

Haar Cascade Algorithm for the Visually Impaired to Detect and Recognize Objects

Priya Samyukthaa. M^[1], Siva Bharathi.K^[2] Sivagami.A^[3]

U.G Student, ECE^[1], U.G Student ECE^[2], Assistant Professor – ECE^[3]

Sri Manakula Vinayagar Engineering College
Puducherry, India

Abstract – The current survey on the population of visually impaired people around the world shows that there is a steady increase in blindness and nowadays much of the research has been focused on visually challenged people. There is a wide range of blind navigation and guidance systems which are used to detect objects, staircase and change in ground level. But these systems are used only for detection. Besides guidance and navigation, there are lots of challenges faced by the visually impaired people. Though radio frequency identification is used to recognize objects, there are certain limitations such as complexity and it is inapplicable in an outdoor environment. Hence an alternate system has been proposed which not only detects but also recognize various objects using a simple Haar cascade algorithm in open source computer vision (OpenCV). The proposed system has Raspberry pi 2 model B, a camera and an audio feedback device. Initially the image captured by the camera is processed and matched with the features stored in the database and when the match occurs the output is conveyed to the user through a voice signal. Thus the proposed system increases the confidence of blind people by giving them the proper information of the objects around them.

Keywords – OpenCV, Haar cascade algorithm, Raspberry pi, database, camera, audio feedback device.

I. INTRODUCTION

Worldwide around 285 million people are visually impaired out of which 39 million are blind and 246 million have severe or moderate visual impairment. There are lot of techniques for guidance and navigation that are been established since 17th century. White cane and guide dogs are popular guidance system in 18th century. Later in 19th and 20th century, after the invention of many electronic devices, there established many detection devices, GPS based navigation system and sensors to detect change in ground level.

In smart stick, there are two types of sensors. They are the ultrasonic and proximity sensors. The ultrasonic sensor is used to detect the object from particular distance by sending ultrasonic waves which is reflected from the object. Hence, based on the reflection the distance is calculated and the output is processed as vibrations. The proximity sensor is used to detect the change in ground level.

In IR sensor based smart stick, a pair of infrared sensor is used. The horizontal one is to detect obstacles in front of the blind in the range of 200cm, whereas the inclined

infrared sensor is to detect obstacles on floor, upward and downward stairs.

The Radio frequency identification (RFID) of objects for visually impaired has been an emerging technology in recent years. The basic RFID system consists of reader and tags. RFID uses radio waves to communicate between a tag, which stores information and a reader, which interprets the information. This system can be incorporated in the form of glove which will be worn by the visually impaired and has a RFID reader. When the hand with the glove is moved towards an object, voice information of the object will be played through the speaker attached to the gloves.

There are certain limitations in these existing systems. The smart stick is limited by multiple reflections that occur due to ultrasonic sensor. Even though the IR sensor based smart stick detect objects effectively, it cannot recognize objects. The RFID based system is costly and not feasible in an outdoor environment. Moreover, the RFID does not recognize same objects with different features.

II. LITERATURE SURVEY

Detection and recognition of objects in real-time has become one of the most important applications in image processing. A variety of techniques have been developing over the years but improvement is still required in order to achieve efficiency and accuracy. A survey of the existing algorithms on detection and recognition are discussed below.

Initially Kazuyuki Miyazawa et al., [1] presented an efficient algorithm for iris recognition using phase-based image matching. A major approach for iris recognition is to generate feature vectors corresponding to individual iris images and to perform iris matching based on some distance metrics. One of the difficult problems in feature-based iris recognition is that the matching performance is significantly influenced by many parameters in feature extraction process, which may vary depending on environmental factors of image acquisition. A new method based on the simplified model of PCNN was proposed by S. Wei, Q. Hong and M. Hou, [2] where the images segment automatically. The parameter settings are studied to ensure that the threshold decay of Simplified Pulse Coded Neural Network(S-PCNN) would be adaptively adjusted according to the overall characteristics of the image.

The PCNN with fast links among neurons can be implemented with the help of FGPA chip. Hualiang Zhuang et al., [3] proposed a PCNN with multichannel (MPCNN) linking and feeding fields for color image segmentation. With respect to the spectral feature vectors and spatial proximity, the computing of the color image segmentation can be implemented in parallel on a FPGA chip. The parallel neural circuits improve the speed of processing drastically as compared with the sequential-code-based counterparts.

Later Abbas Zohrevand, et al., [4] proposed a more efficient algorithm using SIFT method on image to extract key points and the corresponding descriptors. Then the extracted descriptors are represented in form of (Attributed Graph) AG graphs. So, for a scene image and the model of objects the two different graphs referred as scene and model graphs are considered. Hamid Ali Abed AL-Asadi et al., [5] considered the environmental effects those results in change in shape of an object. Hence they proposed an Artificial Fish Swarm Algorithm is a class of an evolutionary optimization technique with three types of classifier combinations using different geometrics shape for the recognition of the plant leaves. Fish Swarm Algorithm is applied on Fourier descriptors to get optimum weights that maximize the recognition rate.

The merging of two main algorithms results in processing of high resolution image at faster rate which was done by Yuli Chen et al., [6] who merged region-based object recognition (RBOR) method to identify objects and SPCNN for color image segmentation. This method performs color image segmentation by a SPCNN for the object model image and test image, and then conducts a region-based matching between them. The proposed SPCNN-RBOR method overcomes the drawback of feature-based methods that inevitably includes background information into local invariant feature descriptors when key points locate near object boundaries. Merging of algorithm is complex and difficult to predict error at different stages.

Hence an alternate object detection and recognition system has been proposed using Haar cascade algorithm in OpenCV which allows the user to detect and recognize the objects. This would allow them to improve their navigational ability and to retain their independence.

III. BASICS OF DETECTION TECHNIQUES

A. Image processing

The analysis and manipulation of digitized image so as to enhance its quality is known as image processing. The various steps involved in image processing is illustrated in Fig. 1 which include,

1) Image Acquisition

Image Acquisition is the initial step involved in the image processing technique. It is nothing but the digitization of the original image with the help of scaling process.

2) Image Enhancement

The main aim of enhancement is to highlight certain features in the image by the process of changing brightness and contrast.

3) Image restoration

Image restoration deals with improving the appearance of the image whose techniques are based on mathematical or probabilistic models of image degradation.

4) Color Image processing

Color image processing is done through color modeling and processing in digital domain in order to improve the color quality.

5) Wavelets and Multiresolution processing

Wavelets are used for representing images in various degrees of resolution. In this the images are subdivided into smaller regions for data compression and pyramidal representation.

6) Compression

Compression is the process of reducing the storage in order to save an image or a bandwidth to transmit it.

7) Morphological processing

Morphological processing deals with the tools used for extracting image components which are helpful in the representation and description of shape.

8) Segmentation

Segmentation procedures include partition an image into its constituent parts. There are two types of segmentation,

- Autonomous segmentation
- Rugged segmentation

9) Representation and description

The representation and description is usually an output of segmentation stage which is a raw pixel data.

10) Object recognition

Object recognition is the process of assigning a label to an object based on the descriptors.

11) Knowledge base

Knowledge base is detailing regions of an image where the information of interest is known to be located.

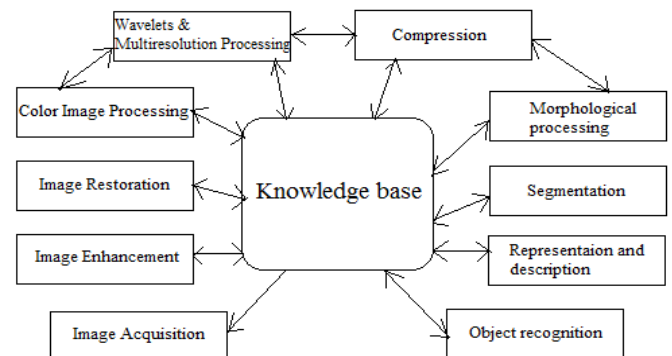


Fig. 1 Block diagram of Image Processing

B. OpenCV

Open Source Computer Vision (OpenCV) was started at Intel in 1999 by Gary Bradsky, and was first released in 2000. OpenCV supports a wide variety of programming languages such as C++, Python, Java, etc., and is available on different platforms including Windows, Linux, Android, and iOS.

Image processing in OpenCV includes the following steps,

1) *Changing Color spaces*

There are more than 150 color-space conversion methods available in OpenCV. The most widely used are BGR \leftrightarrow Gray and BGR \leftrightarrow HSV. The function used is cv2.cvtColor (), cv2.inRange ()

2) *Geometric Transformations of Images*

There exist different geometric transformations to images like scaling, translation, rotation, affine transformation, resizing, etc. These functions are available in cv2.getPerspective Transform.

3) *Image Thresholding*

In Image thresholding, if pixel value is greater than a certain threshold value, it is then assigned a value (which may be white), or else it is assigned another value (which may be black). The function used is cv2.threshold.

4) *Smoothing Images*

Image smoothing is achieved by convolving the image through a low-pass filter kernel which removes the noise. It actually removes high frequency content (e.g.: noise, edges) from the image therefore the edges are blurred when this is filter is applied.

5) *Morphological Transformations*

Morphological transformations normally performed on binary images. It needs two inputs, one is the original image, and the second one is called structuring element or kernel which decides the nature of operation. Two basic morphological operators for transformations are Erosion and Dilation.

6) *Template Matching*

Template Matching is a method used to search and find the location of a template image in a larger image. OpenCV comes with a function cv2.matchTemplate () for this purpose. It slides the template over the input image and compares the template and patch of input image under the template image.

IV. PROPOSED SYSTEM

A. *Haar Algorithm*

Haar feature-based algorithm is an effective method proposed for Face detection. But it is also an efficient method for object detection by the collection of more positive and negative samples. It is a machine learning based approach. Haar features consider adjacent

rectangle shaped regions at a specific location in a detection window, and then adds up the pixel intensities in each specific region and the difference between these sums is calculated. This difference is used to categorize the portions of the image.

In the detection phase, a window of the target size is slided over the input image, and for each section of the image the Haar-like feature is calculated. This difference is then compared to a certain threshold that separates non-objects from objects.

B. *Computation of Haar-like features*

Integral images (summed area tables) can be defined as two-dimensional lookup tables. It can be in the form of a matrix with the same size as that of the original image. Each element of the integral image holds the sum of all pixels located on the up-left region of the original image.

The sum of shaded rectangular area is shown in Fig.2. This allows computing sum of rectangular areas in the image, at any position using four lookups.

$$\text{Sum} = I(C) + I(A) - I(B) - I(D)$$

where the points A,B,C,D belong to the integral image.

The 2-rectangle features need six lookups, eight lookups are needed for 3-rectangle features and 4-rectangle features need nine lookups.

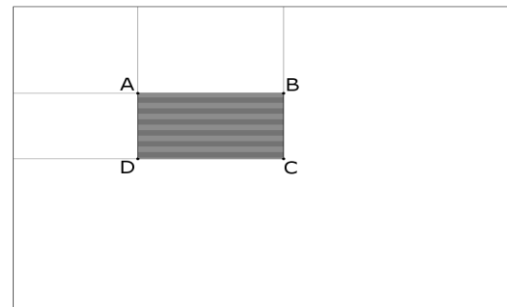


Fig. 2 Finding the sum of shaded rectangular area

C. *Cascade classifier*

The cascade classifier consists of a list of stages, in which each stage consists of a list of weak learners. The system detects objects by moving a window over the image. Each stage of the classifier labels the specific region as either positive or negative. If object was detected then the output is positive and negative means that the specified object was not detected in the image. If the labeling yields a negative result, then the classification of the specific region is thereby complete and the location of the window is moved to the next location.

If the labeling produces a positive result, then the region moves on to the next stage of classification. The classifier gives a final verdict of positive, when all the stages of the classifier, including the last one, give a result, saying that the object to be detected is found in the image

The various stages of cascade classifier is illustrated in Fig. 3.

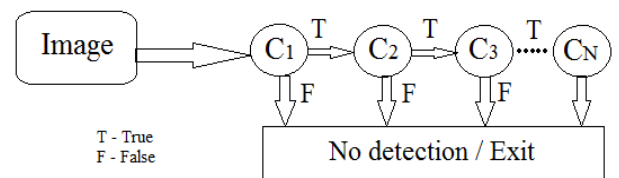


Fig. 3 Stages of cascade classifier

- . There are four cases in output consideration,
- A true positive - the object in question is indeed present in the image processed and the classifier labels it as such which indicates a positive result.
- A false positive - the labeling process falsely determines that the object is located in the image, even though it is not.

- A false negative - the classifier is unable to determine the actual object from the image.
- A true negative – the classifier does not detect an object even it is present in the detection window.

D. Training cascade

There are various steps in training cascade which includes collection of samples, creation of samples using a tool in OpenCV and merging of files.

1) Collection of samples

This process include collection of positive and negative samples which means a lot of images that show the object that has to be detected (positive sample) and even more images without the object which are the negative samples.

2) Positive images

Positive images are the images of an object that has to be recognized. Those images can be captured through camera or collected from internet or else extracted from a video. The collected images should differ in lighting and background.

Once the pictures are collected, it has to be cropped so that only the desired object is visible. All the collected positive images should have almost equal ratio. The positive, cropped images are inserted in the **./positive_images** directory.

3) Negative images

The negative image is the image that looks exactly like positive image, but they do not contain the object that has to be recognized. For a high accurate classifier, around 600 negative images are to be taken.

Once the negative images are taken, that has to be inserted in the **./negative_images** folder of the repository. The list of negative samples is stored in **negatives.txt**.

4) Creating samples

Creating samples is applying transformation and distortion in the positive images. The best way to generate more number of samples is to use a tool of OpenCV which is **opencv_createsamples**.

5) Merging files

The next thing is to merge all the *.vec files present in the sample directory. In order to merge, **mergevec.cpp** tool has to be used, which is included in the **src** directory of the repository, to combine them all into one *.vec file.

The final merged file is known as XML file.

E. eSpeak

eSpeak is a compact open source software speech synthesizer for English and other languages. It supports Windows and Linux. It utilizes a formant synthesis method that allows many languages to be provided in a small size. The speech is clear, and can be used at high speeds.

V. HARDWARE DESCRIPTION

An overview of the hardware components used in the project. It gives detailed description of the hardware used: Raspberry Pi 2 model B, webcam and earphones.

A. Raspberry Pi 2 Model B

Raspberry Pi is a portable and powerful minicomputer. The board length is only 85 mm and width is only 56 mm. Its size is just as big as a credit card but it is a capable PC. It can be used for high-definition video, spreadsheets, word-processing, games and more. The Raspberry Pi Model B provides more GPIO, more USB than version 1. It also improves power consumption, audio circuit and SD card.



Fig. 4 Raspberry Pi 2 Model B Board

i) Power

The device is powered by a 5V micro USB supply. The current required by the Raspberry Pi is dependent on the devices connected to it. A 1A or 1.2A (1000mA or 1200mA) power supply is suitable for running the device.

The power requirements of the Raspberry Pi depend on the utilization of the various interfaces on the Raspberry Pi. The HDMI port uses 50mA, the camera module requires 250mA, and keyboard take around 100-500mA.

ii) USB

The Raspberry Pi Model B is equipped with four USB (Universal Serial Bus 2.0) ports. A USB device connected to the single upstream USB port on BCM2835.

The USB ports enable the attachment of peripherals such as keyboards, mike, webcams that provide the Pi with additional functionality.

The USB host port inside the Pi is an On-The-Go (OTG) host as is the application processor which powers the Pi. OTG in general supports communication to all types of USB device, but for most of the USB devices that might be plugged into a Pi, additional software may be required to provide an adequate level of functionality. This however causes the system software load to increase.

B. USB Web Camera

A webcam is a video camera that streams image in real time to a computer or to a computer network. When "captured" by the computer, the video stream may be saved and processed as required. A webcam is generally connected by a USB cable, or similar cable. In this project we use Creative Live! Cam Socialize webcam for video input. The USB 2.0 capability ensures high frame rate video capture of up to 30 fps at 800 x 600 resolution.

C. Earphones

Earphones are portable and convenient. Earphones are made in a range of different audio reproduction quality capabilities. Those designed for telephone use typically cannot reproduce sound signal with the high fidelity when compared to the expensive units designed especially for music listening by audiophiles. Earphones that use cables typically have either a 1/4 inch (6.35mm) or 1/8 inch (3.5mm) phone jack for plugging them into the audio source.

D. Interfacing Raspberry Pi and Webcam

A standard USB webcam can be used to take pictures and video on the Raspberry Pi. In the proposed system we use Creative Live! Cam Socialize webcam for video input. The following sequences are to be considered while installing a webcam,

- `sudo apt-get install fswebcam` - enter the command `fswebcam` which is to be followed by a filename and a picture will be captured using the webcam, and saved to the filename.
- `fswebcam image.jpg` - filename of the picture taken.
- `fswebcam -r 640x480 image2.jpg` - the webcam used has a resolution of 640 x 480 so to specify the resolution of the image to be taken at, use the `-r` flag:
- `fswebcam -r 640x480 --no-banner image3.jpg` - Addition of `no-banner` flag

E. Interfacing Raspberry Pi and Earphones

The Raspberry Pi has two audio output modes:

- HDMI
- Headphone jack.

It is possible to switch between these modes at any time. If the HDMI monitor or TV has built-in speakers, then the audio can be played over the HDMI cable, however it can be switched over to a set of earphones or other speakers plugged into the headphone jack 3.5mm. If the display has speakers, sound is output via HDMI by default; if not, its output is via the headphone jack.

VI. RESULTS AND DISCUSSIONS

A. Complete setup

The final portable device for object detection and recognition includes a USB web camera, head phones and a power bank interfaced with the Raspberry pi board.



Fig. 5 Complete setup of the proposed system

B. Detection of objects

The following are the five objects detected and recognized and the work can be extended to multiple objects.

Once the object is detected it is matched with features stored in the cascade classifier. If the features are matched then the recognition of the object is given in the form of audio output using speaker or headphones.

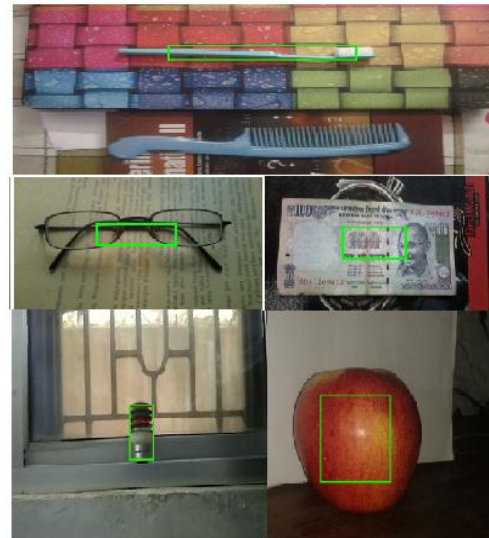


Fig. 6 Detected objects

VII. CONCLUSION

The object detection and recognition system uses a simple Haar cascade algorithm in open source computer vision (OpenCV). This system has Raspberry Pi 2 model B, a camera and an audio feedback device. Initially the image captured by the camera is processed and matched with the features stored in the database and when the match occurs the output is conveyed to the user through a voice signal. Haar algorithm is simple and it promotes high speed detection if more number of positive and negative samples is added. Although the radio frequency identification can be used to recognize objects, there are certain limitations such as complexity and it does not recognize same objects with different features which makes it inapplicable in an outdoor environment. The visually challenged people can easily handle this device as it is portable and easy to operate. This system is also useful in an outdoor environment. Thus the proposed system increases the confidence of blind people by giving them the proper information of the objects around them and enabling them to move independently in indoor and outdoor environment.

VIII. FUTURE ENHANCEMENT

The number of objects to be detected can be increased easily making the system useful in real world environment. The system can also be improved by finding the distance of the detected object, making it easier for the user to locate the object. Text detection and character recognition can also be implemented which could be used in identification of the detected object by the user.

REFERENCES

- [1] Kazuyuki Miyazawa, Koichi Ito, Takafumi Aoki Koji Kobayashi, Hiroshi Nakajima, "An Efficient Iris Recognition Algorithm Using Phase-Based Image Matching", *IEEE Journal on Information Sciences*, vol. 2, Sep 2005.
- [2] S. Wei, Q. Hong, and M. Hou, "Automatic image segmentation based on PCNN with adaptive threshold time constant", *Neurocomputing*, vol. 74, no. 9, pp. 1485–1491, Apr 2011.
- [3] Hualiang Zhuang, Kay-Soon Low and Wei-Yun Yau, "Multichannel Pulse-Coupled-Neural-Network-Based Colour Image Segmentation for Object Detection", *IEEE Transactions on Industrial Electronics*, Vol. 59, No. 8, Aug 2012.
- [4] Abbas Zohrevand, Alireza Ahmadyard, Aliakbar Pouyan and Zahra Imani, "A SIFT Based object recognition using contextual Information", *Iranian Conference on Intelligent Systems (ICIS)*, pp. 1 – 4, Feb 2014.
- [5] Hamid Ali Abed AL-Asadi, Majida Ali Abed "Object Recognition Using Artificial Fish Swarm Algorithm on Fourier Descriptors", *American Journal of Engineering, Technology and Society*, 2015.
- [6] Yuli Chen, Yide Ma, Dong Hwan Kim, and Sung-Kee Park, "Region-Based Object Recognition by Colour Segmentation Using a Simplified PCNN", *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 26, No. 8, Aug 2015.
- [7] Alexander Toshev, Ameet Makadia and Kostas Daniilidis, "Shape-based Object Recognition in Videos Using 3D Synthetic Object Models", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 288-295, May 2009.
- [8] Anuj Srivastava, Eric Klassen, Shantanu H. Joshi, and Ian H. Jermyn, "Shape Analysis of Elastic Curves in Euclidean Spaces", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 7, Jul 2011.
- [9] Seungwon Lee, Junghyun Lee, Monson H. Hayes and Joonki Paik, "Adaptive Background Generation for Automatic Detection of Initial Object Region in Multiple Color-Filter Aperture Camera-Based Surveillance System", *IEEE Transactions on Consumer Electronics*, Vol. 58, No. 1, Feb 2012.
- [10] Reza Oji, "An Automatic Algorithm for Object Recognition And Detection Based on ASIFT Keypoints", *Signal & Image Processing: An International Journal (SIPIJ)* Vol.3, No.5, Oct 2012.
- [11] Dung Phan, Chi-Min Oh, Soo-Hyung Kim, In-Seop Na and Chil-Woo Lee, "Object Recognition by Combining Binary Local Invariant Features and Color Histogram", *Second IAPR Asian Conference on Pattern Recognition*, pp. 460-470, Nov 2013.
- [12] Yang Liu, Youngkyoon Jang, Woontack Woo and Tae-Kyun Kim, "Video-based Object Recognition using Novel Set-of-Sets Representations", *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 533-540, Jun 2014.
- [13] Yan Zhuang, Member, IEEE, Xueqiu Lin, Huosheng Hu, Senior Member, IEEE, and Ge Guo Member, IEEE, "Using Scale Coordination and Semantic Information for Robust 3-D Object Recognition by a Service Robot", *IEEE Sensors Journal*, vol. 15, no. 1, Jan 2015.
- [14] Ali Borji, "What is a Salient Object? A Dataset and a Baseline Model for Salient Object Detection", *IEEE Transactions on Image Processing*, Vol. 24, No. 2, Feb 2015.
- [15] Hyun Oh Song, Ross Girshick, Stefan Zickler, Christopher Geyer, Pedro Felzenszwalb, and Trevor Darrell, "Generalized Sparselet Models for Real-Time Multiclass Object Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 5, May 2015.
- [16] Raspberry Pi Foundation - <https://www.raspberrypi.org>
- [17] Adrian Rosebrock - <http://www.pyimagesearch.com>
- [18] Ashwin Pajankar, *Raspberry Pi Computer Vision Programming*, Packt Publishing, 2015, pp. 30-39.
- [19] Tim Cox, *Raspberry Pi Cookbook for Python Programmers*, Packt Publishing, 2014, pp. 49-52.