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

Sandeep Kumar Yedla
Department of CSE
Indian Institute of Information Technology
Kottayam, India.
Email: sandeep16@iiitkottayam.ac.in

V. M. Manikandan
Department of CSE
SRM University AP,
Andhra Pradesh, India.
Email: manikandan.v@srmap.edu.in

Panchami V.
Department of CSE
Indian Institute of Information Technology
Kottayam, India.
Email: panchamam036@iiitkottayam.ac.in

2020 5th International Conference on Devices, Circuits and Systems (ICDCS) 978-1-7281-6368-0/20/$31.00 ©2020 IEEE 10.1109/ICDCS48716.2020.243571

***Abstract*—Automation is a process of utilizing technology to reduce human efforts. In this paper, we propose a computer vision-based automated system for stock management in the supermarkets. The proposed system will help to reduce the man-power required in a supermarket by continuously monitoring the availability product in the supermarket and reporting the useful information to the concerned person automatically. The key idea behind the proposed scheme is that a few low-cost cameras will be placed in the supermarket which will help to capture the videos of the product racks in the supermarket. The presence of human beings are identified b y u sing a s tructural similarity index (SSIM) based scene change detection technique, further, an object detection technique will be used to count the number of items present in the specific p roduct r ack. I f t he number of items present in a particular rack goes below a threshold limit, a short message service (SMS) and/or email will go the concerned authority. To make it more comfortable, a product identifier (printed) w ill b e k ept j ust b elow t he p roduct racks. An optical character recognition module in the proposed scheme will identify the product identifier a nd i t w ill b e m entioned in the SMS or email which will help the supervisor for scheduling the replacement of the items in the racks. The experimental study is carried out by placing sample items on a rack and the mobile camera is used as an IP camera with the help IP webcam android application for the monitoring purpose. The experimental study shows that the proposed scheme will work reliably in a supermarket environment.**

***Index Terms*—Automation, Object detection, Stock management, Scene change detection;**

I. INTRODUCTION

Nowadays automation techniques are exploiting in its full potential to improve the productivity in manufacturing industries and service sectors. Computer vision techniques are utilized in many places to mimic the behaviour of human vision. The computer vision is a process extracting some useful information from images or videos. In general, the computer vision task involves two activities: object detection and object recognition. The object detection is a process identifying the regions in the image where some meaningful objects are present. The object recognition is the task of classifying the detected objects into certain classes.

In this paper, we use computer vision approach for monitoring the status of the products available in the supermarket. For this purpose, a few cameras need to be placed to capture the real-time videos of the product racks. From the real-time video, whenever the available number of products are going below a threshold limit, it will be identified and that information will be shared to the concerned person.

The novelty of the proposed scheme is that we have used a structural similarity (SSIM) based scene change detection technique and it is used to optimize the number frames required to process for automatic stock verification. The well-known object detection and recognition scheme called YOLO (You Only Look Once) is used in the proposed scheme for identifying the products in the product racks.

The major concepts explored in this paper are scene change detection, object detection, and optical character recognition. So for better understanding, the manuscript is arranged as follows: the related work is discussed in section II, the proposed scheme is discussed in section III. The experimental setup and sample results are discussed in section IV, and finally, section V concludes the manuscript with the possible extensions on the work.

II. RELATED WORK

The proposed scheme uses scene change detection and object detection techniques. This section mainly discusses the well-known scene change detection schemes and object detection schemes available in the literature.

A fast and simple scene change detection algorithm based on the correlation between the frames of the video is discussed in [1]. In this scheme, the first frame is taken as the reference frame and further, the correlation between histograms of the upcoming frames are used as the criteria to detect the scene change. Another histogram-based scene change detection technique is discussed in [1]. The histograms of the blocks from the frames are considered for finding the scene changes. In this scheme, Otsu thresholding technique is used for finding the threshold value to detect the scene change. An entropy-based scene change detection is discussed in [2]. A change detection algorithm based on the structural information is discussed in [3]. In the proposed scheme, we used the SSIM measure between the reference frame and the current frame of the real-time video. The threshold values to identify the scene change is empirically identified as 0.6.

Object detection is an active research area widely used for content-based image retrieval, robot navigation, and product quality verification. A detailed survey on object detection is discussed in [4]. A Hough transform based object detection scheme is introduced [5]. This scheme is mainly introduced for finding circular objects in a given image. Various machine learning approaches are exploited for object detection and object recognition [6]–[8]. A well-known, real-time object detection scheme is introduced in [9] which is known as YOLO (You Only Look Once). Due to high detection accuracy and real-time performance, YOLO is widely accepted and in our proposed scheme also we have used the same.

III. Proposed Scheme

The overview of the proposed scheme is shown in Fig. 1. In the proposed scheme, we need to place the video cameras to monitor product rakes in a shop/supermarket. The captured video frames need to be processed in a pre-defined way.

The sequence of activities in the proposed scheme is given in Algorithm I.

Algorithm I :Computer-vision based supermarket management

1) Capture the real-time video, V, of the product rack. The video can be considered as sequence of N frames, $V = (f_1, f_2, f_3, ..., f_N)$
2) $R_f = f_1$ `//The first frame will be considered as the reference frame for scene change detection`
3) $p = 30 \times F_R$ F_R `is the frame rate of the camera; we are Considering frames in an interval of 30 seconds`
4) While $(f_p \neq NULL)$ `//Need to process all the frames from the given video stream`
5) Find the structural similarity index (SSIM) between f_p and R_f and keep it in F_D. The way to compute SSIM is given Equation (1).
6) If $F_D < T_1$ `//`T_1 `is a threshold value to determine a frame change (presence of human). We empirically identified,` $T_1 = 0.60$
7) Apply object detection and recognition on f_p and identify the number of objects, C, present in the image.
8) if $C < T_2$ `//`T_2 `is a threshold value and the concerned person should get a notification if the number of product available in the rack goes below` T_2
9) Apply optical character recognition to know the unique product identifier, P_{id}, 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, P_{id}, does not have enough stock.
11) $R_f = f_p$
12) $p = 30 \times F_R$ `Considering frames in an interval of 30 seconds`
13) EndWhile

The proposed computer vision based system will capture the video frames throughout the day. The whole proposed framework consists of the following modules:

- Frame extraction
- Scene change identification
- object detection
- Optical character recognition (OCR)
- Communication module

A. Frame Extraction

The real-time video, V, captured by the surveillance camera can be considered as a sequence of N frames and it can be denoted as $V = (f_1, f_2, ..., f_N)$. Each frame in the video can be treated as a colour image in RGB format, where RGB represents red, green and blue colour components. The frames in a frequent interval will be transferred to the next phase.

B. Scene Change Detection

The next phase in the proposed scheme is scene change detection. The key idea is that the further steps like object detection, counting the products in the rack, OCR and automated emailing or SMS sending process can be carried out only after scene change detection. Since the camera is working all the time and we are considering the frames in a frequent interval (every frame in an interval of 30 seconds), whenever the product is taking out by costumers then a scene change will be detected. No need to do all the phases in every frame. The scene change detection helps to optimize the overhead of object detection and recognition tasks. Since we have used SSIM based scene change detection technique, it can be carried out in real-time.

Absolute frame difference is one of the most common approach to detect the scene change or frame change in a video sequence. The absolute frame difference between two images I and G having size of $R \times C$ pixels can be computed as follows:

$$A_{FD} = \sum_{x=1}^{x=R} \sum_{y=1}^{y=C} \left| I_{x,y} - G_{x,y} \right| \quad (1)$$

Based on the situations, we need to fix a threshold value for A_{FD} and if it goes below the threshold value, we can identify it as a scene change. The absolute frame difference approach is tried in the proposed scheme to detect the scene changes, but it was not giving a reliable result due to the lighting changes. This motivated us to use a new scheme for scene change detection which uses SSIM measure.

For identifying the scene change, we have used the SSIM measure between the reference frame and the current frame [10]. The SSIM is an image quality assessment technique to find the structural similarity between two images. We used the same measure to identify the frame changes in the proposed scheme. The structural similarity index (SSIM) is a perception-based model that considers image degradation as a perceived change in the structural information [**?**]. The SSIM between two images I and G can be computed as follows:

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

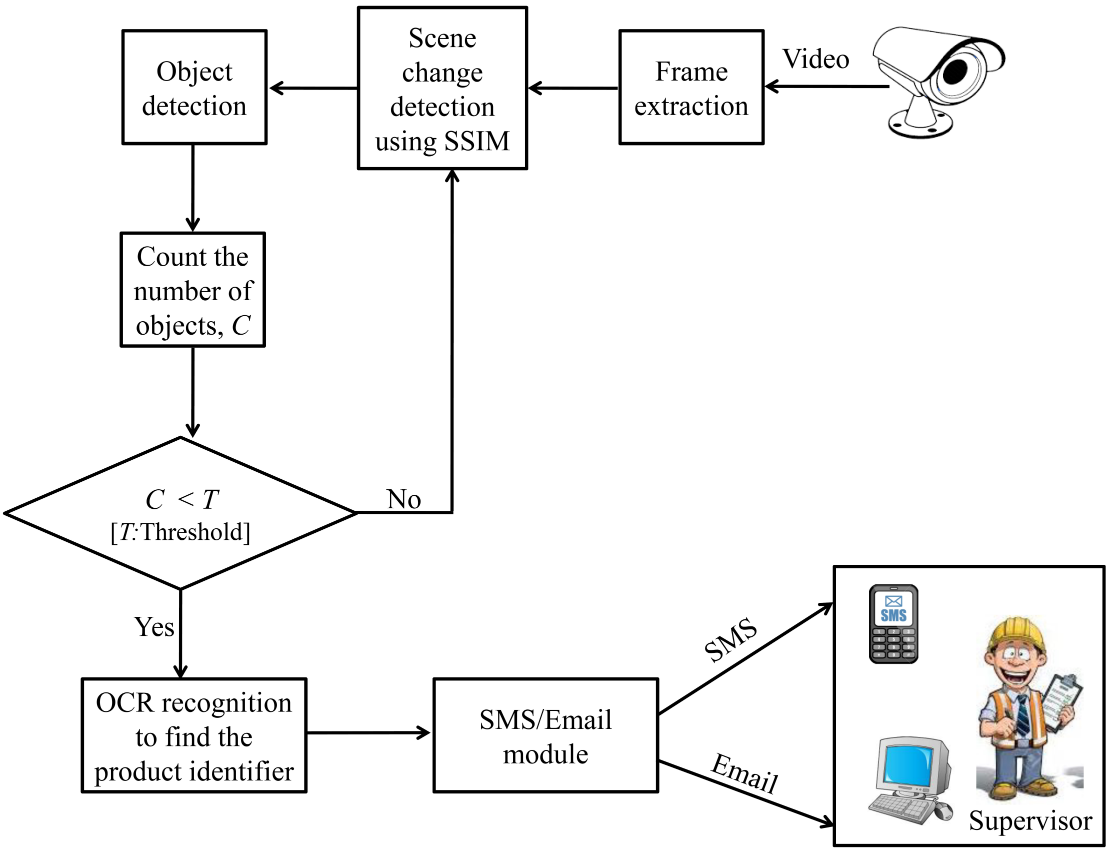

Fig. 1. Overview of the proposed scheme

where μ_I and μ_G are the means, σ_I and σ_G are the standard deviations of I and G respectively, σ_{IG} is the cross-covariance between the images I and G. It may be noted that the variables C_1 and C_2 are used to stabilize the division operation with weak denominator. Generally, $C_1 = (k_1 D)^2, C_2 = (k_2 D)^2$, where D is the dynamic range of the pixel values in the image I and G, by default k_1 and k_2 are considered as 0.01 and 0.03 respectively. In an 8-bit grayscale image, dynamic range D will be 256.

The SSIM value may vary between 0 to 1. If the SSIM value between two images are 1, then it indicates that both the images are almost the same. The low SSIM values between two images declare that both the images do not have the same structural properties. In the proposed scheme, we are finding the SSIM value between reference frame, R_f, and the current frame, f_p. If the SSIM value between R_f and f_p is less than 0.6 then we are considering that in f_p there is a scene change has happened. The scene change may happen due to various reasons: a person may walk between the product rack and the camera or a person is taking out a product from the rack. In all such cases, a scene change will be identified by the proposed scheme. The overview of the SSIM based scene change detection is shown in the Fig 2.

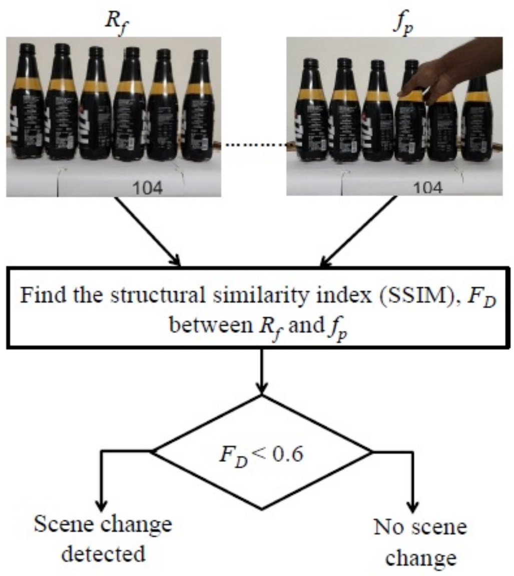

Fig. 2. The overview of scene change detection technique using SSIM

C. Object Detection

The object detection is carried out using YOLO. The YOLO is a real-time object detection scheme which helps to identify

Product 1 Product 2

Product 3 Product 4

Fig. 3. Initial arrangement of products in product rack

Product 1 Product 2

Fig. 4. Scene change detection identified while taking the products from the product rack

the objects in a frame very fastly. A fast YOLO can process 155 frames per second and the result will be an image which contains bounding boxes around the objects.

D. 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 will be kept below the product in the product rack. Based on this assumption, we filtered other text information which may be present in the product itself like company name or other details manufacturing date, expiry date etc.

E. 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.

IV. Experimental Study and Result Analysis

We simulated the supermarket environment in the laboratory by keeping a few sets of items (soft drink bottles) with printed product identifier on the bottom. The algorithms were implemented using Python. For capturing the real-time video, we have used Redmi Note 3 mobile camera (16 Mega Pixels) and the captured frame size is 640×480 pixels. A few sample products arranged as shown in Fig. 3 during the experimental study. Initially, we kept six products in the product rack and we removed products one by one and checked whether the SMS and/or email are going to the given mobile number and/or email when the number of products in the rack goes below the threshold. We fixed 3 (half of the number of items) as the threshold value during the experimental study. The experimental set is done in different lighting conditions since it is very much important in a practical situation.

We mimicked the actual situation that may happen in the supermarket by taking the bottles randomly. While a person is trying to take the bottles, obviously the scene change detector will identify a major frame change. Two frames that have been identified as a scene change are shown in Fig. 4. Note that due to the presence of the hand in the particular frame, the SSIM measure will vary while comparing with the reference frame.

After identifying the scene changes, from the upcoming frames, we will identify the objects in the frames using YOLO. If the number of objects in the frames goes below half of the products in the reference frames, then a message is triggered to the supervisor. The objects detected by YOLO is shown in Fig. 5.

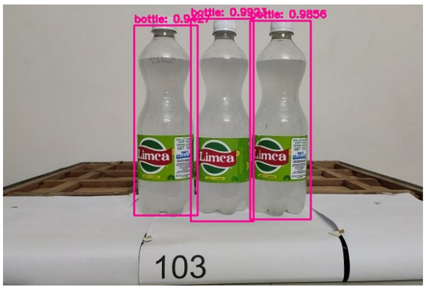

Fig. 5. The objects detected by YOLO object detection technique

Once the system identifies that the number of products in the product rack is less than it will send an SMS and/or email to the respective supervisor. Before sending the warning message to the supervisor, an OCR module will detect the product identifier mentioned below the product rack and that particular information will be included in the message for the convenience. There is a high probability that the product itself may contain some text information, and OCR may detect that also. But, in the proposed scheme we assumed that the product identifier is kept below the product and based on this assumption, we ignored other text information.

The experimental study is carried on a limited number products and the results observed during the experimental study is given in Table I. We have considered five products and assumed that maximum 6 products will be kept in the rack. We assumed the threshold for sending warning message is 3 (half of the maximum number of products that we can keep in a rack). The reliability of the proposed scheme can be measured only by checking whether the system is triggering warning message on time when the number of products in the rack goes beyond the limit.

TABLE I
STATUS OF TRIGGERING THE MESSAGE BASED ON PRODUCT COUNT

Product	**Status of triggering warning message**						
Number of products ->	**6**	**5**	**4**	**3**	**2**	**1**	**0**
Product 1 (101)	No	No	No	Yes	Yes	Yes	Yes
Product 2 (102)	No	No	No	Yes	Yes	Yes	Yes
Product 3 (103)	No	No	No	Yes	Yes	Yes	Yes
Product 4 (104)	No	No	No	Yes	Yes	Yes	Yes

The results are given in Table I shows that the proposed scheme is working reliably in most of the cases.

V. CONCLUSION

A computer vision-based approach for automated monitoring of the products in the supermarket is introduced in this paper. The proposed framework can be adopted in busy supermarkets. The implementation of the proposed scheme in supermarkets will help the managing companies to reduce the required manpower, increase the profit and provides better customer satisfaction. In the proposed scheme, we used a new structural similarity index based scene change detection algorithm to avoid the processing of all the video frames. In this work, we have used one camera to keep track of the products in one single rack, but in a real case, a single camera may be used to cover a large area which contains racks for different products. This issue can be taken into consideration for improving the scheme introduced in this chapter.

REFERENCES

[1] N. I. Radwan, N. M. Salem, and M. I. El Adawy, "Histogram correlation for video scene change detection," in *Advances in Computer Science, Engineering & Applications*. Springer, 2012, pp. 765–773.

[2] V. Manikandan and V. Masilamani, "Real-time scene change detection and entropy based semi-fragile watermarking scheme for surveillance video authentication," in *Proceedings of the 3rd International Symposium on Big Data and Cloud Computing Challenges (ISBCC–16')*. Springer, 2016, pp. 101–110.

[3] H. F. Huifu Zhuang, Kazhong Deng and S. Ma, "A novel approach based on structural information for change detection in sar images," *International Journal of Remote Sensing*, vol. 39, no. 8, pp. 2341–2365, 2018.

[4] P. Loncomilla, J. Ruiz-del Solar, and L. Martínez, "Object recognition using local invariant features for robotic applications: A survey," *Pattern Recognition*, vol. 60, pp. 499–514, 2016.

[5] M. Rizon, Y. Haniza, S. Puteh, A. Yeon, M. Shakaff, S. Abdul Rahman, M. Sugisaka, Y. Sazali, M. M Rozailan, and M. Karthigayan, "Object detection using circular hough transform," 2005.

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[7] M. Najibi, M. Rastegari, and L. S. Davis, "G-cnn: an iterative grid based object detector," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2369–2377.

[8] K. Sarkar, K. Varanasi, D. Stricker, K. Sarkar, K. Varanasi, and D. Stricker, "Trained 3d models for cnn based object recognition." in *VISIGRAPP (5: VISAPP)*, 2017, pp. 130–137.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[10] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 2366–2369.