# Object Detection for Visually Impaired People Using SSD Algorithm

**Abstract:**

Visually impaired people are unaware of the danger that they are facing in their life. They may face many challenges while performing their daily activity even in their familiar environments. Vision is the necessary human senses and it plays the important role in human perception about surrounding environment. Hence, there are variety of computer vision products and services which are used in the development of new electronic aids for those blind people. In this paper we designed to provide navigation to those people. It guides the people about the object as well as provides the distance of the object. The algorithm itself calculates the distance of the object. Here it also provides the audio jack to insist them about the object. Here we are using SSD Algorithm for object detection and calculating the distance of the object by using monodepth algorithm.

<https://ieeexplore.ieee.org/document/9262344>

**Functions**: It helps in Navigation to people, it guides the people, object detection, distance from the person to object, provide audio description.

**Technologies used**: SSD Algorithm, Monodepth algorithm, computer vision, TTS

**Category-Blind Human Action Recognition: A Practical Recognition System**

**Abstract:**

Existing human action recognition systems for 3D sequences obtained from the depth camera are designed to cope with only one action category, either single-person action or two-person interaction, and are difficult to be extended to scenarios where both action categories co-exist. In this paper, we propose the category-blind human recognition method (CHARM) which can recognize a human action without making assumptions of the action category. In our CHARM approach, we represent a human action (either a single-person action or a two-person interaction) class using a co-occurrence of motion primitives. Subsequently, we classify an action instance based on matching its motion primitive co-occurrence patterns to each class representation. The matching task is formulated as maximum clique problems. We conduct extensive evaluations of CHARM using three datasets for single-person actions, two-person interactions, and their mixtures. Experimental results show that CHARM performs favorably when compared with several state-of-the-art single-person action and two-person interaction based methods without making explicit assumptions of action category.

<https://openaccess.thecvf.com/content_iccv_2015/papers/Li_Category-Blind_Human_Action_ICCV_2015_paper.pdf>

**Functions** – It does not make assumptions of actions; it subsequently classifies actions based on its motion primitive co-occurrence patterns.

**Technologies used** – Motion Unit models, Action using MSRC12-Guesture, MU combination model, Hybrid actions3D.

# IoT Enabled Automated Object Recognition for the Visually Impaired

<https://www.sciencedirect.com/science/article/pii/S2666990021000148?via%3Dihub>

## **ABSTRACT**

### **Background**

Visual impairments have become one of the most predominant problems for the last few decades. To keep doing their daily tasks, vision-impaired people usually seek help from others. An automated common object and currency recognition system can improve the safe movement and transaction activity of visually impaired people.

### **Objective**

To develop a system that can identify indoor and outdoor objects, notify the users, and send all information to a remote server repeatedly at a fixed time interval.

### **Methods**

The proposed system assists the visually impaired to recognize several objects and provides an audio message to aware the user. Four laser sensors are used in the system to detect the objects in the direction of the front, left, right and ground. The proposed system uses Single Shot Detector (SSD) model with MobileNet and Tensorflow-lite to recognize objects along with the currency note in the real-time scenario in both indoor and outdoor environments.

### **Results**

Among 375 participants, 82% reacted that the price of the proposed system is reasonable, 13% treated as the cost is moderate and the rest 5% people responded that the cost is relatively high for them. In terms of size and weight, 73% reacted that the size and weight are considerable, 20% treated that the size is not suitable, and weight needs to lessen, and the rest 7% people responded that the system is bulky. Regarding input signal observation, 98% responded that they have heard the sound appropriately and the remaining 2% of individuals missed hearing the signal.

### **Conclusions**

This paper represents an IoT-enabled automated object recognition system that simplifies the mobility problems of the visually impaired in indoor and outdoor environments. The overall accuracy of the proposed system in object detection and recognition is 99.31% and 98.43% respectively. In addition, the proposed system sends all processed data to a remote server through IoT.

# CICERONE- A Real Time Object Detection for Visually Impaired People

<https://iopscience.iop.org/article/10.1088/1757-899X/1085/1/012006/pdf>

Abstract. Our work provides a solution to people who are suffering from partial blindness due to diseases of the eye or due to accidents. Unlike people who are born blind, they depend a lot on other people for doing their day to day activities. Our project is a boon to such people as our end product; a smart walking stick detects the trained objects that are used daily by the person and intimates the person using audio messages. This project changes the visual world into an audio world by informing the visually impaired of the objects in their environment. These people can use this particular prototype for self -navigating their way. A YOLO (You look only once) algorithm is used in our project. Real-time objects in an image are detected with their names represented on a bounding box and these names are converted to speech signals. The conversion to audio signals is done by using an e-Speak tool which forms Googles Text to Speech (gTTS) system. The prototype consists of several modules. The Raspberry Pi camera module takes the image and transfers it to the Raspberry Pi desktop. Then, real-time object detection is carried out by using YOLO network. The detected image is converted to speech by using the gTTS module and the audio result is provided to the user through a headset. For achieving the required portability, the battery backup is being used.

**StakeHolders:** Partial Blind, people who lost vision in accidents, people who are suffering from eye diseases.

**Functions**: It is a smart walking stick detects the trained objects, it provides audio description

**Technologies used:** YOLO Algorithm, e-spark tool, gTTS Module

# Acoustic scene classification based on three-dimensional multi-channel feature-correlated deep learning networks

As an effective approach to perceive environments, acoustic scene classification (ASC) has received considerable attention in the past few years. Generally, ASC is deemed a challenging task due to subtle differences between various classes of environmental sounds. In this paper, we propose a novel approach to perform accurate classification based on the aggregation of spatial–temporal features extracted from a multi-branch three-dimensional (3D) convolution neural network (CNN) model. The novelties of this paper are as follows. First, we form multiple frequency-domain representations of signals by fully utilizing expert knowledge on acoustics and discrete wavelet transformations (DWT). Secondly, we propose a novel 3D CNN architecture featuring residual connections and squeeze-and-excitation attentions (3D-SE-ResNet) to effectively capture both long-term and short-term correlations inherent in environmental sounds. Thirdly, an auxiliary supervised branch based on the chromatogram of the original signal is incorporated in the proposed architecture to alleviate overfitting risks by providing supplementary information to the model. The performance of the proposed multi-input multi-feature 3D-CNN architecture is numerically evaluated on a typical large-scale dataset in the 2019 IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events (DCASE 2019) and is shown to obtain noticeable performance gains over the state-of-the-art methods in the literature.

<https://www.nature.com/articles/s41598-022-17863-z>

Function: Classifies based on the aggregation of spatial-temporal features using CNN.

**Supervised Deep Learning Techniques for Image Description:**

## **Abstract**

Automatic image description, also known as image captioning, aims to describe the elements included in an image and their relationships. This task involves two research fields: computer vision and natural language processing; thus, it has received much attention in computer science. In this review paper, we follow the Kitchenham review methodology to present the most relevant approaches to image description methodologies based on deep learning. We focused on works using convolutional neural networks (CNN) to extract the characteristics of images and recurrent neural networks (RNN) for automatic sentence generation. As a result, 53 research articles using the encoder-decoder approach were selected, focusing only on supervised learning. The main contributions of this systematic review are: (i) to describe the most relevant image description papers implementing an encoder-decoder approach from 2014 to 2022 and (ii) to determine the main architectures, datasets, and metrics that have been applied to image description.

<https://www.mdpi.com/1099-4300/25/4/553>

**Function:** It gives automatic image description using the CNN

**Technologies used**: CNN, RNN, Deep Learning

**Functionalities that we want to implement in our project**

1. Object Detection and object classification-
2. Scenery Description
3. Face Detection
4. Currency Classification
5. Audio Description of Scenery