JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Real-time Scene Change Detection with Object Detection for Automated Stock Verification

Santhosh Kumar P C[1], Selbi M U[2]

Technical Officer, O/o Controller of Technical Examination, Kaimanam, Trivandrum, Kerala, India.[1]

Lecturer, Dept. of Computer Engineering, Govt. Polytechnic College, Kalamassery, Ernakulam, Kerala, India.[2]

Abstract : Automation is the process of reducing human effort via the use of technology. The suggested method will contribute to a reduction in the amount of staff needed at a supermarket by continuously tracking product availability and automatically informing the appropriate party of any relevant information. A scene change detection technique based on the Structural Similarity Index Measure (SSIM) allows for the detection of human presence. The number of products in a given product rack will be counted using an object detection algorithm. In this scheme, the products in the product racks are identified using the well-known object detection and recognition scheme known as YOLO. A brief notice is displayed if fewer than a certain amount of items are present in a specific rack. A brief message and/or email will be sent to the relevant authority if the quantity of goods on a particular rack falls below a predetermined level. A printed product identifier will be kept close to the product racks to make it more comfortable. The product identifier will be identified by an OCR module in the suggested system, aiding the supervisor in planning when to replace the products in the racks.

Index Terms **- Automation, Scene change detection, Object Recognition, You Only Look Once, Frame Extraction.**

I. INTRODUCTION

Today, automation techniques are fully utilized to increase productivity in both the manufacturing and service sectors. In many instances, computer vision techniques are used to simulate the functions of human eyesight. The computer vision technique is a way of getting some valuable data out of pictures or movies. The two main components of the computer vision task are object detection and object recognition. The method of object detection involves locating the areas of an image where some significant objects are present. The task of categorizing the identified items into certain classifications is known as object recognition.

This system uses computer vision to track the availability of the products in the store. A few cameras must be mounted for this reason in order to record videos of the product racks in real-time. Every time the number of products offered drops below a predetermined threshold, it can be seen on the real-time video, and the affected person will be informed. The suggested methodology is unique in that it uses a scene change detection method based on structural similarity (SSIM) in order to reduce the number of frames needed for automatic stock verification. The method for detecting the items in the product racks uses the well-known YOLO (You Only Look Once) object detection and recognition algorithm.

Scene Change Detection, Object Detection, and Optical Character Recognition are the main concepts explored. A few communication modules are also explored; these may be paid for by private companies, but a trial version can be accessed and used for this project. A traditional mail approach is also investigated, based on data from the code or integers from the code that are converted into strings and attached to warning mail for use in alerting purposes.

II. RELATED WORK

The Fast Region-based Convolutional Network technique (Fast R-CNN)[2] for object detection is suggested in this research. Fast R-CNN uses a number of advances to increase detection accuracy while speeding up training and testing. In Fast R-CNN, we provide the CNN the input image, and it produces convolution feature maps. The regions of suggestions are extracted using these maps. We next restructure all of the suggested regions into a fixed size using a RoI (Regions of Interest) pooling layer so that it may be fed into a fully linked network.

Modern object detection networks rely on region proposal techniques to make location assumptions for objects. In this study, we introduce the Region Proposal Network (RPN), which collaborates with the detection network to share full-image convolutional features and enable almost cost-free region suggestions. To extract all the items, the program needs to run through a single image numerous times. The performance of the systems that come after it depends on how the earlier systems performed because they operate one after the other.

A quick YOLO can process 155 frames per second and construct bounding boxes based on confidence and support calculations, making it one of the best advancements for fast processing speed when compared to other detection methods. Although YOLO has a lower propensity to predict false positives on background, it generates more localization errors. Since the entire detection pipeline consists of a single network, detection performance can be tuned from beginning to end.

Building systems for monitoring humans and comprehending their appearance, activities, and behavior, as well as developing sophisticated user interfaces for interacting with humans, have become increasingly the focus of recent computer vision research. Recognizing motion of objects in the two provided photos is the aim of motion detection. Additionally, recognizing motion in things can help with object recognition.[8].

This uses fully convolutional, region-based networks to recognise objects accurately and quickly. Our region-based detector is completely convolutional, with practically all computation shared across the entire image, in contrast to earlier region-based detectors like Fast/Faster R-CNN that apply an expensive per-region subnet work hundreds of times.[1].

III. PROPOSED SYSTEM ARCHITECTURE

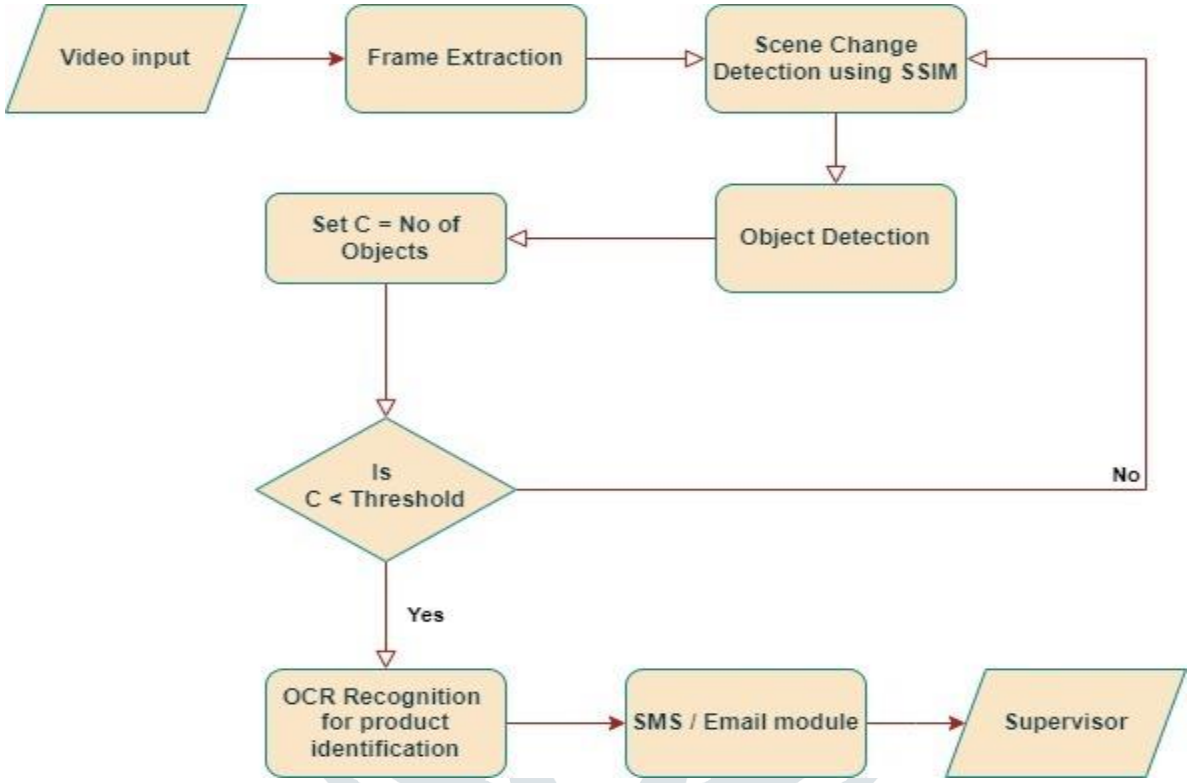


Figure 1 : Working of Proposed system

Figure1 depicts the proposed scheme's overall layout. In this plan, we need to install video cameras to keep an eye on product racks in a store for security. The video frames that were taken need to be processed in a predetermined manner. The algorithm and system's operation are explained below as the order of steps in a step-by-step process.

ALGORITHM

1. Capture the real-time video, V, of the product rack. The video can be considered as sequence of N frames, V=(f1,f2,f3,...,fN).

2. Rf (Reference frame) = f1 (The first frame will be considered as the reference frame for scene change detection).

3. p=30×FR,(Here FR is the frame rate of the camera; we are Considering frames in an interval of 30seconds).

4. While (fp not equal to NULL).

5. Find the structural similarity index (SSIM) between fp and Rf and keep it in FD (frame difference).

6. If FD < T1//T1 is a threshold value to determine a frame change (presence of human).

7. Applyobjectdetectionandrecognitiononfpandidentifythenumberofobjects,C,present in the image.

8. If C<T2 (here,T2 is a threshold value and the concerned person should get a notification if the number of product available in the

rack goes below T2).

9. Apply optical character recognition to know the unique product identifier, Pid, that will be present in the bottom of the product rack.

10. Send an email and/or SMS to the responsible person to inform him that the product with product identifier, Pid, does not have enough stock.

IV. WORKING

This system introduces a computer vision-based method for automatically monitoring the products. The device records videos of the product racks in the supermarket to continuously monitor their availability. The following step is to extract each frame from the recorded video, designate the first frame as the reference frame, and then compare the other frames to the reference frame to detect scene changes. To verify that things are below the threshold limit, an object detection approach is employed. The system's optical character recognition (OCR) module helps the supervisor schedule the replacement of the items in the racks by identifying the product.

4.1. Frame Extraction

The real-time video, V, that the security camera recorded can be viewed as a series of N frames. It is possible to view each frame of the movie as an RGB-formatted color image. The frames will be moved to the following step at regular intervals.

4.2. Scene Change Detection

Scene change detection comes next in the suggested method. A scene change will be seen anytime a product is removed by customers because the camera is always recording and we are reviewing the frames often (15 seconds between each picture). The overhead of item detection and recognition jobs is reduced because to the scene change detection. It can be done in real-time because we employed SSIM-based scene change detection technology.

A perception-based approach called the structural similarity index (SSIM) takes image deterioration into account as a perceived change in the structural information. Following are the steps for computing the SSIM between two photos, I and G:

Theoretical SSIM Formulae

$$SSIM = \frac{(2\mu_I\mu_G + C_1)(2\sigma_I\sigma_G + C_2)}{(\mu_I^2 + \mu_G^2 + C_1)(\sigma_I^2 + \sigma_G^2 + C_2)}$$

4.3. Object Detection

The YOLO Algorithm is used in our suggested system to detect objects. YOLO is a real-time object recognition technique that makes it easy to quickly identify the things in a frame. A quick version of YOLO can process 155 frames per second, producing a picture with bounding boxes surrounding the objects.It is a smart Convolutional Neural Network (CNN) and one of the most effective real-time object detection algorithms. Because it only requires the image (or video) to transit once through its network, it is named that way as opposed to earlier object detector algorithms like Region-based algorithms (R-CNN or its update Fast R-CNN). These antiquated techniques looked at various areas of the image one by one to discover the objects that were there. By applying reasoning at the level of the big picture, YOLO transformed that.

4.3.1. OVERVIEW OF YOLO

The program examines each grid individually and labels any labels that include objects along with their bounding boxes. The labels of the empty grid are denoted as zero. The label is taken to be Y. Y has 8 possible values.

- **Pc –** Represents whether an object is present in the grid or not. If present **Pc=1**, else**0**.
- **bx, by, bh, bw –** are the bounding box values of the objects(if present).
- **C1,C2,C3 –**are the classes. Let's say the classes are human, car and chair respectively. If the object is a car then **C1** and **C3** will be 0 and **C2** will be 1. Elements of label **Y.**
- If two or more grids contain the same object then the center point of the object is found and the grid which as that point is taken.
- For this, to get the accurate detection of the object we can use two methods. They are Intersection over Union and Non-Max Suppression.
- In IoU, it will takes the actual and predicted bounding box value and calculates the IoU of two boxes by using the formulae,

IoU = Area of Intersection (overlap) / Area of Union.

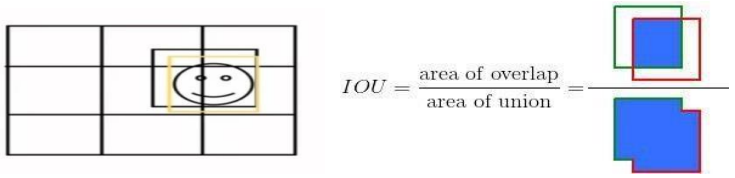


Figure 2 : IoU calculation

- If the value of IoU is more than or equal to our threshold value (0.5) then it's a good prediction. The threshold value is just an assuming value. We can also take greater threshold value to increase the accuracy or for better prediction of the object.
- The other method is Non-max suppression, in this, the high probability boxes are taken and the boxes with high IoU are suppressed. Repeat this until a box is selected and consider that as the bounding box for that object.

4.3.2. YOLO ARCHITECTURE

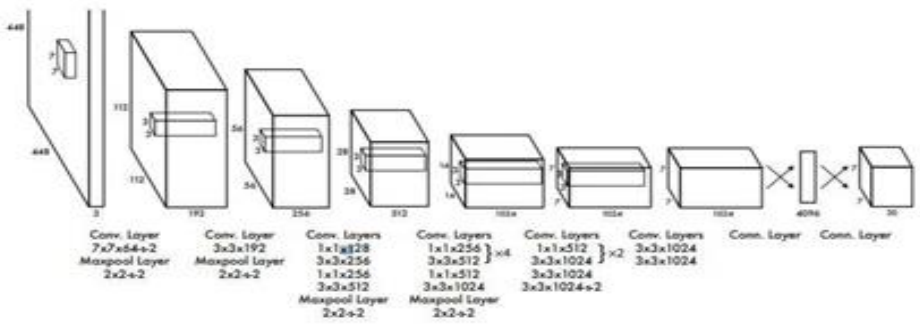


Figure 3 : Yolo Architecture

4.4. Optical Character Recognition

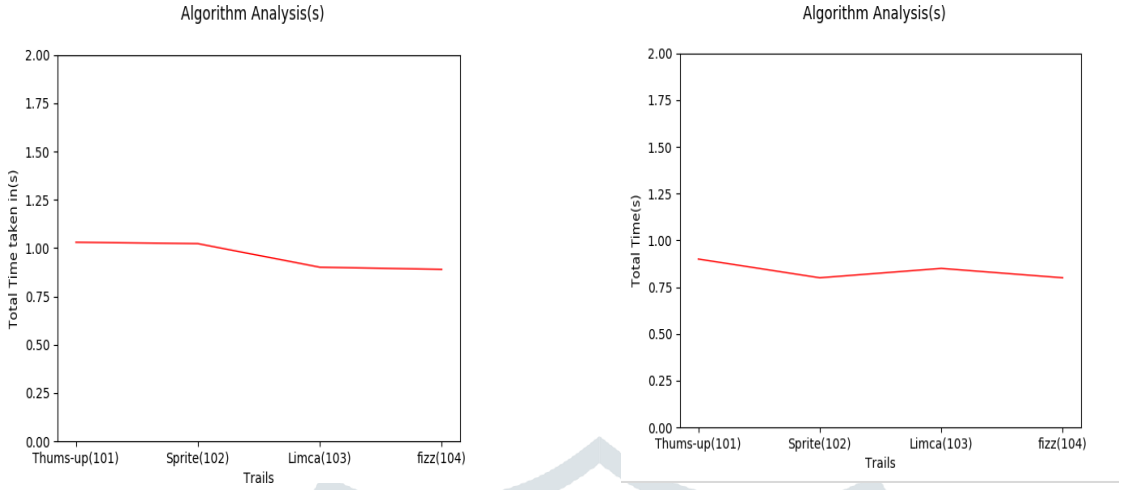
- An OCR scheme is used in the proposed scheme to identify the product name or unique product identifier mention on the product rack.
- We assumed that the product identifier will be kept below the product in the product rack.
- And in our proposed system, we used the Python-Tesseract (pytesseract) tool for Optical Character Recognition (OCR) phase.
- Python-Tesseract or pytesseract is an OCR tool for python. It will read and recognize the text in images, license plates, etc.. (Any image with text).

4.5. Communication Module

- In this phase, the number of products in the selected rack will be counted and if it goes beyond the threshold limit an SMS or an email will be triggered to inform the supervisor.
- The product identifier or product name also will be included in the SMS or email that will help the supervisor to plan the refilling process.
- Python already has a library that lets us connect to an SMTP server. This library is called smtplib and comes included with python. We used ‘smtlib’ to send E-mails.
- SMTP (Simple Mail Transfer Protocol) is an application-level protocol (on top of TCP) used to communicate with mail servers from external services, like an email client on your phone.
- We used Sinch’s API for SMS. As a global CPaas (Communications Platform as a Service provider), Sinch enables businesses to send text messages to large numbers of customers. By providing Application Program Interfaces (APIs),making it easy for developers to integrate their apps to Sinch, a business client can send out text messages to customers.

V. PERFORMANCE EVALUATION

CPU Vs GPU Performance



Processing time on CPU *Processing time on GPU*

Figure 4 : CPU Vs GPU Performance

The graphs show that the system performs at nearly one second and less than one second when using a CPU, and almost the same while using a GPU, where its slightly less than one at around 0.8 to 0.9 seconds. Therefore, based on this information, we can say that the suggested system is scalable and trustworthy even when using a CPU.

VI. CONCLUSION AND FUTURE WORK

The creation and introduction of a computer vision-based solution for automated product monitoring in the supermarket. The suggested framework is adaptable to crowded stores. The management businesses of supermarkets will benefit from the execution of the suggested strategy by having less labor needed, more profits, and improved consumer happiness. To avoid analyzing all of the video frames in the suggested technique, we employed a new scene change detection algorithm based on the structural similarity index, which facilitates real-time optimization. In this project, we utilized a single camera to monitor the products in a single rack, but in a real situation, a single camera might be used to monitor a sizable area that has multiple racks containing various products. This system's design allows for further development and can be used to capture many items with distinctive identification.

The scope of our suggested plan is superior in the long run for lowering the amount of labor needed in a busy supermarket or hypermarket for stock checking and alert systems. Future upgrades to the system could include the addition of new functions.i.e., the proposed solution can also be integrated with the current billing system. A database can be used to implement an android application that can show the current stock status, send out alert messages when the stock gets low, and also add product details like price and expiration date so that the system can be changed to send alerts when products are about to expire. Therefore, refilling is simple and quick, which will enhance the customer's shopping experience. The biggest benefit is that we can easily build, maintain, and expand upon this system with new features. Therefore, it has a lot of potential for the automation of supermarkets in the future.

REFERENCES

[1] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-fcn: Object detection via region-based fully convolutional networks. Advances in neural information processingsystems,2016.

[2] Ross Girshick. Fast r-cnn. Proceedings of the IEEE international con-ference oncomputervision,2015.

[3] Venugopal Gundimeda, Ratan S Murali, Rajkumar Joseph, and NT NareshBabu. An automated computer vision system for extraction of retail food product metadata. First International Conference on Artificial Intelligence and Cognitive Computing, 2019.

[4] Wei-Ying Ma and Bangalore S Manjunath. Edge ow: a technique for boundary detection and image segmentation, IEEE transactions on image processing, 2000.

[5] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only lookonce:Unified,real-time objectdetection.IEEE,2016.

[6] SayantaniSaha and SarmisthaNeogy. A case study on smart surveillance applicationsystem using wsn and ip webcam. In 2014 Applications and Innovations in MobileComputing (AIMoC),pages36{41.IEEE,2014.

[7] NishuSingla. Motion detection based on frame difference method. InternationalJournal ofInformation &ComputationTechnology,2014.

[8] Srikrishna Srivastava Varadarajan, MuktabhMayank. Weakly supervised object localization on grocery shelves using simple fcn and syn-thetic dataset. arXivpreprint arXiv:1803.06813,2018.