Mody University International Journal of Computing and Engineering Research
2 (2), 2018, 92-96
http://www.publishingindia.com/micer

Analysis on Object Recognition Algorithms and Models to Assist Visually Impaired People

Samiksha Choyal[1*] and Ajay Kumar Singh[2]

[1]Research Scholar, Department of Computer Science and Engineering, Mody University of Science and Technology, Lakshmangarh, Sikar, Rajasthan, India. Email: samikshachoyal95@gmail.com
[2]Assistant Professor, Department of Computer Science and Engineering, Mody University of Science and Technology, Lakshmangarh, Sikar, Rajasthan, India. Email: aksingh.cet@modyuniversity.ac.in
*Corresponding Author

Abstract: Object Recognition play a pivotal role in image processing. Object identification is itself a tedious task as there exist abundant objects which are identical. Vision is one of the most essential senses of the living being, especially a human. The biggest challenging task for a blind person is to move independently without any obstacle. Hence, to recognize an object at various places and scenarios is tough. Visual impairment has been one of the crucial areas of research and as a result a lot of electronic devices and applications have been developed. These devices and applications help blind person to effortlessly identify an object and to take immediate action on the object which is ahead. This paper gives the blueprint of the research and development that has been done for visually impaired people for recognizing objects. The various algorithms or methods that are used in object identification are listed and explained. This paper will have the researchers to know about the latest work that is done in object detection or recognition.

Keywords: Computer vision, Object detection, Object recognition, Visually impaired.

I. Introduction

Object detection has been an influential area of research in computer vision. Object detection was launched in 70s, but in the 90s, it began to track when computers grew vigorous with plentiful application. It is simple for humans, but tough for computers to recognize objects [1]. Adding different aspects of objects and the obscure surroundings make object detection is more ambiguous. All over the universe, there are approximately 36 million people without a vision. This number is not confined it increases rapidly year after year [2]. The intrinsic ability to detect, make a distinction and segregate objects rapidly allow humans to take immediate decisions corresponding to what is seen [3]. The detection of a real-world object is problematic as the objects are hard to model and the variations in texture and color of an object. Also, the background is not constrained where the objects are present [4]. Hence, Fig. 1 describes the fundamental model of object recognition for blind people. The initial input to the object detection system or model is the image or in the case of videos it is a scene. An algorithm or model is selected and then in the next step it search for the matching object image.

In the next step if an object is found an audio output is generated [5].

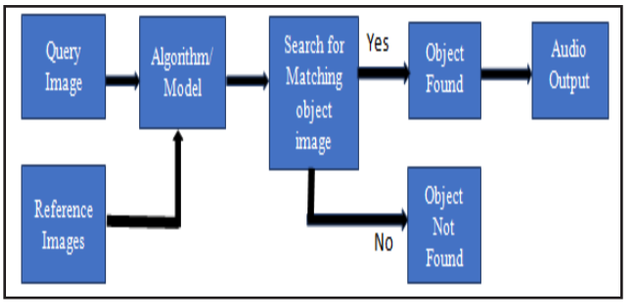

Fig. 1: Fundamental Model of Object Detection for Blind

The technology has served as a lifeline for many people.

The interaction with the objects in the surrounding is restricted for a visually disabled person [6]. One of the important issues for blind people is navigation**.** Despite of the fact that there are several possible substitutes such as memorizing or use a white cane to carry out with the everyday routines. The information provided by them is limited and is not able to detect overhanging hindrance. A simple thing to be considered that it is laborious for them to look for various objects or visit unusual places [7]. The technology has helped a blind person to access independently, figure out and explore the surroundings. But, still challenge remains for the reasons below. First, in various

architectural environments, diversity of design and appearance of objects could be observed. Second, for various object models there are small interclass differences. Third, objects in outdoor scene or nature, have rich texture and color, but mostly the indoor objects are made by man thus have a little texture. Another one is only a few parts of objects is picked up from a wide view variation when there is a movement made by blind user [8]. The advancement in digital cameras, smart phones has made it practical for blind individuals in the development of products based on camera which incorporate computer vision technology. The wearable devices common for blind people these days are depicted in Fig. 2. These are [9]:

- To detect obstruction ultrasonic waves are used in Ultrasonic smart glasses.
- A wearable tactile harness-vest display gives guidance for directional movement with the help of six vibrating motors.
- The ultrasonic sensors are associated to a belt along with the computer which gives direction / instructions through audio. The system demonstrates the user with tactile images around the environment. Further, it converts the optical scene in the form of acoustic or tactile information to grant safe and swift walk.
- Helmet mounted with chips and speakers which amplifies echo of objects located in space is created by ultrasonic sounds.

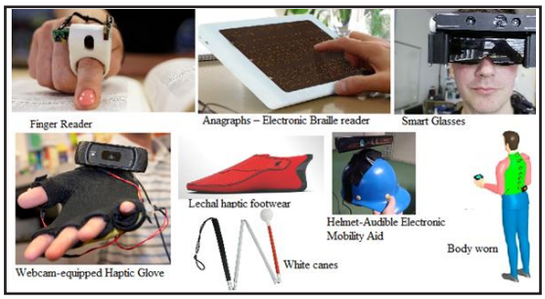

Fig. 2: Wearable Devices for Blind

The other application areas of using various object recognition algorithms are military, robotics, home serve, industries. Fig. 3 shows the graphical representation of the contribution made in different time spans.

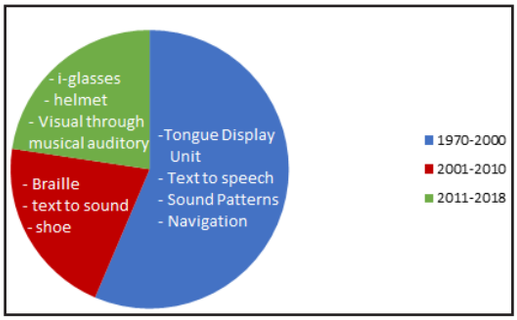

Fig. 3: Technologies Developed for Blind in Different Time Span

II. Approaches and Algorithms: Literature Survey

So far there exist many hardware and software devices or applications for assistance of blind people in their daily life. Such devices and applications are based on various algorithms for recognizing an object for the visually impaired. Now, in this literature some of these algorithms are reviewed. Feature extraction is essential for object detection algorithm to develop accuracy and efficiency. The Scale-Invariant Feature Transform (SIFT) is one such algorithm for detection and description of local feature [10]. This algorithm works basically in four steps. First step is to detect maxima or minima also called as interest points in scale space. In second step location and scale are determined and then key points are selected. These keypoints are most resisted points in an image. The next step is the assignment of orientations of key points depending upon the gradient direction of local image. In the last step alternations in shape distortion and illumination could be seen so that computation of keypoint descriptors is performed [11]. Jabnoun in [12] proposes an algorithm which highlights SIFT in detecting and matching key points for recognition of the objects. The extrema are calculated for each octave from Differences of Gaussian. The Difference of Gaussian (DOG) in equation (1) is calculated by change in two adjacent scales that are separated by k.

$$D(X,\sigma) = (G(X,k\sigma) - G(X,\sigma)) * I(X) \quad (1)$$

Here,

I: input image

X: point X (x, y)

σ: the scale

G (X,**σ**): variable-scale Gaussian

The key points of objects using the SIFT algorithm are build and local dissimilarity map is calculated. A set of everyday object images was selected, and feature points were obtained from each image. Further, keypoint descriptors are obtained for each frame as this recognition of objects in done for video. Key-points of frames are matched with the objects and identification of the object detected is done. Identical and non-identical frames are identified for the next step. A video file is launched for various objects that are detected. There were 35 percent true positive and 65 percent false positive depending on 3 scales, whereas on the scale of 5 there were 95 true matches and 5 percent false matches as shown in Table I.

Table I: Matches Depending on Scale Number

No. of Scales	True Positive(%)	False Positive (%)
3	35	65
5	95	5

The illumination level varies from one video scene to another and decrease in the number of correct matches could be

observed. But, the conclusion drawn is that SIFT is constant to illumination changes and invariant to rotation, translation and scaling. The performance of SIFT descriptor is better as compared to local descriptors. Though there are a few limitations are that in an image with noise this algorithm can find less key points, hence the image used must be clear and sharp for performance to be good. The matching reliability decreases if the keypoints are more in count for a query image in comparison to reference image.

Another detector-descriptor feature-based algorithm for detection is Speeded Up Robust Feature (SURF) which manages the properties of an image by accelerating the process of key-point localization [13]. Chucai Yi in [14] states that SURF's standard version gives a faster outcome as compared to SIFT. The excellence in performance in terms of accuracy and time of this detector depends upon Hessian matrix as defined in an image I with a point x (x, y) given as in equation (2).

$$\text{'H}\,(x,\sigma) = \begin{bmatrix} \text{Lxx}\,(x,\sigma) & \text{Lxy}\,(x,\sigma) \\ \text{Lxy}\,(x,\sigma) & \text{Lyy}\,(x,\sigma) \end{bmatrix} \tag{2}$$

where, $L_{xx}(x, \sigma)$, $L_{xy}(x, \sigma)$ and $L_{yy}(x, \sigma)$ are the convolution of Gaussian second order derivative.

SURF descriptor on the other hand can be based on Haar wavelets. After the calculation of response of wavelet, a square patch division takes place in sub blocks of 4×4. Each sub block is further used in the construction of a descriptor depending around the interest points keeping spatial info for the final descriptors. Testing was done on multiple input images and successful results were obtained. The features matching in the query image to the image taken as reference as shown with white lines. The detection of an object takes place when the points matching is higher in number than the thresholds. There were no false results obtained even though the objects were blocked or hindered by other objects. But, when the object is taken through camera this algorithm was not able to find ample number of matching points and thus determine the expected object. The various conditions under which images are captured also matter. Hence, this algorithm gave an accuracy of 69 percent. The performance of a descriptor based on SIFT was higher as compared to the SURF- based descriptor with dimensions 128 and 64 respectively. One more point where the algorithm failed is it could not find enough matches in an image on pre-trained thresholds because of less number of the distinguishable features.

Till now the above two algorithms were used in indoor objects and in gray-scaled images, but the next approach is followed for outdoor objects as well. S. R. Bagwan in [15] made use of the hybrid algorithm to get fast results and reduce computation time. A dedicated object recognition android application named 'VisualPal' is built for blind individual which identifies the object when major colors are used, and brightness is maximized. The built-in camera of mobile phone is used for scanning and detecting an object in the image taken by the camera. A constraint of carrying a hardware or device along with them has been overcome. The communication with the blind person takes place with the verbal message recorded previously. The combination of Euclidean Distance and Artificial Neural Network is used in this hybrid algorithm. The color information along with the path of maximum brightness is used by this application for detection. The mathematical form of Artificial Neural Network and Euclidean Distance measures is represented below:

A neuron with k inputs transforming a set $A \subset IR^n$ of input is a function (3).

$$F : IR^n \times A \ni (\vec{u}, \vec{a}) \to F(\vec{u}, \vec{a}) = f\big((\vec{u}, \vec{a})\big) \in IR \tag{3}$$

Here, v is weight vector

f: IR → IR is known as activation function of a neuron. A neuron is called linear if operator f is linear. The function is in (4).

$$F^* := F\big(\vec{U}, .\big): A\ni \vec{a} \to F^*(\vec{a}) \in IR \tag{4}$$

is called trained k-neuron on A.

The Euclidean distance between point U (U1, U2,…) and a point V (V1, V2,…) is given by the function (5).

$$d = \sqrt{\sum_{i=1}^{n}(u_i - v_i)^2} \tag{5}$$

The author has also described a graph with the comparison of accuracy when the combination of ANN and Euclidean is used and the ANN alone. The accuracy obtained is 97.5 percent with the hybrid algorithm.

You Only Look Once (YOLO) is a unified approach for detection of objects is presented by Redmon in [16]. YOLO is considered as a fastest object detector which can be used for general purpose applications which depend on robust, fast detection of objects. A neural network foresees class probability and bounding box directly from an image in one assessment. Because the detection pipeline network is single, optimization can be done from end-to-end on completion of detection. The performance is outrun than other methods of detection, which include R-CNN. The object detection is framed as a single regression hence complex pipeline is not required. With the help of this method system will tell what the object is and the location of it. This approach considers the whole image during the training and testing system unlike other techniques. Therefore, it encodes implicitly the contextual information of classes along with their appearances. The features of the whole image are used for prediction of bounding box for various classes of images at the same time. The division takes place in this system as a S×S grid for the input image. The grid cell has the responsibility to detect object if the object's center falls in the grid cell. The confidence score and bounding boxes are predicted for each grid cell. The implementation of this model is done on PASCAL VOC 2007, Picasso and People-Art dataset. This network finds features from the initial convolutional layer while probability and coordinates of outcome are predicted by fully connected layers. In this network twenty-four convolutional layers succeeded by two fully connected layers. YOLO is connected with webcam to check that the performance is maintained in real-time and also the time taken for fetching

images and displaying detection as they move around. With the combination of Fast R-CNN and YOLO an improvement of 2.3 percent is obtained. The quantative results obtained are shown in Table II.

Table II: Results Based on Different Datasets

	VOC 2007 AP	Picasso AP	People-Art AP
YOLO	59.2	53.3	45
R-CNN	54.2	10.4	26
DPM	43.2	37.8	32

The drawback of this model is that it tries hard with mini objects which arrive in groups. The incorrect localization also becomes origin of error.

A camera-based system for hand held objects is presented by Deshpande in [2] for visually impaired to read text printed on them. After the text is read, it is converted in the form of speech. Maximally Stable External Regions (MSER) feature is used for detecting the text which is more stable. The pattern of text is localized with Optical Character Recognition (OCR). Two steps are used in the processing of data in this prototype. First, the text patterns are detected from the image with Object of interest. The second step is text localization for obtaining text regions and then recognizes them. Finally, the output is in the form of audio for the user. The characteristic function for a region is given as in equation (6).

$$f(t) = |I(t) - I_0|/((1/t)\int|I(t) - I_0|) \tag{6}$$

The region of approximation uses ellipse of the second order which is further used at the moment of the region depicted in the equations below:

$$m_{uv} = \iint x^u y^v f(x,y)dxdy \tag{7}$$

$$v = Bu \tag{8}$$

$$\sum_2 = B\sum_1 B^T \tag{9}$$

$$u = [x,y]^T \tag{10}$$

where, u is relative to the center of mass.

The high performance can be achieved with MSER and OCR algorithm for detecting and recognizing text in a different background. The analysis of text detection algorithms is depicted in Table III. The bright and dark intensity pixels play an important role in text recognition. It does not perform well with motion blur.

Table III: Analysis of Text Detection Methods

Detector	Feature Type	Scale Independent
FAST	Corner	No
Minimum Eigen Value Algorithm	Corner	No
Corner Detector	Corner	No
SURF	Blob	Yes
BRISK	Corner	Yes
MSER	Region with uniform intensity	Yes

III. Conclusion

This paper reviewed different algorithms and models for object recognition. These methodologies are used for object detection and recognition. The approaches used for recognizing an object for the blind include SIFT, SURF, Hybrid, YOLO. For text recognition of various objects can be done with OCR and MSER. The images are captured with a camera and auditory display will assist blind individuals to know about the objects or text that has been identified. The system is designed for making life of blind person comfortable and easy. Hence, an algorithm with high performance and more efficiency will be chosen for making own software for visionless people to identify objects and text with own dataset in real-time. The various limitations such as color, background, time will be kept in mind while designing the system.

References

[1] S. Prasad, and S. Sinha, "Real-time object detection and tracking in an unknown environment," *World Congress on Information and Communication Technologies*, pp. 1056-1061, 2011.

[2] S. Deshpande, and R. Shriram, "Real time text detection and recognition on hand held objects to assist blind people," *International Conference on Automatic Control and Dynamic Optimization Technology (ICACDOT'2016)*, pp. 1020-1024, 2017.

[3] B. A. G. de Oliveira, F. M. F. Ferreira, and C. A. P. da S. Martins, "Fast and lightweight object detection network: Detection and recognition on resource constrained devices," *IEEE Access*, vol. 6, pp. 8714-8724, 2018.

[4] A. Jothimani, S. Edward, G. K. Divyashree, and Laavanya, "Object identification for visually impaired," *Indian Journal of Science and Technology,* vol. 9, no. s(1), December 2016.

[5] K. U. Sharma, and N. V. Thakur, "A review and an approach for object detection in images," *International Journal of Computational Vision and Robotics*, vol. 7, no. 2, pp. 196-237, 2017.

[6] S. Gautam, K. S. Sivaraman, H. Muralidharan, and A. Baskar, "Vision system with audio feedback to assist visually impaired to grasp objects," *Procedia Computer Science*, vol. 58, pp. 387-394, 2015.

[7] N. Bari, N. Kamble, and P. Tamhankar, "Android based object recognition and motion detection to aid visually impaired," *International Journal of Advances in Computer Science and Technology*, vol. 3, no. 10, pp. 1-5, 2014.

[8] B. Kaur, and J. Bhattacharya, "A scene perception system for visually impaired based on object detection and classification using multi-modal DCNN," *Journal of Neurocomputing*, 2018.

[9] Y. Tian, X. Yang, C. Yi, and A. Arditi, "Toward a computer vision-based wayfinding aid for blind persons to access unfamiliar indoor environments," *Machine Vision and Application*, vol. 24, no. 3, pp. 521-535, 2013.

[10] R. Jafri, S. Abid, and A. Hamid, "Computer vision-based object recognition for the visually impaired in an indoors environment: A survey," Springer, 2013.

[11] A. P. George, "Object recognition algorithms for computer vision system: A survey," *International Journal of Pure and Applied Mathematics*, vol. 117, no. 21, pp. 69-74, 2017.

[12] H. Jabnoun, F. Benzarti, and H. Amiri, "Object detection and identification for blind people in video scene," *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA),* December 2015.

[13] R. Chincha, and Y. Tian, "Finding objects for blind people based on SURF features SURF-based," *IEEE International Conference on Bioinformatics and Biomedicine Workshops,* pp. 526-527, 2011.

[14] C. Yi, R. W. Flores, R. Chincha, and Y. Tian, "Finding objects for assisting blind people," *Network Modeling and Analysis in Health Informatics and Bioinformatics*, vol. 2, no. 2, pp. 71-79, 2013.

[15] S. Md. R. Bagwan, and L. J. Sankpal, "VisualPal: A mobile app for object recognition for the visually impaired," *IEEE International Conference on Computer, Communication and Control (IC4'2015)*, 2015.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.