Moving object detection and tracking Using Convolutional Neural Networks

|  |  |
| --- | --- |
| Santhosh K  Department of Computer Science Engineering  Sai Vidya Institute of Technology,  Bangalore, India  santhoshk.18cs@saividya.ac.in  Suhas M  Department of Computer Science Engineering  Sai Vidya Institute of Technology,  Bangalore, India  suhasm.18cs@saividya.ac.in  Tejas Manu S  Department of Computer Science Engineering  Sai Vidya Institute of Technology,  Bangalore, India  tejasmanumanus.18cs@saividya.ac.in  Vaishnavi A Kumar  Department of Computer Science Engineering  Sai Vidya Institute of Technology,  Bangalore, India  vaishnavia.18cs@saividya.ac.in | Prof. Kshama S B  Department of Computer Science Engineering  Sai Vidya Institute of Technology,  Bangalore, India  kshama.sb@saividya.ac.in |

***Abstract*— India is an agricultural country where a variety of fruits and vegetables are grown and produced for consumption as well as export. India is in the second position after China in the production of fruits. Texture, Color & freshness are the important parameters for fruit quality identification. Color recognition is a very important process in ripeness detection. The ripeness detection is an external quality factor. It is also important to check for any bacterial/fungal growth on the fruit/vegetable as due to these various factors, defective fruits can be recognized. The shelf life of fruit is also an important factor to take into consideration. Without determining the shelf-life of fruit and vegetables, there is a large amount of fruit/vegetable being wasted or rotten in the warehouse facilities and stores. Shelf-life prediction is a method that determines how long an item lasts until its “end of life”. For fresh produce, this usually means the time until an item is no longer acceptable to sell to a consumer.**

**Our goal is to develop a model where it is able to scan the fruit/vegetable in real-time and analyze it to tell if it is good or bad based on external factors such as color, freshness, texture, absence of defects and also predict the approximate shelf life of the product and display it to the user. This will be achieved using hardware such as a pi camera, Raspberry pi, LCD screen to display the output to the user. Determining the shelf life of the products will help in making a better decision on managing the fruits and storing the product that has a longer shelf life in the warehouse and transporting them to farther regions whereas products with small shelf life are shipped to local shops or sold faster thereby avoiding food wasted**

I. INTRODUCTION

According to the National, Horticulture Database India produced 99.07 million metric tonnes of fruits and 191.77 million metric tonnes of vegetables, which accounts for around 15% of the world’s vegetable production.

It's estimated that approximately 20% of produce or more gets thrown out for cosmetic reasons like weird shapes or blemishes on a peel you don't even eat. That's one in five fruits and vegetables getting tossed into landfills even though they're just as nutritious and delicious to eat. Fruits and vegetables are a vital part of the human diet to keep them healthy and free of diseases and deficiencies. Depending on their age and sex, federal guidelines recommend that adults eat at least 1½ to 2 cups per day of fruit and 2 to 3 cups per day of vegetables as part of a healthy eating pattern.

Many factors are taken into consideration during the sorting process of fruits and vegetables. These factors are internal quality factors and external quality factors. The external quality factors are texture, shape, colour, size and volume, and internal quality factors are taste, sweetness, flavour, aroma, nutrients, carbohydrates present in that fruit.

Sorting of these fruits and vegetables were being done manually by humans which required a lot of time and labour with places and chances of human error namely leaving an infected/rotting fruit with the rest of the fruits which were qualified as good, leading to the infection of the rest of the good fruits and therefore spoiling the produce. This is just one of the scenarios that can occur due to human neglect. Our model can be used to avoid such scenarios.

It also predicts the shelf-life of fruits and vegetables which is also an important factor to avoid food wastage. By determining the shelf-life of the products, we can determine which products can be stored in the warehouse for some time or be transported over long distances be it for export or storage. i.e., the products with a longer shelf-life. While the products with shorter shelf life can be transported to local stores and sold in a short period thereby avoiding wastage.

# II. LITERATURE SURVEY

There are different approaches had been presented by different researchers starting from background subtraction to CNN. Some of the human tracking methods have been presented in this section.

Human tracking consists of three basic steps for pedestrian tracking: Human detection from sequence of frame, tracking and analysis of the tracking for particular purpose. Deep learning-based low-cost machine vision system for grading the fruits based on their outer appearance or freshness. Various state-of-the-art deep learning models and stacking ensemble deep learning methods were applied to two data sets of fruits. The results of this study show that Efficient Net [CNN](https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/neural-networks) models and their stacked combinations have the highest accuracy in grading the test set and real samples as compared to the other deep learning models. [8]

A framework for learning and classifying bananas is developed first. It uses neural network technology to detect the fruit's ripening stage.

Due to the complexity of the banana fruit's ripening stages, it is necessary to develop image processing tools that can identify the various fresh incoming bunches.

The goal is to create an image processing system that can detect the different stages of the fruit's ripening process. This method would help determine the optimal eating quality and the price of bananas. [9]

Computer vision is a widely used technique for processing images. It has enormous potential in terms of image processing and farming.

In this paper, we study the various aspects of machine learning for the classification of fruits and vegetables. Through a variety of data sources, we found that SVM achieves better accuracy than other machine learning techniques.

In this paper, we perform the Recognition and classification of fruits and vegetables and also detection of disease in fruits and vegetables among the horticulture products under the agriculture field using computer vision.[11].

This paper proposed a classification model for maturity status classification of papaya fruits in two approaches, machine learning and transfer learning approach.

Overall, the VGG19 is better as VGG19 is based on transfer learning, there is no requirement of feature extraction and feature selection process. Although the transfer learning approach needs complex architecture, high training time and large datasets it is one time only.

However, the achieved accuracy in both machine learning and transfer learning is 100% and beat the previous method i.e. 94.7% of accuracy. [12] .

# A deep learning-based framework for fruit classification was proposed in this work. Two CNN models were investigated in the proposed framework, a small CNN model, and a VGG-16 fine-tuned model.

# The VGG-16 fine-tuned model achieved excellent accuracy on both datasets.

# The light CNN model also achieved excellent accuracy on dataset 1 with data augmentation

# The performance of the two models has been compared with two other methods in the literature. It was found that the two proposed models outperformed the two existing methods on dataset.[13].

# III. METHODOLOGY

The proposed CNN based moving object detection algorithm consists of two phases: Object detection and tracking. The generalized block diagram of the proposed system is shown in Fig. 1.

Input

Video

Frame

Extraction

Object detection

using Tensor

flow

Get object

Location

Object Tracking

Using CNN

Fig. 1. Block Diagramof proposed system

In this system, the video is feed to the system as an input. Frames are extracted for further processing. The two main algorithms object detection and object tracking is process through deep learning methods. The object detection is explained in detail in below flow.

The object detection using computer vision algorithm is affected by different aspects like light variation, illumination, occlusion and system has difficulty to detect the multiple objects. Hence in this paper, Tensor flow based object detection algorithm has been used.

Start

Import Necessary

Libraries

Initialize Detection Graph and load

configuration from trained model

Initialize image, box and tensor class

Read Image

Apply Tensor Flow

Extract Object Location

Object Detection

Stop

Fig. 2. TensorFlow Based Object detection flowchart

TensorFlow based object detection API is an open source platform. It is built on the top of TensorFlow which make simple to construct, train and detection models. The process of tensor flow based object detection is presented in Fig. . In this approach, firstly the necessary libraries are imported. Then import the pre-trained object detection model. The weights are initializing along with box and tensor class. After initialization of all the parameters of the tensor flow model, the image in which object to be detected is read. Apply the loaded tensor flow model on the image, the TensorFlow based model test the image and return the location (x, y, w, h) of the object in the image. This is the process of object detection of TensorFlow object detection algorithm. The success rate of this approach is better and it is applicable to RGB images.

After detecting the object, their locations are important to start the tracking process. Instead of using conventional computer vision based algorithm, in this approach Convolutional Neural Network (CNN) based tracking algorithm is used. The Flow of CNN based object tracking algorithm is as shown in Fig.3.

Load weights and

checkpoints

Get Object location and

frame number from object

detection algorithm

Pass initial bounding box to

the tracker

Pass next frame and box to

neural network

Extract New bounding Box

If last frame

Stop

Yes

No

Fig. 3. Flowchart for object detection

The object tracking is the important step in computer vision algorithms. For tracking to be robust, requires object knowledge and understanding like motion and its variation over time. Tracker must be able to its model and adopted for new observations.

In this approach first load the weights of the pre-trained model. The model is capable of incorporating the temporal information. Rather than focusing on the objects in the testing time, the pre-trained model which is trained on large variety of objects in real time. This lightweight model has ability to track the object at the speed of 150 frames per second. Also it is able to remove the remove the barrier of occlusion.

In this approach, the object locations obtained from the TensorFlow based object detection algorithms are passed to the CNN based object tracking algorithm. The initial positions are learned by the model and the same points are search in the net frames by testing process of CNN model.

# IV. RESULTS

The proposed algorithm is tested on variety of video sequences. The experimentation is divided into two parts, the object detection and tracking. The algorithm is implemented in python and tested on nine different video sequences, 3.6GHz Laptop. Without optimization the algorithm runs with good FPS.

The results of the proposed algorithm are presented in qualitative and quantitative manner. The qualitative analysis is as shown in Fig. 4 for different sequences.



Fig. 4. Qualitative analysis of proposed system of the cdv sequence.



Fig. 5. Qualitative analysis of proposed system of the mdv sequence.

The quantitative analysis is performed using sensitivity, specificity and accuracy parameter. These parameters are calculated using True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

x TP: moving object correctly identified moving object x FP: Stationary object incorrectly identified as moving object

x TN: Stationary object correctly identified as Stationary object

x FN: moving object incorrectly identified as Stationary object

The mathematical representation of the quality metrics is given as:

## A. Sensitivity

It is the ratio of truly object present in the scene who are correctly identify as an object. This term present the number of positive samples correctly identified. Higher the true positive element higher is the sensitivity.

ܶܲ

(1) = ݕݐ݅ݒ݅ݐ݅ݏܵ݁݊

ܰܨ+ܲܶ

## B. Specificity

It is the ratio of truly stationary object present in the scene that are correctly identify as a stationary object. This term present the number of negative samples correctly identified. Higher the true positive element higher is the Specificity.

ܶܰ

(2) = ݕݐ݅ݒ݅ݐ݅ݏܵ݁݊

ܰܨ+ܰܶ

## C. Accuracy

Accuracy is the overall performance of the system including sensitivity and specificity.

ܶܲ+ܶܰ

(3) = ݕܿܽݎݑܿܿܣ

ܰܨ+ܲܨ+ܰܶ+ܲܶ

TABLE I. QUALITATIVE ANALYSIS OF THE PROPOSED SYSTEM

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Database video sequences** | | **No. of frames** | **Sensitivity** | **Specificity** | **Accuracy** |
| Brv | | 1201 | 0.9885387 | 0.9837067 | 0.9775 |
| Cdv | | 2030 | 0.9597742 | 0.9018933 | 0.928079 |
| Cpv | | 239 | 0.8874172 | 0.797619 | 0.841004 |
| Gfv |  | 1482 | 0.9730539 | 0.9766537 | 0.946694 |
| Mdv |  | 2400 | 0.9160305 | 0.8788789 | 0.915833 |
| Mev |  | 551 | 0.8856089 | 0.8931034 | 0.905626 |
| Pev |  | 820 | 0.9467593 | 0.9331395 | 0.88916 |
| Psv |  | 2938 | 0.8630665 | 0.9254984 | 0.905037 |
| T |  | 386 | 0.8732394 | 0.9217391 | 0.870466 |
|  | **Average** | | 0.9214987 | 0.9124702 | 0.908822 |

# V. CONCLUSION

In this paper, novel approach for object detection and tracking has been presented using convolutional neural network. The moving object detection is performed using TensorFlow object detection API. The object detection module robustly detects the object. The detected object is tracked using CNN algorithm. Considering human tracking as a special case of detection of objects, spatial and temporal classes the facilities were learned during offline training. The shift variant architecture has extended the use of conventional CNNs and combined the global features and local characteristics in a natural way. The proposed approach achieves the sensitivity of 92.14%, specificity of 91.24% and accuracy of 90.88%.

# ***References***

1. A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, “Background and foreground modeling using nonparametric kernel density estimation for visual surveillance,” Proc. IEEE, vol. 90, no. 7, pp. 1151– 1163, 2002.
2. Enrique J. Fernandez-Sanchez \*, Javier Diaz and Eduardo Ros, “Background Subtraction Based on Color and Depth Using Active Sensors”, Sensors 2013, 13, 8895-8915; doi:10.3390/s130708895.
3. Prajakta A Patil, Prachi A Deshpande, “Moving Object Extraction Based on Background Reconstruction”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 4, April 2015
4. Jialue Fan, Wei Xu, Ying Wu and Yihong Gong,” Human tracking using Convolutional Neural Network,” in Proc. IEEE transactions on neural networks, vol.21, NO.10, Oct 2010
5. Junda Zhu, Yuanwei Lao, and Yuan F. Zheng, “Object tracking in structured environment for video surveillance applications”, IEEE transactions on circuits and systems for video technology, vol.20, February 2010.
6. D. Koller, J. Weber and J. Malik, “Robust multiple car tracking with occlusion reasoning,” Proc. Third European Conference on Computer Vision, 1994, pp. 189-196, May 2-6 1994
7. Logesh Vasu and Damon M. Chandler, “Vehicle tracking using a Human-Vision-based Model of Visual Similarity” IEEE 2010.
8. Chen, Y, X. Yang, B. Zhong, S. Pan, D. Chen, and H. Zhang, “Cnn tracker: Online discriminative object tracking via deep convolutional neural network”. Applied Soft Computing, 2016.
9. Feris, R., A. Datta, S. Pankanti, and M. T. Sun, “Boosting object detection performance in crowded surveillance videos”. In IEEE Workshop on Applications of Computer Vision, pp. 427-432, 2013.
10. Krizhevsky, A., I. Sutskever, and G. E. Hinton, “Imagenet classi\_cation with deep convolutional neural networks”, 2012
11. Ji, S., W. Xu, M. Yang, and K. Yu, “convolutional neural networks for human action recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence 35(1), 221-231, 2013.
12. Fan, J., W. Xu, Y. Wu, and Y. Gong, “Human tracking using convolutional neural networks”. IEEE Transactions on Neural Networks 21(10), 1610-1623, 2010.
13. Hong, S., T. You, S. Kwak, and B. Han, “Online tracking by learning discriminative saliency map with convolutional neural network”, arXiv preprint arXiv:1502.06796 .
14. Wang, N., S. Li, A. Gupta, and D. Yeung, “Transferring rich feature hierarchies for robust visual tracking”, Computing Research Repository abs/1501.04587, 2015.

1. Wang N., S. Li, A. Gupta, and D. Yeung, “Transferring rich feature hierarchies for robust visual tracking”. Computing Research Repository 2015, abs/1501.04587.