection in various environments. However, those features can make the decision in detection confusing due to false recognition of a particular situation or scene. For example, the information from color feature cannot be used when a scene is dark because it leads to false recognition. Hence, the problem of integrating multiple features is crucial when using multiple features. It should be noted that the effective features for object detection are depended on the object's properties and the environment where the object is laid. For instance, a colorful objects or objects that have rich textures can be detected using color and/or texture information. However, textureless objects such as white dishes are difficult to be detected using color and/or texture information. Moreover, the environments are also affected the object detection. For example, a yellow plushy normally easy to be detected using color information. However, if the background shares the same color with the plushy or due to bad illuminating condition, the detection will be difficult.

On the other hands, textureless objects will be easier to be detected in the colorful background, because its textureless has an important meaning in detection on such situation. It can be said that it is difficult to determine the feature should be used in detection beforehand because it is difficult to know what kind of objects are laid in unknown scenes. Therefore, the proposed method utilizes a likelihood as an input in weighting process of the features. This weighting process can be used as feature selection while performing object detection.

Figure 1 illustrates the architecture of proposed method. In the learning phase, multiple features that consist of color, texture, and depth information are extracted. The object-learning scenario used in [12] is adopt; where the user shows the object to the robot during the learning phase. The reference images are divided into several parts to solve the problem of detection in the arbitrary pose. To realize scale invariant, the features are extracted after rescaling the reference images. In the detection phase, at the first time, particle are located randomly on the scene. After resampling the particles several times, the particles will focus on a certain position; this position is a candidate of target objects. Finally, probability map is generated to determine the target object's region. For object tracking, the processes on detection except the initialization step are similar. The initialization step of object tracking is the detected position of the object.

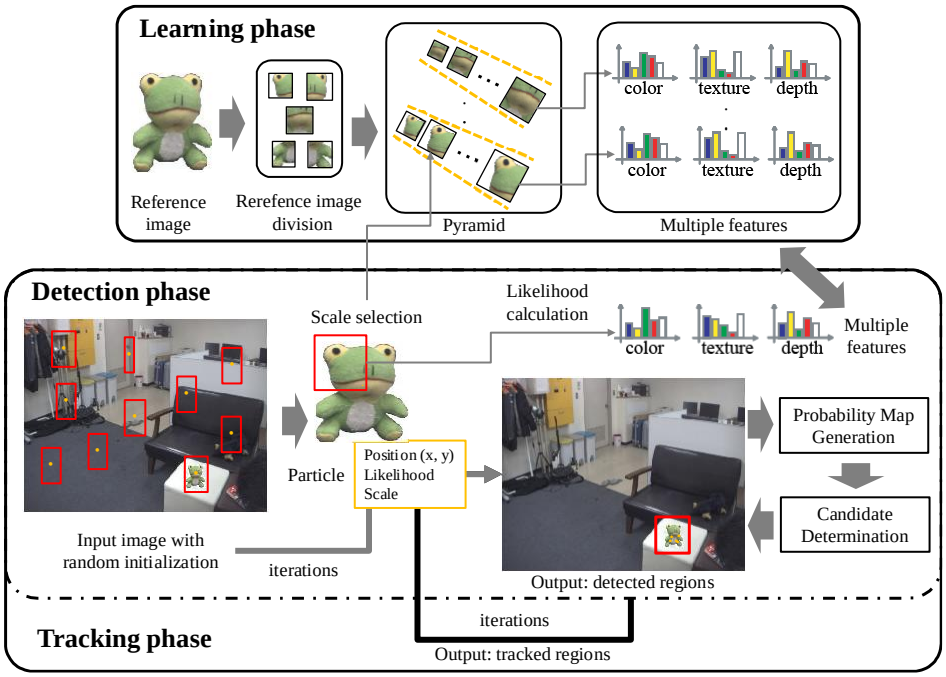


Figure 1. Object detection and tracking system.

3. Extraction of Multiple Features

In this study, the features that are invariant to scale, rotation, and translation; are needed. The histogram-based feature is employed on our proposed method because it makes features rotation- and translation-invariant. Moreover, the depth information can be used for normalizing the scale, which results in the scale-invariant features. Since it is difficult to realize the feature that has view invariance, this property is ensured by matching of all features from various viewpoints, which are accumulated in the learning phase. The features employed in this study will be discussed. Color and texture information are used to distinguish among objects of similar shape but different color or texture. Color information is calculated as a color histogram of hue (H) and saturation (S) in the HSV color space, which is chosen considering its robustness to illumination changes. The values of H and S at each pixel inside the target object are quantized into bins of size 32 and 10, respectively. For implementation, a 320-dimensional feature vector is utilized.

As for texture information, a circular census transform histogram (CCT) is proposed. This feature is an extension of census transform histogram (CENTRIST) [13]. Since CENTRIST calculate feature by comparing the pixel to its 8-neighborhood pixels, the computational cost is low. However, CENTRIST is not rotation invariant and still lack in feature representation due to its simplicity. To detect objects in various scenes, the CCT, that is the feature which can compensate the weakness in CENTRIST; is introduced. Unlike the original CENTRIST that uses 2-valued thresholding (i.e., 0 if less than its neighborhood pixel and 1 otherwise), our proposed CCT uses 3-valued thresholding (i.e., 0: if the difference value of pixel and its neighborhood is between the negative predetermined threshold and the positive one, 1: if the difference value of the pixel and its neighborhood is less then the negative predetermined threshold, and 2: if the difference value of the pixel and its neighborhood is more then the positive predetermined threshold). To achieve rotation invariant, the comparison of original CENTRIST is modified from using 8-neighborhoods into circular-P-neighborhoods. Here, the circular-P-neighborhoods can be found by calculating P posititon (x_p, y_p) of "virtual pixel" p as follows.

$$x_p = x_c + R\,cos(2\pi p/P), \quad y_p = y_c + R\,sin(2\pi p/P), \tag{1}$$

where, x_c and y_c represent the position of a target pixel, and R is a radius from the target pixel which is predetermined. The value of "virtual pixel" is calculated using interpolation. In this

study, P is set to 8. The combination of 8-possibilities comes from comparison with 3-valued thresholding are put into the same bin. Finally, texture information is represented as an 834-dimensional feature vector.

The histogram of depth (HOD) [11] is used as a depth information. This is a histogram of depth values of all pixels in an object region. To detect the object in the scene, HOD is modified. First, in the learning phase, the mean distance $\bar{z}$ of the object region is calculated. Next, $z_{min} = \bar{z} - k$ and $z_{max} = \bar{z} + k$ are determined; where k is the largest distance from $\bar{z}$. Then, voting the depth value z into bin b is done as follows.

$$b = \left\lfloor N_d \frac{z - z_{min}}{z_{max} - z_{min}} \right\rfloor, \quad (2)$$

where, N_d is number of bins.

Moreover, there is a possibility that background can have a bad effect on the calculation of HOD inside particle's region. To tackle such problem, the depth information is filtered by selecting only the depth that its value is between the minimum depth value that is inside the region and predetermined threshold.

4. Particle Filter Based Object Detection and Tracking

The important processes of this study will be discussed as follows.

4.1. Features Integration

In this paper, particle filter processing, in general, is used to perform object detection. For particle n, the particle's likelihood P_n is calculated as follows.

$$P_n \propto exp\{-w^c \lambda^c (D_n^c)^2 - w^t \lambda^t (D_n^t)^2 - w^s \lambda^s (D_n^s)^2\}, \quad (3)$$

where, D^c, D^t, D^s are represented respectively, the Bhattacharyya distance of color, texture, and depth information. The Bhattacharyya distance of feature vector $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ (both of them are M-dimensional) that is represented as $D(\boldsymbol{h}_1, \boldsymbol{h}_2)$ is calculated as follows:

$$D(\boldsymbol{h}_1, \boldsymbol{h}_2) = \sqrt{1 - \sum_{m=1}^{M} \boldsymbol{h}_1(m) \times \boldsymbol{h}_2(m)}. \quad (4)$$

moreover, λ^* is a predetermined coefficient to adjust the variance of distances and dependent on the type of a feature. The weight of each feature w^*, is updated as following.

$$w^* = min\left(1/_{min(\lambda^* (D_n^*)^2)}, 1\right). \quad (5)$$

For a given feature, if the distances of all particles are large, such feature will not afford effective detection toward the scene. This will decrease the integrated likelihood of all particles. Hence, as for the dark scene which is color information cannot be used due to extremely low intensities, the distances of all particles will be large and the weight of color information will be small that can nullify automatically the color information. On the other hands, if the color of the objects and the background is similar, which is the case when the distances are all small in every place where the objects are located; all of the particles will have the same likelihoods of color information that is close to one which means that the color information is nullified.

4.2. Determination of Window Size of the Particle based on Distance

In general, the window size of each particle is one of the parameters of a particle filter which is processed by sampling it. Thanks to the 3D visual sensor, the depth information can be utilized to calculate the window size of the objects as well as the window size of the particles. Therefore, the parameters of the particle filter are its position on the scene (x, y). Moreover, to realize the scale invariant, the pyramid method of reference images is adopt (i.e. providing several reference images that are captured in several scales). Then select the closest scale by comparing the distance between the reference image and the particle size. Finally, likelihood calculation of the feature is done by using the selected scale.

4.3. Reference Image Division

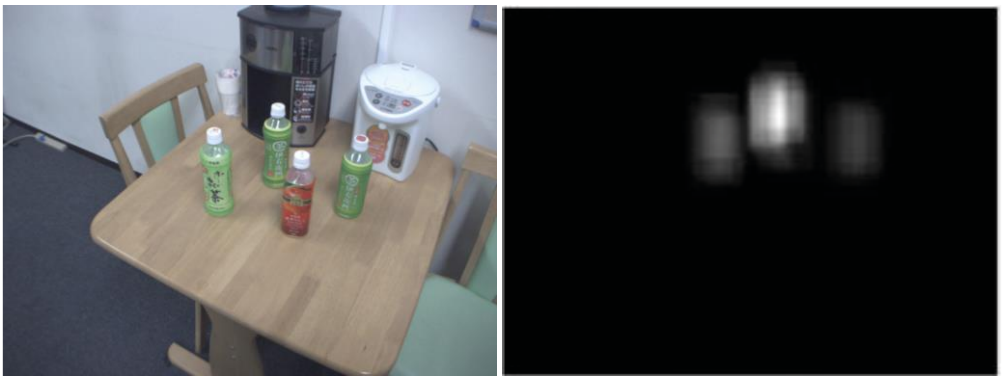
Although the features used in this study is rotation invariant, they are not three-dimensional-rotation invariant. This type of rotation is necessary for particle filter to detect the objects which are in general laid in arbitrary positions. For example, if the plastic bottle lying horizontal (or not stand) on the table, it will be difficult to be detected even using a rotation invariant features. To solve such problem, the reference images are divided into several rectangles as shown in Figure 1. By dividing the reference image into several parts, the problem of three-dimensional-rotation in object detection can be solved; as well as the occlusion problem because the partial part can be detected using the proposed method.

4.4. Candidate's Regions Determination

In this study, the region of candidates in object detection is represented by a probability map. For N particles each of which has likelihood P_n, the probability map of position (x, y) on the scene is calculated as follows.

$$P(x,y) = \frac{1}{P_{max}} \sum_{n=1}^{N} \frac{P_n(x,y) - P_{min}}{P_{max} - P_{min}}, \quad (6)$$

where, P_{min} and P_{max} represent respectively minimum and maximum likelihood of object detection for a given scene. For a given input image in Figure 2 (a), the probability map is generated as shown in Figure 2 (b). The position of a particle (x, y) that has probability $P(x, y)$, which is higher than a predetermined threshold, is chosen to be a candidate's region.



(a) (b)

Figure 2. Target object determination: (a) input image, and (b) corresponding probability map.

4.5. Object Tracking

Object tracking is realized by putting the initial position of detected objects. This can be done sequentially after object detection because the same framework is used.

5. Experiments

Experiments were conducted to validate the proposed object detection and tracking.

5.1. Experimental Setting

Experiments were carried out in the living room, using 18 objects as shown in Figure 3 (b). A user showed each object in various angles to the robot, and a database which contains feature vectors of 10 frames per object was generated by the robot in the learning phase (see Figure 3 (a) for object learning scenario). In the detection phase, several scenes were used to test the proposed method.

(a) (b)

Figure 3. (a) Object learning scenario: a user is showing the target object to the robot [12], whereas the red rectangle depicts the 3D visual sensor proposed in [11], and (b) examples of objects used in the experiment.

5.2. Evaluation of Object Detection

In this study, to validate the proposed object detection, 20 scenes were captured and 100 synthesized scenes were generated (hereafter, these scenes were defined as normal-scenes). These scenes were generated by locating targets object randomly in captured scenes. Moreover, 100 scenes were also provided with a different color (hereafter, these scenes were defined as different-color-scenes) and 100 dark scenes (hereafter, these scenes were defined as dark-scenes), which resulted in 300 scenes in total were used in this experiments. Figure 4 shows the example of synthesized scenes.

(a) (b) (c)

Figure 4. Examples of synthesized scenes: (a) normal scene, (b) different-color-scene, (c) dark-scene. Each of which is corresponded.

The synthesized scenes were then inputted to the detection system, and recall, precision, and F-measure of detected objects were calculated. Here, if “true detected” is defined as the regions detected by the system that belong to the true regions, then recall is the fraction of “true detected” over true regions, whereas the precision is the fraction of “true detected” over detected regions. Meanwhile, F-measure is is the harmonic mean of precision and recall which can describe both the values at once. The higher all of those values are the better the detection system performs. Our proposed method was compared to the previous method. Here, the previous method is a method that used the color information only. Table 1 shows the results of object detection in various types of scenes. In normal-scenes (see life side of Table 1), there is a slight improvement in recall and a large improvement in precision and F-measure. However, in different-color-scenes (see middle side of Table 1) and dark-scenes (see right side of Table 1), the proposed method outperformed the previous method which cannot detect objects almost at all. Thanks to the depth information, a reliable information in dark- and different-color-scenes still can be used. The results also show that proposed method can adapt by selecting features that are appropriate when the environments changed. Several examples of object detection is shown in Figure 5 (a). It can be seen that the detection of multiple objects can be done by the proposed method as shown in Figure 5 (b) (top-side). Figure 5 (c) (bottom-side) shows the detection results of object with different pose with learning phase.

Table 1. The Detection Rate in Various Types of Scenes.

Detection	In normal scenes			In scenes with different color			In dark scenes		
Method	Recall	Precision	F-measure	Recall	Precision	F-measure	Recall	Precision	F-measure
Previous	0.825	0.623	0.651	0.059	0.059	0.059	0	0	0
Proposed	0.863	0.825	0.837	0.775	0.488	0.552	0.605	0.483	0.516

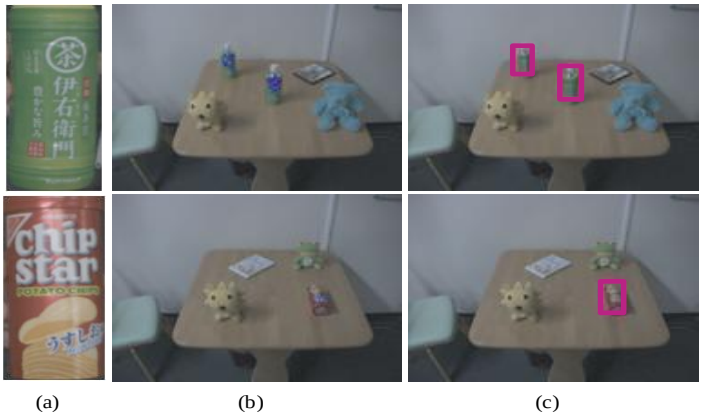


Figure 5. Examples of object detection results: (a) target object, (b) input scene with the final state of particles, and (c) detected objects.

5.3. Discussion

As mentioned above, to perform object detection with various types of objects in various environments, a feature selection system is needed to choose the suitable features for object detection. To demonstrate this, our proposed object detection and tracking was tested in the scenes with illumination that gradually changes from normal to dark. Figure 6 shows the object tracking results in such condition. One can see that object tracking can be done by the proposed system (see Figure 6 (top)). The weights of features were changed during the tracking phase as shown in Figure 6 (bottom). In the scenes with a normal illuminating condition, the weight of both of color information and depth information was high, whereas the weight of texture information was changed from high in the first frame to low. In the tracking phase, the pose of objects was difficult to detect using texture information due to lack reference images collected in the learning phase. However, in the dark scenes, the weight of color and texture information was low, whereas the depth information was high; because the reliable information was depth information. Overall, object tracking can be performed by the proposed method in the various environments.

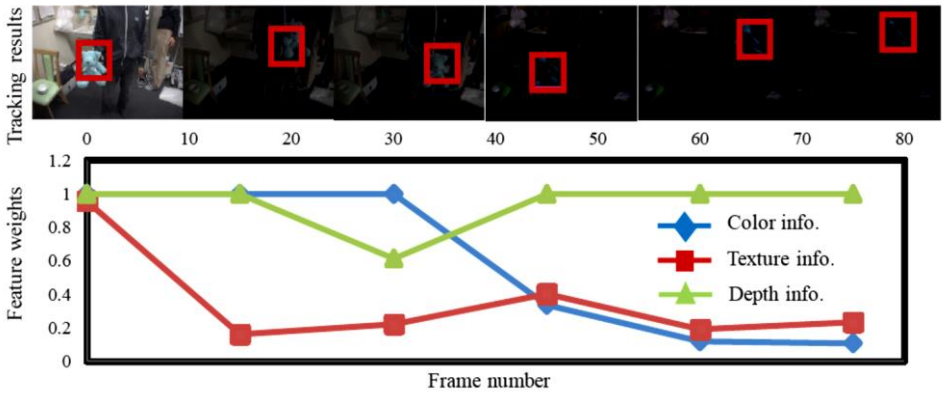


Figure 6. Examples of object tracking results. Top-side of the figure depicts the tracked target objects in various illuminating condition, whereas the bottom-side illustrates the corresponding weights.

6. Conclusions

In this paper, a method to detect and track the target objects based on a particle filter with integrated multiple features, has been proposed. The proposed method used color, texture, and depth information that is combined by an adaptive weighting mechanism which considered both the object's types and the complexities of scenes. Thanks to the proposed method, the detection of various types of objects in various scenes can be done. The results showed that in a normal scene our proposed method outperformed the previous method which is based on color information only. In the extreme scenes, that are the scenes with different color and dark scene, the previous method was not able to detect the objects, whereas our proposed method was able to perform object detection. The experimental results have also shown that the proposed method can detect the objects located in arbitrary poses. Moreover, our proposed method can also perform object tracking in such scenes.

The current object detection system can be considered as detection of an instance. Developing not only instance detection but also object detection of a category is our future work. Moreover, inspired by work on [14], integration and improvement of detection systems through hierarchical object detection will be studied in the future.

References

[1] Okuma K, Taleghani A, de Freitas N, Little JJ, Lowe DG. A Boosted Particle Filter: *Multitarget Detection and Tracking.* 2004; 28–39.
[2] Czyz J. *Object Detection in Video via Particle Filters*. 18th Int Conf Pattern Recognit. 2006: 820–3.
[3] Czyz J, Ristic B, Macq BM. A particle filter for joint detection and tracking of color objects. *Image Vis Comput.* 2007; 25:1271–1281.
[4] 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; 67–72.
[5] Girshick R, Donahue J, Darrell T, Malik J. *Rich feature hierarchies for accurate object detection and semantic segmentation*. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2014; 580–587.
[6] Girshick R. *Fast R-CNN*. Proc IEEE Int Conf Comput Vis. 2015 International Conference on Computer Vision, ICCV 2015: 1440–8.
[7] 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; 91–99.
[8] Dong W, Wang Y, Jing W, Peng T. An Improved Gaussian Mixture Model Method for Moving Object Detection. *TELKOMNIKA Telecommunication Computing Electronics Control.* 2016; 14(3A): 115.
[9] Sumardi, Taufiqurrahman M, Riyadi MA. Street mark detection using raspberry pi for self-driving system.*TELKOMNIKA Telecommunication Computing Electronics Control*. 2018; 16(2):629–634.
[10] Attamimi M, Araki T, Nakamura T, Nagai T. Visual recognition system for cleaning tasks by humanoid robots. Int J Adv *Robot Syst*. 2013; 10.
[11] Attamimi M, Mizutani A, Nakamura T, Nagai T, Funakoshi K, Nakano M. *Real-time 3D visual sensor for robust object recognition*. In: IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings. 2010.
[12] Attamimi M, Mizutani A, Nakamura T, Sugiura K, Nagai T, Iwahashi N, et al. *Learning novel objects using out-of-vocabulary word segmentation and object extraction for home assistant robots*. In: Proceedings - IEEE International Conference on Robotics and Automation. 2010.
[13] Wu J, Christensen HI, Rehg JM. *Visual Place Categorization: Problem, dataset, and algorithm*. In: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. 2009. p. 4763–4770.
[14] Attamimi M, Nakamura T, Nagai T. *Hierarchical multilevel object recognition using Markov model*. In: Proceedings - International Conference on Pattern Recognition. 2012.