IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# ASSISTANCE FOR VISUALLY IMPAIRED IN OBJECT DETECTION AND RECOGNITION

Sharan L Pais, Akshatha A M, Akshitha Poojary, Manish R, Nanda Kishore V
Assistant Professor, Student, Student, Student, Student
Department Of Information Science and Engineering
Alva's Institute of Engineering and Technology, Moodbidri, India

**Abstract**

This paper introduces a camera-based portable assistive text reading framework to assist blind people in reading text labels and product packaging from hand-held items in their everyday lives. Three functional elements comprise the device framework: First, scene capture—with a small camera, the text that the user wants to read is captured as an image and sent to the image or data processing system.

**Introduction**

There are over a billion visually impaired people on the planet. Even in developed countries like the United States, a large portion of this population is blind. According to a national health interview study conducted in 2008, over 85 percent of adult Americans are blind. In recent years, through developing camera-based devices that combine computer vision technology with already-existing products such as optical character recognition, advancements in computer vision, digital cameras, and portable computers assist these individuals.

Few devices are capable of providing better access to common hand-held items such as product packages and text-printed objects. Developing devices that are more compact and advanced will help people live independently and achieve economic and social independence.

The location of the point of interest inside the camera view is the most difficult aspect of an assistive reading device for blind people. A camera with a large angle is used as an estimated solution to focus the target inside the camera view. Text from the surrounding area is often used. To extract hand-held objects from an image, we proposed a motion-based method for isolating the region of interest, with text recognition limited to that location.

**Related work**

It's based on a good binarization and enhancement technique, then a good connected component analysis method. The final binary images, which primarily consist of text regions, are described using connected component analysis.

Assistive Text and Product Labels Using a Portable Camera Reading from hand-held objects for blind people-propose a motion-based approach for defining a region of interest in a video by asking the user to shake the object to isolate objects from the cluttered context.

A context subtraction method based on a mixture of Gaussians is used to remove the moving object region in this method. To obtain text information, text localization and recognition are performed in the extracted

ROI. We propose a novel text localization method to automatically localise the text regions from the object ROI.

In an Adaboost model, a localization algorithm is learned by studying gradient features of stroke orientations and distributions of edge pixels. Off-the-shelf optical character recognition software binarizes and recognises text characters in the localised text regions. The understood text codes are spoken aloud to blind people.

Compact Camera Based Assistive Text Product Label Reading and Image Identification for Hand-Held Items for Visually Challenged People- proposes a camera-based assistive text reading framework, as well as processing and obtaining object information, to assist blind people in reading text labels and product packaging from hand-held objects in their everyday lives. Raspberry Pi is at the heart of it.

**Preliminaries**

(i) Optical Character Recognition (OCR)

The identification of printed or written text characters by a machine is known as optical character recognition (OCR). This entails character-by-character photo scanning of the text, review of the scanned-in image, and then translation of the character image into character codes, such as ASCII, which are widely used in data processing.

Our eyes and brain perform optical character recognition without us even realising it when we read from a computer screen. Our eyes understand the patterns of light and dark that make up the characters (letters, numbers, and punctuation marks) written on the screen, and our brain deduces what the other person is trying to mean.

Computers are capable of doing this as well, but it is extremely difficult for them to do so. The first issue is that computers lack eyes, so if we want them to read anything like a page from an old book, we must first provide them with a picture of that page, which can be created using an optical scanner or a digital camera. The page we generate in this manner is a graphic file (often in the form of a JPG) that, to a machine, is identical to a photograph of any monument or any other graphic: it's a meaningless pattern of pixels (the coloured dots or squares that make up any computer graphic image). To put it another way, the machine has a view of the screen rather than the text—it can't read the words on the page like we can. OCR is the method of converting a scanned JPG of a printed or handwritten page into text (in other words, creating a TXT or DOC file).

(ii) Raspberry pi

The Raspberry Pi is a small, low-cost device the size of a credit card that connects to a computer monitor or TV and uses a regular keyboard and mouse. The Raspberry Pi is slower than a modern laptop or desktop computer, but it is still a fully functional Linux computer that can perform all of the functions required of it while using very little power.

System on a Chip (SoC) is a technique for putting all of the electronics needed to operate a computer on a single chip. Instead of having separate chips for the CPU, GPU, USB controller, and RAM, everything is crammed into a single, compact package. The Raspberry Pi was designed to run Linux, and several Linux distributions now have a Raspberry Pi-optimized edition.

The Raspberry Pi is a credit card-sized device that runs on the Broadcom BCM2835 system-on-a-chip (SoC). This SoC features a Video core IV GPU and a 32-bit ARM1176JZFS processor clocked at 700MHz, as well as 256MB of RAM in a POP box above the SoC. A 5V micro-USB AC adapter supplies power to the Raspberry Pi.

HDMI and composite video outputs, two USB 2.0 ports, a 10/100 Ethernet port, an SD card slot, a GPIO (General Purpose I/O Expansion Board) connector, and analogue audio output are all available on the Raspberry Pi (3.5mm headphone jack). The Raspberry Pi does not have any on-board non-volatile memory, which is typically used to store boot loaders, Linux kernels, and file systems in more conventional embedded systems, due to cost considerations.

**Methodology**

Figure 1 shows a block diagram of the steps to follow for text reading with OCR, as well as audio output.

Image Acquisition:

The image acquisition aspect creates photographs from scenes containing objects of interest. For image acquisition, a widely accessible and low-cost webcam is used.

Pre-processing and Gray Scale Conversion:

Image pre-processing techniques such as noise filtering are used to make the device more stable, allowing it to operate in noisy environments. The input is converted to Gray Scale in order to reduce the overall processing time of the process.

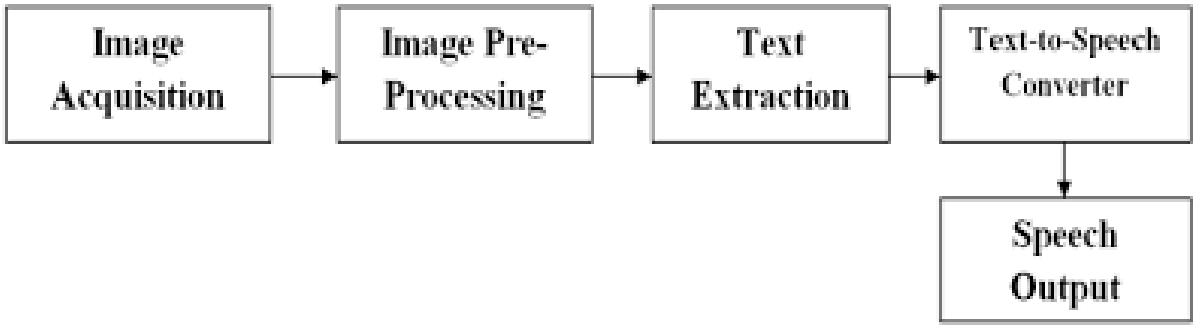


Figure1. Block diagram of project using OCR

Pre-processing of document images is a method of improving image quality by using advanced image processing techniques. Its aim is to improve and extract useful information from images for further processing. Thresholding and noise reduction are two pre-processing functions that are done here.

Edge detection:

Edge detection is a collection of mathematical methods for locating points in an image where the brightness of the image abruptly shifts or where there are discontinuities. Edges are a series of curved line segments that are commonly used to arrange such points. In the detection of edges, the Canny Edge Detection Algorithm is used.

The algorithm runs in various separate steps:

- Smoothing
- Finding gradients
- Edge tracking by hysteresis (Final edges are determined by suppressing all edges that are not connected to a very certain (strong) edge).

Thresholding:

It is the most basic method for segmenting an image from a grayscale image. To make a binary image, you can use thresholding. The Otsu method is used for thresholding. The threshold is chosen by Otsu's method by decreasing the within-class variation.

Automatic Text Extraction:

The area containing the label text is then detected using an automated text extraction algorithm. Two novel function maps are used to extract text attributes based on stroke orientations and edge distributions, respectively, to manage complex backgrounds. Automatic text extraction employs the most secure external area possible.

Optical Character Recognition:

Prior to the generation of insightful words from the localised text regions, text recognition is done using off-the-shelf OCR. A text region designates the smallest rectangular area within which characters can be accommodated, with the text region's border touching the text character's edge boundary. OCR, on the other hand, performs better when text regions are given proper margin areas and binarized to separate text characters from context. For OCR, we consider using a Template Matching Algorithm. The OCR produces a text file that contains the product label (its name) in text format. The aim of the audio output feature is to notify the blind user in the form of speech or audio of recognised text code.

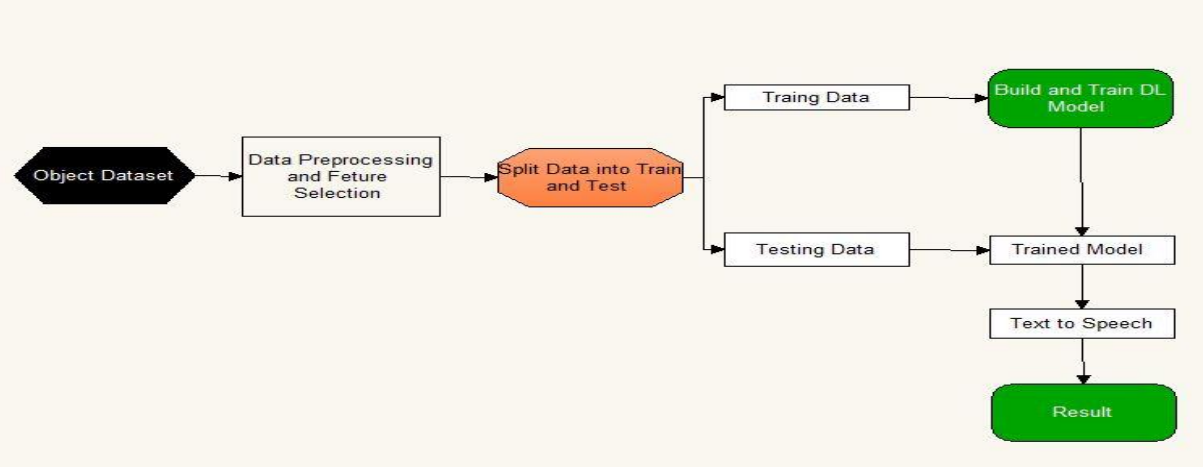


Figure2. Proposed System

Text Identification and Audio output:

The product's text is recognised and translated into audio form after the background text is extracted. Via a miniature camera fixed on spectacles, we can catch a picture of a product or some other image with text on it. In order to extract text from an image, several operations must be carried out after the image has been acquired.

First, pre-processing steps such as image filtering and noise reduction are carried out. After filtering, the image is converted to grayscale and thresholding operations are carried out. These operations are used to improve and extract valuable information from photographs for later processing.

Pre-processing measures including image filtering and noise reduction are done first. The picture is then transformed to grayscale and thresholding operations are performed. These operations are used to enhance and extract useful information from images in preparation for further processing.

**Results**

Results using OCR: It is preferable to transform the original-coloured image into a grey image in order to reduce the difficulty of performing morphological operations on the image. The conversion results are shown in Figure 2

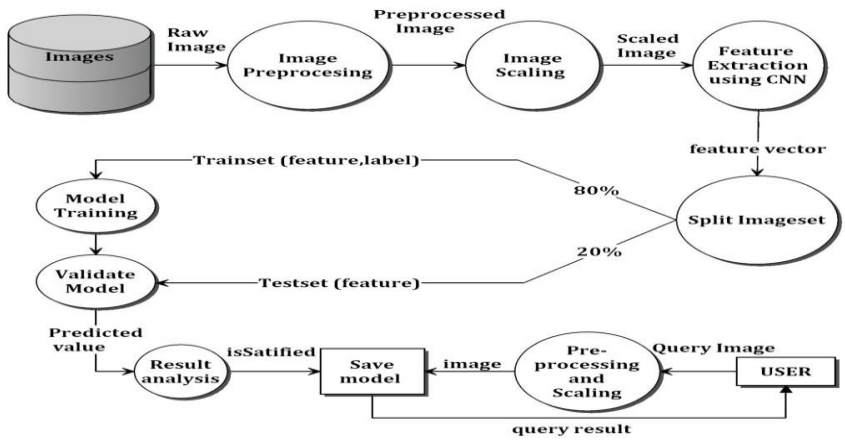


Figure 3. Dataflow Diagram

The conversion to inverted binary is completed, as shown in Figure 3.
As a result, the image's context is suppressed, and we have a plain text to work with.

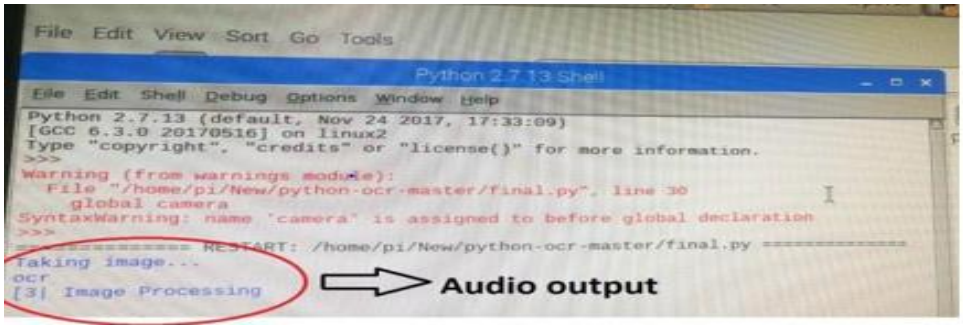


Figure 4. Audio output

The final text recognition performance is shown in Figure 4, which is highlighted on the panel. Following that, the text is processed and translated into audio output. We will hear it if we plug in our earphones.

**Conclusion:**

As a result, this paper introduces a camera-based assistive text reading framework to assist blind people in reading text labels and product packaging on hand-held items in their everyday lives. The OCR approach yielded good results for all texts with a font size of 1 inch or greater on non-complex backgrounds. Model matching may also be used to achieve effective outcomes. This approach is better suited for banknotes. Everyone, including blind people, uses currency notes, and their identification accuracy of 82%. We would use more powerful algorithms in the future so that our system can recognise small text as well as text with a variety of complex backgrounds.

**References:**

[1] B. Epshtein, E. Ofek, Y. Wexler, "Detecting text in natural scenes with stroke width transform",Proceedings of Computer. Vision PatternRecognition., pp. 2963-2970, 2010.

[2] K. Kim, K. Jung, J. Kim, "Texture-based approachfor text detection in images using support vector machines and continuously adaptive mean shift algorithm", IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1631-1639, Dec. 2008.

[3] L. Ma, C. Wang, B. Xiao, "Text detection in natural images based on multi-scale edge detection and classification", Proc. Int. Congr. Image Signal Process., vol. 4, pp. 1961-1965, 2006.

[4] X. Chen and A. L. Yuille, “Detecting and reading textin natural scenes,” in Proceedings of Computer. Vision Pattern Recognition, 2004, vol. 2, pp. II-366– II-373.

[5] M. Shi, Y. Fujisawab, T. Wakabayashia, F. Kimura, "Handwritten numeral recognition using gradient and curvature of gray scale image", Pattern Recognit., M. Shi, Y. Fujisawab, T. Wakabayashia, F. Kimura, "Handwritten numeral recognition using gradient and curvature of gray scale image", Pattern Recognit.,vol. 35, no. 10, pp. 2051-2059, 2002.

[6] J. Zhang, R. Kasturi, "Extraction of text objects in video documents: recent progress", Proc. IAPR Workshop Document Anal. Syst., pp. 5-17, 2008.