Bilge International Journal of Science and Technology Research
Web : http://dergipark.gov.tr/bilgesci - **E-mail:** kutbilgescience@gmail.com

Received: 03.12.2018
Accepted: 08.03.2019
DOI: 10.30516/bilgesci.491557

ISSN: 2651-401X
e-ISSN: 2651-4028
3 (1), 58-62, 2019

3D Object Detection Using a New Descriptor with RGB-D

Erkut Arıcan[1*], Tarkan Aydın[1]

Abstract: Object detection is a very important study area in computer vision. Many research use only RGB images to find objects. In our work, we present new descriptor for object detection using RGB-D's Depth image data. We combine RGB image with depth image to create new feature vector. The introduced features feeds Bag of Visual Words algorithm to classify images of the objects. Result shows us to RGB-D images are given better accuracy results to comparing with RGB image.

Keywords: RGB-D, Depth Image, Machine Learning, Bag of Visual Words

1. Introduction

There are many fields in computer vision and object detection is the most popular one. A long time ago, finding object was a challenging process. In last decades, many works show that finding object became an easy process for computer vision but mostly RGB information without depth is used. In literature, it can be found many studies using descriptor which are very preferred in computer vision (Lowe 2004; Bay et al. 2006; Shechtman and Irani 2007; Calonder et al. 2010; Rublee et al. 2011)

Nowadays, technology is going better and cheaper therefore it can be found and access 3D image easily. You can create dataset using Microsoft Kinect (Microsoft) , Intel RealSense 3D Camera (Intel) , etc. or you can download many 3D image dataset like SUN3D (Xiao et al. 2013), Berkeley 3D Object Dataset (Janoch et al. 2011), RGB-D Object Dataset (Lai et al. 2011) and NYU Dataset (Silberman et al. 2012) . In literature, as observed some 3D descriptors such as Huang's study (Huang and You 2012) and (Arıcan and Aydın 2017). In (Huang and You 2012), authors propose 3D local descriptor using self-similarity and point cloud information while in (Arıcan and Aydın 2017), authors used depth oriented gradients for object detection.

In this paper, we propose a new descriptor for 3D object detection using RGB and Depth image together. We combine 3 important information and create a new descriptor. Bag of Visual Word (Csurka et al. 2004) algorithm is our base algorithm and we add Local Self Similarity (Shechtman and Irani 2007) and Depth information to create a new 3D descriptor for object detection.

Our paper continue as follows: background information about Bag of Visual Words and Local Self-Similarity is given in Background Section. Then proposed method is explained in Section Material and Method. Results are showed in Result Section. Finally, in the last section, we summarized our study and mention future works.

2. Background

In this section, we explain Bag of Visual Words (BoW) and Local Self Similarity (LSS) methods which compose base method with Depth information to our study.

2.1. Bag of Visual Words

Csurka and et. all (Csurka et al. 2004) developed a descriptor for identifying the object. Authors define 4 main steps for Bag of Visual Words.

[1] Bahcesehir University, Faculty of Engineering and Natural Sciences, 34353, İstanbul, Türkiye
*Corresponding author (İletişim yazarı): erkut.arican@eng.bau.edu.tr

Citation (Atıf): Arıcan, E., Aydın, T. (2019). 3D Object Detection Using a New Descriptor with RGB-D. Bilge International Journal of Science and Technology Research, 3 (1): 58-62.

These steps are; 1) determining of image patches; 2) creating a vocabulary; 3) creating bag of keypoints and 4) using multi-class classifier for determining categories for input images.

2.2. Local Self Similarity

In Local Self Similarity (Shechtman and Irani 2007), they use self-similarities for same object but different image characteristics. They create image patch and region part to create correlation surface. When they create the correlation surface, they use binned log-polar representation and create a descriptor. You can see LSS method in Figure 1.

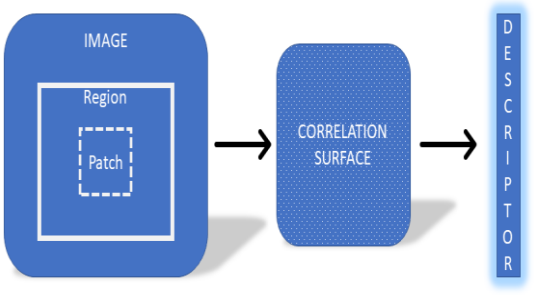

Figure 1. Local Self Similarity

Authors used Sum of Squared Distance method for creating correlation surface, this part is another key point to our work.

3. Material and Method

Feature extraction is the crucial step in feature-based machine learning methods. In our study, Ideas of Local Self Similarity was used for Region and Patch matrix and also calculation of SSD. In Algorithm 1, we explain our method step by step. In the first step, we read RGB image in grayscale format and found SURF points. We selected each SURF points as a center point and extracted 4x4 patch matrix and 12x12 region matrix. We use template matching algorithm for these 2 matrices with Sum of Square Distance (Eq.1) and create SSD matrix in the size of 9x9. This size will also define Depth image matrix's size. In Figure 2, you can see an example of template matching.

$$d_i(I_j, T) = \sum_{i=1}^{n} |I_{i,j} - T_i| \qquad (1)$$

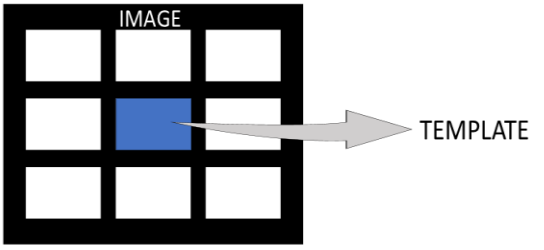

Figure 2. Template Matching Example

Afterwards, we extract SURF points from Depth image, then we select same SURF points as a center points and create 9x9 matrix. From now on depth matrix is called DM. An important part of our study is combining this depth information to local self-similarity method. For the combining process, firstly we normalize DM and convert SSD matrix as a double and then create a new matrix using Equation 2 as follows

$$DS = (SSD * e^{DM})^T \qquad (2)$$

We convert this matrix in a single row to create our new descriptor.

Algorithm 1. Our Method

1. Read RGB image in grayscale
2. Detect SURF points
3. For each SURF points selected as a center
    a. Extract 4x4 patch matrix from image
    b. Extract 12x12 region matrix from image
    c. Calculate SSD
    d. Read Depth image in grayscale
    e. Extract 9x9 matrix from depth image (DM)
    f. Create new matrix using SSD and DM
    g. Convert matrix to 1-D vector
    h. Return new descriptor

For classification of the features Bag of Visual Words was used (Bay et al. 2006) and the vocabulary size is selected 20.

4. Results

We use RGB-D Object Dataset (Lai et al. 2011) to test our algorithm. Example of RGB and Depth image from dataset can be seen in Figure 3 and 4. In this Kinect style dataset, there are 300 common household objects, 51 categories. We create implementation to our algorithm using MATLAB (MATLAB).

Figure 3. RGB Image Example from Dataset

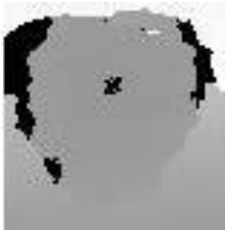

Figure 4. Depth Image Example from Dataset

In Table 1 and Figure 5, you can see comparison between our method and BoW + SSD results for different number of labels. Each label set selected from 51 categories. Biggest set include 10 labels which is explain more detail in Table 2.

Table 1. Comparison Methods Different Number of Labels

	Accuracy			
Method	4 Labels	5 Labels	6 Labels	10 Labels
BoW + SSD	61%	48%	46%	43%
Our Method	**64%**	**51%**	**57%**	**53%**

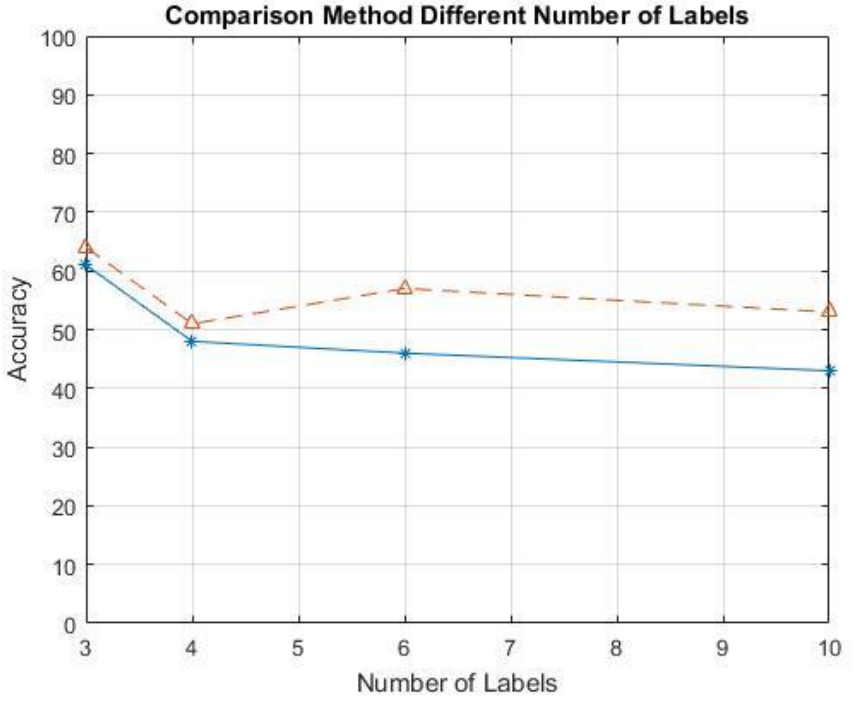

Figure 5. Comparison Methods Different Number of Labels

In Table 2, you can see in detail each label accuracy rate comparison of 10 labels in Table 1 using BoW+SSD and Our Method. These categories which are apple, ball, bell pepper, binder, bowl, cap, cell phone and soda can, give much better accuracy rate than BoW + SSD method.

Table 2. Comparison Methods Different Object in 10 Labels

	BoW + SSD	Our Method
Apple	0.0934	**0.1593**
Ball	0.4553	**0.5813**
Bell Pepper	0.0842	**0.2105**
Binder	0.4695	**0.6901**
Bowl	0.4540	**0.8333**
Calculator	0.6552	0.6092
Cap	0.8457	**0.9096**
Cell Phone	0.2945	**0.3620**
Cereal	0.7018	0.6608
Soda Can	0.2340	**0.2819**

It is clear that our algorithm gives better performance than Bag of Visual Words + SSD algorithm using Table 1 and Table 2.

5. Discussion and Conclusions

Using RGB is very popular and often preferred in object detection. On the other hand, depth images are becoming easily accessible information, so we contribute literature as proposing a new descriptor for object detection using RGB-D images. Combining Sum of Square Distance method with depth image gives better accuracy than using only RGB. For future work, we plan to improve our method to give much better performance and better running time.

Acknowledgements

This study is a part of Bahcesehir University Doctoral Programme's PhD Dissertation.

This study extended version of full text paper in International Conference on Science and Technology (ICONST 2018) hold from September 5 to 9, 2018, in Prizren, Kosovo.

References

Arıcan E, Aydın T (2017). Object Detection With RGB-D Data Using Depth Oriented Gradients. In: Book of Proceedings - International Conference on Engineering and Natural Sciences

Bay H, Tuytelaars T, Van Gool L (2006). SURF: Speeded up robust features. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp 404–417

Calonder M, Lepetit V, Strecha C, Fua P (2010). BRIEF: Binary robust independent elementary features. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 6314 LNCS:778–792. doi: 10.1007/978-3-642-15561-1_56

Csurka G, Dance C, Fan L, et al (2004). Visual categorization with bag of keypoints. Int Work Stat Learn Comput Vis. doi: 10.1234/12345678

Huang J, You S (2012). Point cloud matching based on 3D self-similarity. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on. pp 41–48

Intel Intel RealSense. https://www.intel.com

Janoch A, Karayev S, Jia Y, et al (2011). A category-level 3-D object dataset: Putting the Kinect to work. Proc IEEE Int Conf Comput Vis 1168–1174. doi: 10.1109/ICCVW.2011.6130382.

Lai K, Bo L, Ren X, Fox D (2011). A large-scale hierarchical multi-view RGB-D object dataset. In: Proceedings - IEEE International Conference on Robotics and Automation. pp 1817–1824

Lowe DG (2004). Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60:91–110. doi: 10.1023/B:VISI.0000029664.99615.94

MATLAB MATLAB. https://www.mathworks.com/products/matlab.html

Microsoft Kinect. https://dev.windows.com/en-us/kinect

Rublee E, Rabaud V, Konolige K, Bradski G (2011). ORB: An efficient alternative to SIFT or SURF. Proc IEEE Int Conf Comput Vis 2564–2571. doi: 10.1109/ICCV.2011.6126544

Shechtman E, Irani M (2007). Matching local self-similarities across images and videos. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, pp 1–8

Silberman N, Hoiem D, Kohli P, Fergus R (2012). Indoor segmentation and support inference from RGBD images. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 7576 LNCS:746–760. doi: 10.1007/978-3-642-33715-4_54

Xiao J, Owens A, Torralba A (2013). SUN3D: A database of big spaces reconstructed using SfM and object labels. Proc IEEE Int Conf Comput Vis 1625–1632. doi: 10.1109/ICCV.2013.458