International Journal of Current Engineering and Technology
©2017 **INPRESSCO**®, All Rights Reserved

E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
Available at http://inpressco.com/category/ijcet

Research Article

Real-Time Objects Recognition Approach for Assisting Blind People

Jamal S. Zraqou[#^], Wissam M. Alkhadour[#^] and Mohammad Z. Siam[!^*]

[#]Multimedia Systems Department, [!]Electrical Engineering Department, [^]Isra University, Amman-Jordan

Accepted 30 Jan 2017, Available online 31 Jan 2017, **Vol.7, No.1 (Feb 2017)**

Abstract

Blind assistance is promoting a widely challenge in computer vision such as navigation and path finding. In this paper, two cameras placed on blind person's glasses, GPS free service, and ultra-sonic sensor are employed to provide the necessary information about the surrounding environment. A dataset of objects gathered from daily scenes is created to apply the required recognition. Objects detection is used to find objects in the real world from an image of the world such as faces, bicycles, chairs, doors, or tables that are common in the scenes of a blind. The two cameras are necessary to generate the depth by creating the disparity map of the scene, GPS service is used to create groups of objects based on their locations, and the sensor is used to detect any obstacle at a medium to long distance. The descriptor of the Speeded-Up Robust Features method is optimized to perform the recognition. The proposed method for the blind aims at expanding possibilities to people with vision loss to achieve their full potential. The experimental results reveal the performance of the proposed work in about real time system.

Keywords: *Objects detection, objects recognition, features detection and matching, face detection, detectors, filters, machine learning, image representation, and focusing*

1. Introduction

[1]Globally, it is mentioned that about 38 million people are blind and a further 110 million persons are at a serious risk of becoming blind for different causes such as trachoma, glaucoma, cataract, onchocerciasis, and xerophthalmia. Visual computing can be used to help blind persons tackle some problems of perceiving some objects in cluttered scenes. Moreover, interacting with surrounding environments to find dropped or misplaced personal items for them becomes the most critical challenge.

Computers are being developed in a short time with increased speed, hardware, software, lower cost, and the increased availability of access technology. Hence, several works on assistive technologies are created to enable localization, navigation and object recognition. The best interface then can be customized based on a user request whether that is vibrations, sounds or the spoken word. Figure 1 shows an example of face(s) detection implemented by (Bellotto *et al*, 2010) which is used in this work to help the blinds for the existence of human being. Figure 2 shows a road symbol sign detection that can be used to recognize the type of information displayed on a sign to the blind.

Object recognition is closely tied to the segmentation process, which means that without segmentation; object recognition is not possible. Several steps must be executed in order to perform the recognition: Model database, feature detector, hypothesizer, and hypothesis verifier, as shown in Figure 3.

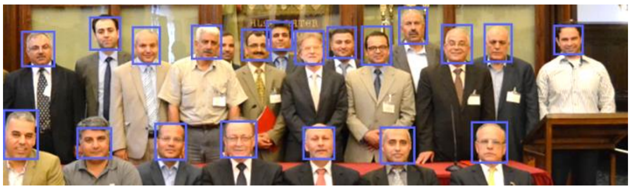


Figure 1: Faces detection in a cluttered scene

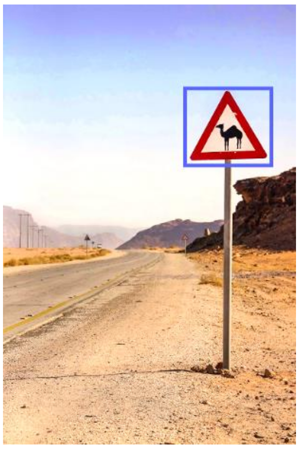


Figure 2: Roadway symbol sign detection

*Corresponding author: **Mohammad Z. Siam**

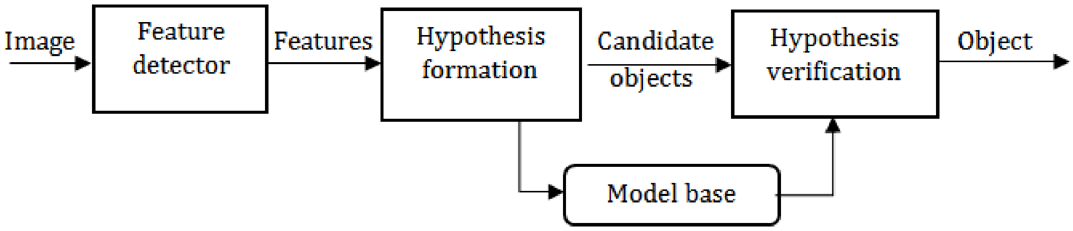


Figure 3: Components of objects recognition

The model database comprises all the models acknowledged to the system. The type of information in the model database is based on the method employed for the recognition. The models of objects are often represented by an abstract feature vector. A feature is an important attribute for describing and recognizing the object such as size, color, and shape.

Applying feature detector on an image identifies the required features to generate object hypotheses. The hypothesizer uses the detected features in the image to assign likelihood to every object that appears in the scene. Object models are used by the verifier to verify the hypothesis in order to define objects likelihood. Finally, the correct object is chosen based on the highest likelihood.

Although there are several related works on objects detection from sequence of images (videos), they generally focus on detecting one specific class of objects (Kang *et al*, 2016), such as pedestrians (Tian *et al*, 2015), cars (Li *et al*, 2014), or humans with actions Jain *et al*, 2014). This provides the main problem of applying the general object detection. In addition, the large appearance changes of objects in videos could reduce the performance of their methods.

Such requirements motivate us to apply object detection and recognition on live camera to assist blinds. Well-known feature descriptors and matching methods are employed to achieve the aim of this work. Speeded-Up Robust Feature (SURF) has shown impressive performance on object detection, which outperforms previous methods in terms of speed with accurate matches.

In this work, object recognition in cluttered scenes is proposed to notify a blind person about his/her frontal scene. The contribution of this paper can be clearly seen in the following: 1) A customized method for building features database for common objects that exist in a blind's environment is created. 2) A utilized matching process is performed using sign of Laplacian for the underlying interest point and the GPS service that marks objects based on their locations. 3) A special method for generating 3D information using two images of the same scene and detecting obstacles using ultra-sonic sensor is presented.

The remaining sections of this paper are organized as follows: Section 2 provides an extended literature review of recent research about object detection and recognition. Section 3 introduces the proposed approach and the contributed work. Section 4 illustrates the experiments that were conducted on the evaluated dataset. The results of our experiments are provided in Section 5. Concluding remarks and recommendations for future work are presented in Section 6.

2. Related Work

In (Kang *et al*, 2016), a complete multi-stage pipeline framework based on deep CNN detection and tracking for object detection in videos was proposed. The still-image object detection was combined with generic object tracking for tubelet proposal. Several perturbation and scoring schemes were evaluated and analyzed based on the tubelet proposal. A convolutional network was presented to incorporate consistency, and the consistent performance improvement was achieved over still-image detections. This work suppers from a kind of failure by keep tracking a failure detected object. The failure detection process could be caused by a large scale changes of target objects. Hence, their work is not accepted by the blind by giving a continuous false information.

The authors in (Seema *et al*, 2016) suggested using a smart system that guides a blind person about obstacles that could not be detected by his/her cane. However, the proposed system was designed to protect the blind from the area near to his/her head. The buzzer and vibrator were used and employed as output modes to a user. This is useful for obstacles detection only at a head level without recognizing the type of obstacles.

A modification of several systems used in visual recognition was proposed in (Dollár *et al*, 2014), where the authors used fast-feature pyramids and provided findings on general and specific (pedestrian) object detection systems. The results showed that the proposed scheme can be strictly used for wide-spectrum images. However, it does not succeed for narrow-spectrum images. Hence, their work cannot be used as efficient general objects detection.

In (Felzenszwalb *et al*, 2010) a combination of multiscale deformable-part models was adopted with an object-detection system to be used for variable object classes. However, the proposed system does not permit part-level mixture models or parts reusing. The presented results showed a slight performance on the car category but not on the person category of the longer-established 2006 dataset.

The authors in (Gould *et al*, 2009) wanted not to consider the multiclass image segmentation and object detection as separated tasks. Instead, they proposed a model based on joining the two tasks by building a hierarchical region-based approach to joint object detection and image segmentation. However, this model used a greedy-inference approach that may be enhanced to allow for more global steps.

An object-detection scheme based on applying convolutional neural networks with high capacity called R-CNN was proposed in (Girshick *et al*, 2014). A training model in the case of large networks was used when having rare labeled training data. This is only an area based object detection method that an illumination variation in the image could generate false object detection.

In (Nazli Mohajeri *et al*, 2011) the authors suggested a two-camera system to capture photos from several points of view to help blind people detect obstacles while moving via comparing the two photos. However, the proposed system was only tested under three conditions and for three objects. Specific obstacles that have distances from cameras of about 70 cm were detected. The results showed some range of error. Blind helping systems need to cover more cases with efficient and satisfied results.

The authors in (Dionisi *et al*, 2012) designed a device that is based on the radio–frequency identification (RFID) technology to help blinds look for different objects. A case study of looking for medicines in a home's cabinet has been considered. The designed device can provide blinds with information about how far the medicine is and makes the search easier via providing an acoustic signal. This work is biased to specific objects in predetermined scenes.

A scheme coined as You Only Look Once (YOLO) has been proposed in (Redmon *et al*, 2016), which is an object-detection scheme based on a regression problem instead of using classifiers. YOLO uses only one neural network to foresee both class probabilities and bounding boxes in single evaluation straight from full images.

In (Wang *et al*, 2007) the authors provided an object-detection scheme that relates bottom-up image segmentation and top-down recognition. However, the suggested scheme does not work when severe obstructions are there where most of the local information is degraded.

Based on the literature, no general objects detection methods were founded. Hence, customized objects that are common in a blind's environment are detected and used to create model database. The proposed method generates unique features for each object which are invariant to scales, illumination variation, and rotation, as shown in the next section.

3. The Proposed Approach

In this work, object recognition approach is presented by applying several steps: Object detection, creating unique descriptor for each object, retrieving from model database, and matching. A model database is a prerequisite stage required to be built in order to apply the matching process. It contains features for all common objects in the environments of the blind. The steps of the proposed approach are illustrated in the next subsections.

3.1Objects Detection

This step is important for finding instances of real world objects for a certain class, such as humans, faces, buildings, bicycles, or cars. First, the work presented in the library of OpenCV is used to detect faces in order to notify the blind about the number of person(s) in front of him/her as shown in Figure 1. Second, Feature-based detection for other objects is employed to achieve the recognition as shown in Figure 4. Finally, disparity map and ultra-sonic sensor are used to provide 3D information and detecting any obstacles. Based on our experiments, SURF method presented in (Bay *et al*, 2006) outperforms the former methods in terms of speed, distinctiveness, and robustness. The features descriptor is invariant to illumination changes, multi-scales, and rotation.

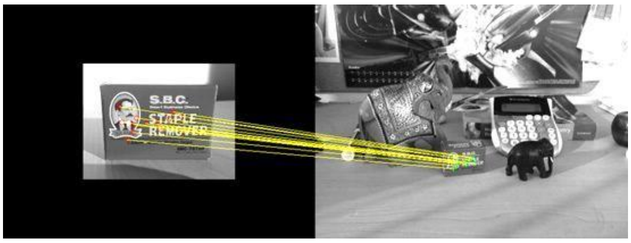


Figure 4: Object detection using feature extraction

3.2 Building Models Database

All objects that exist in a blind's environment are manually extracted and identified by the user to apply machine learning. The model database saves the features for each object as shown in Equation 1, which is used later to apply the matching process. The extracted features are then saved in the database as shown in Table 1. In order to reduce mismatches and computational time, GPS service is used to determine the location for each object. Hence the comparisons are only applied to the objects that exist in an involved area.

$$v_{\rightarrow \text{Object1}} = \sum dx\,, \sum dy\,, \sum |dx|\,, \sum |dy|$$

Equation 1: A sub-region vector is represented by 4x4 matrix

Table 1: Database structure for instances in the real world

Object ID	Features vector	GPS coordinates
Chair	$v_{\rightarrow \text{Object1}}$	(Latitude1, Longitude1)
Door	$v_{\rightarrow \text{Object2}}$	(Latitude2, Longitude2)
…	…	…
Object(N)	$v_{\rightarrow \text{ObjectN}}$	(LatitudeN, LongitudeN)

The database of models is created by applying the following:

1) For each object in the input image: Do the next steps.
2) Extract the SURF features descriptor.
3) Get the GPS coordinate.
4) Identify the object by the user.
5) Save the extracted info.

3.3 Objects Recognition

Based on the models database, fast indexing is performed using the sign of the Laplacian for the underlying interest point and the GPS service to specify the involved areas. Typically, as performed in (Bay *et al*, 2006), the interest points are found at blob-type structures and the sign of Laplacian differentiates bright blobs on dark backgrounds from the reverse situation. This feature and the GPS area-based service are utilized to apply the proposed work at no extra computational cost. It should be noted that we only compare features if they have the same type of contrast and in same location. Therefore, this information allows faster matching and provides a slight increase in the performance as shown in Figure 5.

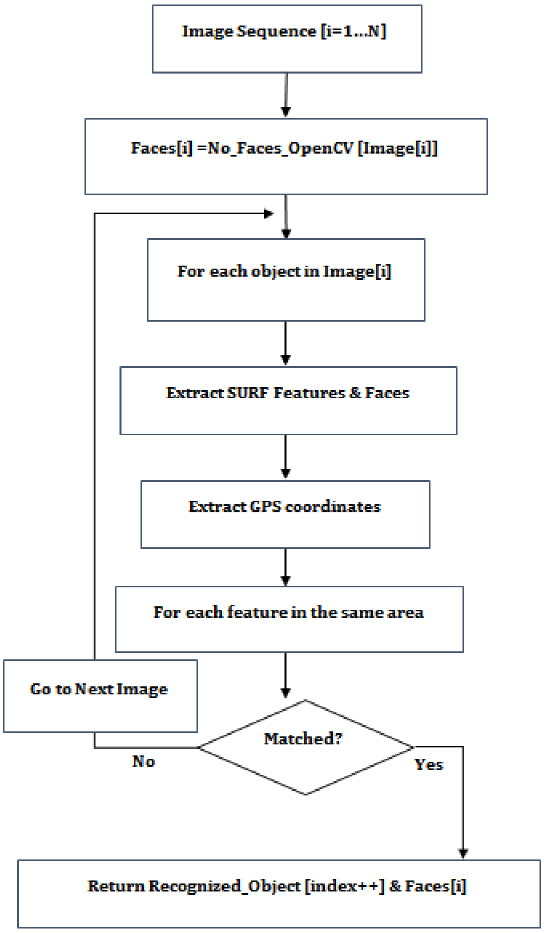


Image Sequence [i=1…N]
Faces[i] =No_Faces_OpenCV [Image[i]]
For each object in Image[i]
Extract SURF Features & Faces
Extract GPS coordinates
For each feature in the same area
Go to Next Image
Matched?
No
Yes
Return Recognized_Object [index++] & Faces[i]

Figure 5: Flow chart of the proposed method

3.4 Depth and Obstacles Detections

By having two images of same scene, it is possible to get depth information from that in an intuitive way as shown in Figure 6. In order to understand this technology it would be as blinking eyes, one at a time, alternating between the left and right eye. This reveals that the closer objects would appear to jump about their location more than the further objects. This small shift denotes that the objects move away as shown in Figure 7. Hence, the brighter and darker shades in the disparity map represent the lesser and greater distances from the camera, respectively. In this work, two cameras are employed to provide two images of the same scene. The blind then can be notified about the closer and further objects. Ultra sonic sensor is used to retrieve the distance between the camera and the closest obstacle for medium and long distances.

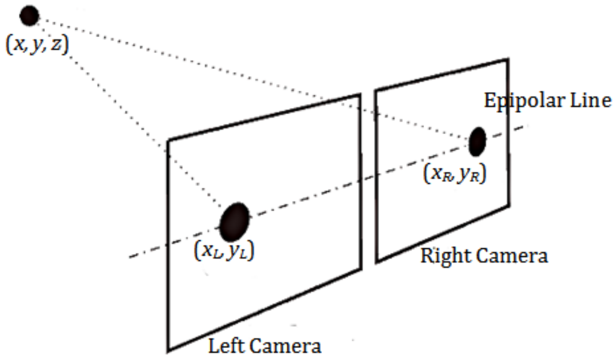
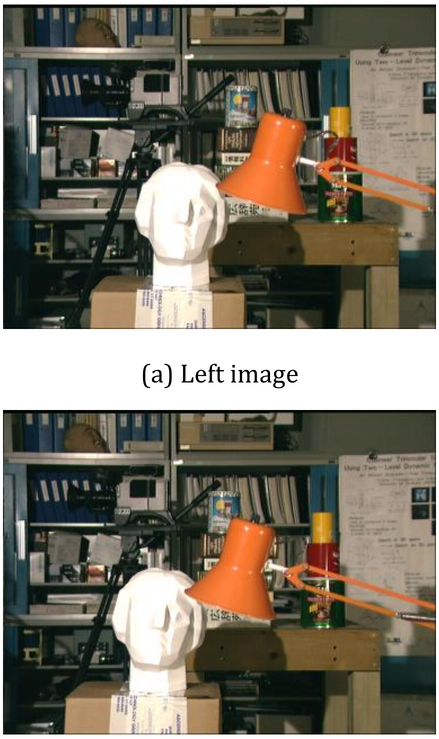
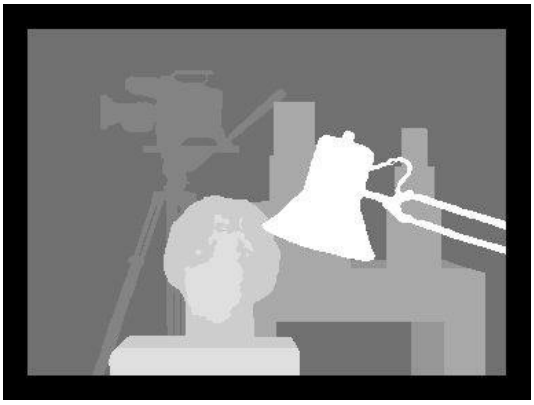


Figure 6: 3D point extraction using two images taken from left and right cameras of the same scene



(a) Left image

(b) Right image



(c) Disparity map

Figure 7: Disparity image in (c) is created by matching each pixel in the (a) left image with its corresponding in the (b) right image

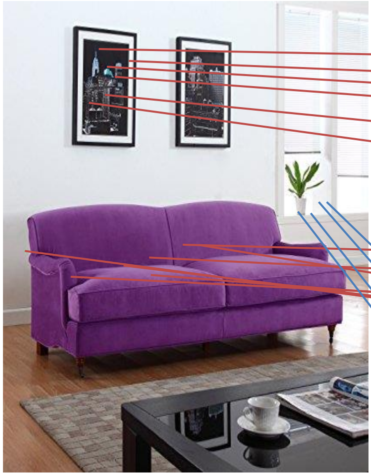
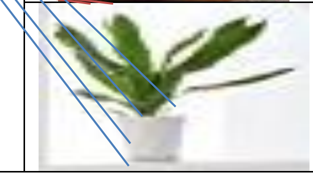
4. Experiments

The results are presented based on our own dataset. The dataset includes objects captured from images of the real life. Initially, the objects are gathered and identified by ourselves. The dataset consists of 300 images of 25 objects.

The tested images comprise 180 images that were taken from the right camera. The resolution of images is about 600 x 500 pixels. Objects recognition from the model database proceeds as follows:

The GPS coordinate is extracted, and it retrieves all objects that are associated with it. The images in the test set are compared to all objects in the database having the same location coordinates. The objects that have acknowledged features under the location term from the database are then chosen as recognized objects as shown in Table 2. The matching is applied by calculating the Euclidean distance between the descriptor vectors of the input object and all objects having the same location in the database. If the distance is closer than 0.8 times the distance of the second nearest neighbor then matching pair is considered to be detected. This threshold value was adapted based on the best result that has been achieved. The output is composed of concatenated strings for both English and Arabic languages. The API of Google cloud speech was used to convert text to audio. This tool was chosen because it supports over 80 languages. Hence, the proposed approach can be globally used based on the supported languages by Google cloud.

Table2: The process of testing the proposed approach. The features of objects for the involved scene (GPS based location) only are extracted and matched with the reference image

Image of real scene	Only the identified objects in the matched area are extracted	Features descriptor in models database	GPS	
			Latitude	Longitude
		Features of paint	X_1	Y_1
		Features of couch	X_2	Y_2
		Features of flower vase	X_3	Y_3

5. Results

The experimental results showed that the proposed approach is able to notify a blind about the surrounding environment such as persons, chairs, doors, tables, or screens. The results of conducted experiments are illustrated in Table 3 by counting the number of correct recognized objects. All conducted experiments are classified in groups counted per object (i.e. number of all faces in the 300 images in the ground truth data is 30 faces, and it was 27 by the proposed work) as shown in Figure 8. The cluttered images covered different scenes of a blind's environment such as bed room, guest room, kitchen, street, and restaurant. The conducted experiments revealed that more than 90% of the detected objects were recognized. The computational time was efficient and varied based on the number of objects in the same scene.

Table 3: The performance of the proposed approach after detecting objects from images taken from the real world

Frame no.	Image scene	Number of objects in real life		The proposed approach		Computational time (msec)
		Faces	Others	Faces	Others	
1	Guest Room	4	8	4	7	1300
2	Bed Room	0	6	0	4	600
3	Street	3	5	3	4	900
4	Restaurant	8	12	7	9	1600
…	…	…	…	…	…	…
300	Kitchen	5	6	4	5	850

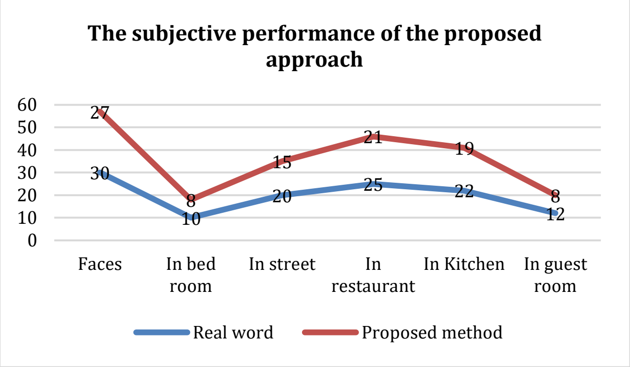


Figure 8: The performance of the proposed method by showing the results of detected objects in 300 images for 25 objects. The comparison is between ground truth objects and the detected objects by the proposed approach

Conclusions and Future Work

In this paper, we have proposed objects recognition approach for helping blinds people. The framework efficiently recognizes objects of interest from cluttered scene. The matching process is utilized using the sign of Laplacian for the underlying interest point and the GPS service to create groups of objects based on their physical existence. Ultra-sonic sensor is used to detect obstacles at medium and long distances, and two camera are employed to classify closer and wider objects regarding to the camera. Future work will aim at recognizing the detected faces. This will help the blind to not only know how many persons are there in his/her frontal scene but also to specifically know their names. In addition, more complex algorithms need to be incorporated to extract more information from cluttered scene, such as: Color of flowers, type of cars, correct key for a door etc.

References

Bellotto, Nicola, and Huosheng Hu (2010), Computationally efficient solutions for tracking people with a mobile robot: an experimental evaluation of bayesian filters. Autonomous Robots 28, no. 4: 425-438.

Kang, Kai, Wanli Ouyang, Hongsheng Li, and Xiaogang Wang (2016), Object detection from video tubelets with convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 817-825.

Tian, Yonglong, Ping Luo, Xiaogang Wang, and Xiaoou Tang (2015), Pedestrian detection aided by deep learning semantic tasks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5079-5087.

Li, Bo, Tianfu Wu, and Song-Chun Zhu (2014), Integrating context and occlusion for car detection by hierarchical and-or model. In European Conference on Computer Vision, pp. 652-667. Springer International Publishing.

Jain, Mihir, Jan Van Gemert, Hervé Jégou, Patrick Bouthemy, and Cees GM Snoek (2014), Action localization with tubelets from motion. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 740-747.

Seema Udgirkar, Shivaji Sarokar, Sujit Gore, Dinesh Kakuste, and Suraj Chaskar (2016), Object Detection System for Blind People. International Journal of Innovative Research in Computer and Communication Engineering 4, no. 9.

Dollár, Piotr, Ron Appel, Serge Belongie, and Pietro Perona (2014), Fast feature pyramids for object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 36, no. 8: 1532-1545.

Felzenszwalb, Pedro F., Ross B. Girshick, David McAllester, and Deva Ramanan (2010), Object detection with discriminatively trained part-based models. IEEE transactions on pattern analysis and machine intelligence 32, no. 9: 1627-1645.

Gould, Stephen, Tianshi Gao, and Daphne Koller (2009), Region-based segmentation and object detection. In Advances in neural information processing systems, pp. 655-663.

Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik (2014), Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587.

Nazli Mohajeri, Roozbeh Raste, and Sabalan Daneshvar (2011),An Obstacle Detection System for Blind People. In Proceedings of the World Congress on Engineering (WCE).

Dionisi, Alessandro, Emilio Sardini, and Mauro Serpelloni (2012), Wearable object detection system for the blind. In Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International, pp. 1255-1258. IEEE.

Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi (2016), You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779-788.

Wang, Liming, Jianbo Shi, Gang Song, and I-Fan Shen (2007), Object detection combining recognition and segmentation. In Asian conference on computer vision, pp. 189-199. Springer Berlin Heidelberg.

Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool (2006), Surf: Speeded up robust features. In European conference on computer vision, pp. 404-417. Springer Berlin Heidelberg