International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-08, Issue-02, MAY 2022

Artificial Intelligence Based Ann Object Recognition and Detection

MAGESH KUMAR R, REVATHY . N.

ABSTRACT - Millions of people live in this world with disabilities in understanding the environment due to visual impairments. Although they can develop choices in dealing with daily schedules, they endure some road challenges and social awkwardness. For example, it is very difficult for them to find a specific room in a new environment. And blinding and seemingly impaired people have a hard time telling if a person is talking to them or someone else in the middle of a conversation.

Machine learning is an application of false knowledge (AI) that gives frameworks the ability to learn and progress naturally from encounter without being uniquely modified. Machine learning focuses on improving computer programs that can access and use data to learn for themselves. Artificial cognition is an approach to train a computer, robot or object to think how intelligent human thinking is. AI can provide insight into how the human brain thinks, learns, chooses and works while trying to solve problems. And finally, this reflection produces brilliant programming systems. The goal of AI is to improve computer abilities in relation to human information to illustrate, think, learn and solve problems. AI research goals are thinking, information representation, organization, learning, normal language management, performance and ability to move and control objects. Within the general knowledge segment there are long-term goals. The approaches include measurable strategies, computational insights, and conventional coding AI. Amid AI research on visual and numerical optimization, artificial neural systems and strategies based on knowledge, probability and finance, we use numerous tools and so.

Keywords – Artificial Intelligence, ANN, AI, Deection.

I. MODULE OVERVIEW

- Video streaming
- Key frame extraction
- Co- variance extraction
- Artificial Neural Network
- Object detection and recognition
- Voice synthesis

II. VIDEO STREAMING

Object detection is a classic problem in computer vision: the task of determining whether the imagery contains a specific object and general object detection approaches are seen to exploit feature extraction. The features that have received the most attention in recent years are local features. .The main idea is to focus on the areas that contain the most meaningful information. Detection of an object cannot normally be achieved with high accuracy as most applications allow detection of objects captured as an image rather than in a live video feed. The proposed system captures the image through continuous video transmission instead of capturing the image of each object every time. Object recognition and recognition has become easier when recognizing from images, but in the case of streaming video, the processing speed needs to be high to recognize the entire object in one frame. The functional process consists of capturing the images through the camera and processing them through image processing algorithms.

III. KEY FRAME EXTRACTION

The frame extraction play the very important role in several video process applications like content primarily based video retrieval, shot detection, segmentation, CC cameras , etcetera The frame conversion may be get with the seconds of the video that are get with the video. every frame will be analyzed to understand the item within the scene. The frame conversion is that the method of extracting the photographs from the video wherever the sequences of images can be delivered as frames with the given video. during this paper, a replacement technique for key frame extraction is presented. The theme uses associate degree aggregation mechanism to combine the visual options extracted from the correlation of RGB color channels, color histogram, and moments of inertia to extract key frames from the video. associate degree adaptational formula is then used to mix the results of the present iteration with those from the previous. the employment of the adaptive formula generates

 DOI : 10.35291/2454-9150.2022.0186

a swish output operate and conjointly reduces redundancy. The results are compared to a number of the opposite techniques supported objective criteria. The experimental results show that the projected technique generates summaries that are nearer to the summaries created by humans.

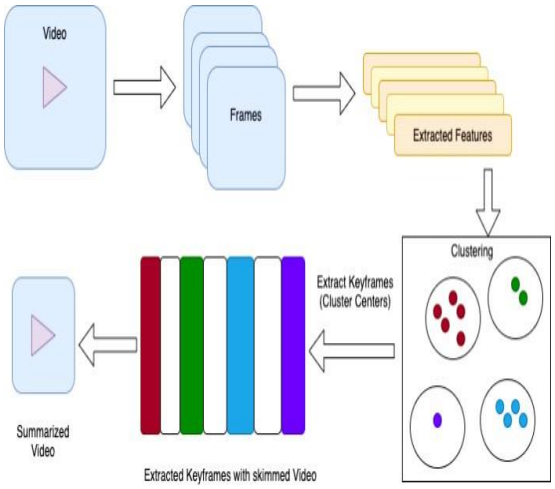


Fig 6.1 Summarized key frame extraction from video streaming

1) Uniform Sampling: Uniform sampling is one of the most common key frame extraction methods. The idea is to select every kth frame of the video, where the value of k is determined by the length of the video. A common length choice for a short video is 5% to 15% of the original video, which means you select every 20 frames at 5% or every 7 frames at 15% of the length of the upscaled video. For our experiment we decided to use all 7 frames to summarize the video. This is a very simple one Concept that retains no semantic relevance. Uniform sampling is often viewed as the generalized basis for video summarization.

2) Image histogram: Image histograms represent the overall distribution of an image. It gives us the number of pixels for specific brightness values, ranked from 0 to 256. Image histograms contain important information about images and can be used to extract key frames. We extract the histogram of all frames. Based on the difference between the histograms of two frames, we decide whether the frames have significant differences between them. We conclude that a significant dissimilarity in image histogram between frames indicates a rapid scene change in the video that may contain interesting components. When the histograms of two consecutive frames differ by 50% or more, we extract that frame as a key frame for our experiments

3) Scale Invariant Feature Transform: Scale Invariant Feature Transform (CO VARIANCE) was one of the most prominent local features used in computer vision applications for each channel, resulting in a 16X16X16 tensor. For computational reasons, a simplified version of this histogram was calculated, treating each channel separately, resulting in feature vectors for each frame belonging to R 48. The proposed nesting step for grouping is slightly different. Therefore we modify VSUMM a bit with the features extracted from VGG16 in the second fully connected layer and bundle it with kmeans.

4) ResNet16 on Image Net: When we read about the VSUMM approach, we decided to take a different approach. We chose ResNet16, trained on an image network, with different filter ranges and sliced from the last loss layer to preserve the embeddings of each image (dimension 512). We extracted frames from the videos and passed them through ResNet16, and after getting the embeds for each frame in the video, we pooled them using two algorithms: K-Means and Gaussian Mixture Models. The cluster number was assumed to be 15% of the video frame numbers. Later, we'll select the frames closest to the center of the groups as key frames. COVARIANCE EXTRACTION The main goal of using functions instead of raw pixel values as input to a learning algorithm is to reduce/increase the within/out of class variability compared to the raw input data and thus facilitate classification. The types of features that can be extracted from the image depend on the type of image, the granularity desired, and the context of the application. Once features have been extracted, their representation depends on the technique used. The feature extraction process must be precise in order to extract the same features from two images showing the same object.

5) Global image descriptor: General image properties are usually based on color indices, and the most well-known global color descriptor is the color histogram.

6) Local image descriptiors: Local peculiarities are those that have received the most attention in recent years. The main idea is to focus on the areas that contain the most discriminated information.

Semi-local image descriptor:

Most form descriptors fall into this category. This descriptor is based on the precise extraction of contours of shapes in the image or region of interest. In this case, image segmentation is often useful as a pre-processing step.

IV. ARTIFICIAL NEURAL NETWORK FOR CLASSIFICATION

The artificial neural network classification system is used to classify the points of the covariance descriptor. Feature matching using fixed features has gained considerable importance due to its application to various recognition problems. These techniques have allowed us to join images

 DOI : 10.35291/2454-9150.2022.0186

independently of the various geometric and photometric transformations between images. Machine learning is used to recognize and label objects. The artificial neural network is used for training and testing purposes. A region-based segmentation is used and then segmented to identify the object. Object detection and detection has become easier when detected from images, but in the case of streaming video. First, we take a pre-trained artificial neural network. This model is then retrained. We train the last layer of the network based on the number of classes to recognize. The third step is to get the region of interest for each image. We then reshape all of these regions to match the input size of ANN. After getting the regions, we train ANN to classify objects and backgrounds. For each class we train a binary ANN. Finally, we train a linear regression model to generate tighter bounding boxes for each object identified in the image.

Step (1) Obtain the set of key-points of objects

a. choose an oversized set of images of daily objects.

b. Extract the CO-VARIANCE feature purposes of all the photographs inside the set and procure the CO-VARIANCE descriptor for every feature point extracted from each image.

Step (2) Obtain the key points descriptor for the first video frame.

a. Extract the characteristic points of the COVARIANCE from the given image.
b. Record the COVARIANCE descriptor for each characteristic point.
c. Match the key points of the box to those of the objects and identify the objects detected.

Step (3) For the next frame:

a) If it contains the same objects, they will not be recognized.
b) New objects are recognized and identified.
c) This step uses a different method to identify similar and different frameworks for further treatment.
d) Step (4) A video file is launched for each object detected in the video to inform blind people of the identity of the objects

V. OBJECT DETECTION AND RECOGNITION

The object recognition procedure is carried out in the verification process. The region of interest is used to capture the object in the scene. Detected objects are marked with the object name. Object recognition is therefore carried out here. the process of finding instances of real-world objects such as faces, bicycles, and buildings in images or videos. Object recognition algorithms typically use function extraction and learning algorithms to recognize instances of a category of objects. It is commonly used in applications such as image retrieval, security, surveillance, and automated vehicle parking systems. Detecting a reference object (left) in a crowded scene (right) using feature extraction and matching. RANSAC is used to estimate the position of the object in the test image.

Local features and their descriptors are the building blocks of many computer vision algorithms. Its applications include image registration, object detection and classification, tracking, and motion estimation. These algorithms use local features to better handle scaling, rotation, and occlusion**.** The features we find are described as being invariant to changes in size, rotation, and position. These are pretty powerful features that are used for a variety of tasks. We use a standard Java implementation of SIFT. The SIFT descriptor is a 128-dimensional description of a pixel blob around the key point.

shows the object features extraction at each key frame

VI. VOICE SYNTHESIS

Speech synthesis is added for speech-based output systems. Here, the classified objects are known with labels, with the object's labels being output as voice output. This helps blind people recognize the object using the language process. Speech recognition essentially uses speech. Wave form analysis techniques. The speech pattern is generated per word or phrase, which is then used to recognize speech by comparing the learned pattern and the newly received pattern. This comparison process is performed using a formula related to cross-correlation. However, speech signals contain a mixture of noise components, and signals at a different frequency are also received as input. In this study, a statistical model called Hidden Markov Model (HMM) is used to accurately find the desired voice. HMM assumes that the system to be modeled is a Markov process with unknown parameters and determines which parameters are hidden from the parameters observed above. It is used to estimate the model parameters in the learning phase and then find the estimated parameters in the newly arriving speech. Google announced that its speech recognition technology is becoming more sophisticated every year by learning parameters from large amounts of data. and said that the error rate in 2017 was only 4.9%. Based on these results, the research in this study was conducted by applying Google's voice recognition service to the voice recognition function.

 DOI : 10.35291/2454-9150.2022.0186

VII. PROPOSED SYSTEM

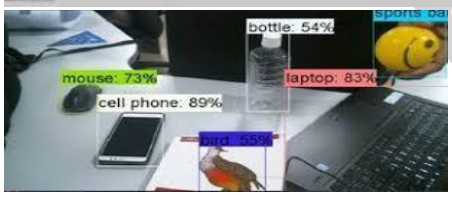
The developed system helps visually impaired people to navigate independently using real-time object recognition and identification technology. The system uses an image processing technique to recognize the object and speech synthesis to produce a voice output. The fast RANN (Region-based Artificial Neural Networks) algorithm was implemented to detect the object with high accuracy. The information from the captured image is passed on to the visually impaired people as a voice output via a speech synthesizer to help them with their mobility. Image recognition on moving objects has been an important area of computer vision research that has been worked on extensively and integrated into residential, commercial, and industrial settings.

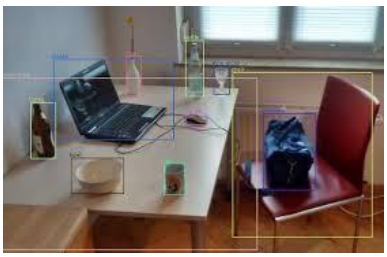
ADVANTAGES

Blind individuals will use this method for the road crossing areas and as a identification system

- They don't would like help invariably once they start
- Object labeling with voice synthesis helps the user to spot the things before them
- The accuracy is achieved with 97% wherever the object is totally detected

SCREENSHOT:





VIII. CONCLUSION

In this article, we design an object recognition system using a deep learning object recognition technique and a speech recognition technology. This system's speech synthesis provides convenient functions for the visually impaired. As one of the areas where deep learning technology can be applied, our study was conducted with a focus on how to effectively help blind people. As a result, speech recognition and voice guidance technologies were added to the system and tested for performance. This studio it can be widely used to provide blind people with privacy and convenience in daily life. In addition, it is expected to be used in industrial areas with poor visibility, such as coal mines and sea beds, to greatly assist industrial production and development in extreme environments.

SCOPE FOR FUTURE ENHANCEMENT

In future system the implementation is employed in the good telephone system wherever the blind folks will simply access victimization their smart phone system. Some limitations to the present system are that the smart phone on that this application will be used will need to be switched on and may have enough battery. it's to be carried by the user with him/her all the time. A wearable device is a lot of convenient as our hands become free in this case.

BIBLIOGRAPHY

BOOK REFERENCE:

- Qibin Hou HYPERLINK "https://orcid.org/0000-0002-8388-8708" ; Ming-Ming Cheng HYPERLINK Object Detection with Short Connections, IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 41 , Issue: 4 , April 1 2019)
- Shuze Du ; Shifeng Chen, Salient Object Detection via Random Forest, IEEE Signal Processing Letters (Volume: 21 , Issue: 1 , Jan. 2014)
- Xiaozhi Chen HYPERLINK "https://orcid.org/0000-0003-2703-5302" ; Kaustav Kundu ; Yukun Zhu ; Huimin Ma HYPERLINK "https://orcid.org/0000-0001-5383- 5667" ; Sanja Fidler ; Raquel Urtasun, 3D Object Proposals Using Stereo Imagery for Accurate Object Class Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 40 , Issue: 5 , May 1 2018)

WEB REFERENCE:

- https://ncrb.gov.in/en/crime-india
- https://www.php.net/
- https://www.w3schools.com/php/php_intro.asp

 DOI : 10.35291/2454-9150.2022.0186