International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

Digital Image Processing Using OpenCV to Detect Face and Objects

Prof. Amol Dumbare[1], Sudeshna Thakur[2], Neha Patil [3], Soumya Patil[4], Nidhi Hegde[5]

[1] *Professor, Dept. of Computer Science, Pimpri Chinchwad College of Engineering and Research, Maharashtra, India.*

[2,3,4,5] *Dept. of Computer Science, Pimpri Chinchwad College of Engineering and Research, Maharashtra, India.*

Abstract - *The future of the world is reliable on latest technologies and innovative ideas. The innovative concepts help to upgrade our knowledge. Technology is changing how people purchase grocery products, clothes, and other products. With retailers hopping on the cashier-less bandwagon in an effort to attract time-crunched shoppers, emerging technologies are helping them make it possible. The retail sector is changing dynamically with all the videos captured in unconstrained real-world scenarios is an interesting problem with many potential applications. This paper presents an efficient approach to digital image processing techniques, face detection algorithms, and object detection algorithms. In this paper we talk about the future of Retail Shops where in people can just take the item they want and leave the store without waiting in the checkout line. The AI that controls the environment follows each customer, noting each item taken and any that may be put back, so no errors are made and there are no charges for products not brought home.*

***Key Words*: Digital Image Processing, OpenCV, Face Detection, Object Detection, stereo cameras. Sensor fusion such as weight detection, Computer Vision. NFC, RFID**

1. INTRODUCTION *(Size 11 , cambria font)*

In recent years, online retail has exploded at an outstanding rate, yet the majority of people still actually prefer to shop-in store. Moving to the online shopping completely in particular categories is not likely possible. In such scenarios, buying the items from the shelves store-in seems to be a more attractive proposition. This system is a latest kind of shopping sense with no checkout required. This advanced technology is an attempt to bridge the gap between online and offline world. In this technology, a customer can skip long checkout queues without wasting time. The customer can get the app, enter into the store, take the required items, and leave. Pictures are captured when the customer enters the store, pick any products from the shelf, and leave the store with the products. With the help of facial recognition, the customer information is fetched which may include images of the customer, details such as height, weight and customer biometrics. That allows the surveillance system of the store to identify the customer and track them as they move throughout the store. Camera picks up images of the shelves that the customer approaches, the items they pick up and whether the picked item stayed with the customer or kept back in the shelf. These actions of the customer are automatically detected by the weight sensors placed at shelves while keeping a track of the products in the virtual cart. A virtual shopping cart is maintained for each individual user account. The surveillance system in coordination with multiple cameras identifies the customer details. It allows them to exit the store without a physical checkout by automatically charging their account after exit.

Computer vision is a fast-growing field and has brought various applications to a wide range of disciplines. It is strongly correlated to machine learning and image processing. Thereby suggests a more well-performed implementation of vision system in the recent advancement of algorithm. A highly optimized library of machine learning i.e.: OpenCV uses machine learning algorithms for searching faces within an image. Since faces are complicated, there are no existing tests that will predict if it found a face or not. The methodology of face detection can be applied to detect the eyes, nose tip, mouth etc. landmarks for localization. It can be utilized further for face geometrical normalization. The haar-like algorithm is also useful for feature selection or feature extraction for an object in an image. It can be used with the help of edge detection, line detection, center detection for detecting eyes, nose, mouth, etc. in the picture. It can also select the essential features in an image extracting them for face detection. Recognize and locate facial features: Fetch the coordinates of the eyes, ears, eyebrows, cheeks, lips, nose, and mouth of every face. Recognize facial expressions: Discover whether a person has his eyes closed or is smiling. Track faces over video frames: Get an identifier for each individual person's face which is detected. One can perform image manipulation on a person in a video stream to imply that the identifier remains consistent across invocations. In order to process video frames in real-time, face detection is performed on the device. It is fast enough to be used in real-time applications, for example, video manipulation.

1.1 LITERATURE REVIEW

There are variety of techniques that have been used to detect the face and objects. Some of the techniques that we studied are as follows:

1] VIOLA JONES ALGORITHM: The primary step of the Viola-Jones face detection algorithm is to convert the input image into an integral image. This is achieved by making each pixel equal to the complete sum of all pixels above and to the left of the concerned pixel. This permits for the calculation of the sum of all pixels inside any given rectangle using only four values. These values are the pixels within the integral image that coincide with the corners of the rectangle within the input image.

2] LBP FEATURE EXTRACTION: The concept behind using the LBP features is that the face images can be often seen as composition small patterns which are invariant to grey scale transformation. Combining these small patterns, a full description of the face image is obtained. The initial LBP operator labels the pixels of an image by thresholding the 3-by-3 neighborhood of every pixel with the middle pixel value and considering the result as a binary number.

3]FPGA based implementation: The proposed system is often applied in a video surveillance and tracking application which was tested for the information received from a camera. This implementation makes a complete system level hardware design described in a hardware description language and validated on the affordable DE2-115 evaluation board.

4] Object detection for computer vision using image segmentation: Image Segmentation for object detection is the most significant part of computer vision where the system has to identify objects differently from the background whether it is a face, hand, man or simply static objects. This method successfully identified the search object from the chaotic background. However, the strategy lacks its identification where the item seems multipart like human body.

5] FACE-NAME GRAPH MATCHING WITH NUMBER OF CLUSTER SPECIFIED: On the basis of the investigations of the noises generated during the affinity graph construction process, we build the name and face affinity graph in rank ordinal level and use ECGM with specially designed edit cost function for face-name matching. During this, a name affinity graph from script analysis and calculation of the confidence in the model, for example, assessing the uncertainties faces and pictures are identified the movie character. The more scenes where two characters appear together, the closer they are, and therefore the outcome of larger weights of the edges of the model itself for movie character identification.

6]FACE-NAME GRAPH MATCHING WITHOUT NUMBER OF CLUSTER SPECIFIED: This method requires no specification for the face cluster number. Standard affinity propagation [29] is used for face tracks clustering. The similarity input s(i, k) is set because the Earth Movers Distance (EMD) between face tracks. All face tracks are equally suitable as exemplars and therefore the preferences s (k, k). There are two types of messages, availability and responsibility, changed between face tracks With availability a(i, k) initialized to be zero, the responsibilities r(i, k) are computed and updated using the rule. The message-passing procedure converges when the local decisions remain constant for particular number of iterations. In our case, high cluster purity with sizable number of clusters is encouraged. Since there are no restrictions on the one-to-one face-name correspondence, the graph matching method expected to deal with the situations where several face clusters correspond to the character name. Hence, a graph partition step is conducted before graph matching.

OBJECT RECOGNITION MODEL BUILDING [6]

In this paper the author Duy Nguyen Phuong evaluate the requirement of the hybrid active vision system, object recognition model is firstly constructed. 20 random objects are selected from COCO dataset: airplane, bicycle, bird, boat, bottle, bus, car, cat, chair, cow, dining table, sheep, sofa, train, etc. Following literature review, MobileNet SSD is then used because the training and evaluating algorithm; and also, the training process is executed with TensorFlow. The pictures of the chosen items are labelled manually and used to train. The aim of the pan-tilt robotic frame is motion of the camera according to the movement of the object so that the center of the tracking bounding box will always be at the center frame of the camera. A PID model is implemented on the offset x and y of the bounding box center and also, the frame center. To deal with the problem with saccade movement, a delay is introduced to the control loop that is proportional to the calibrated value of the motors. A threshold is additionally applied on the control system for maintenance of the system overall stability.

MOTIVATION AND OVERVIEW OF DAVI MODEL [7] The author for this model explains the way to compute pixel disparities, we used the algorithm of Tanimoto et all. The depth at each pixel of an S3D image is obtained by triangulation of the computed pixel disparities. The resulting depth map is then used because the input stimulus to DAVI. Output accommodation is generated after processing this input stimulus and vergence response maps that are computed using the DAVI transfer function model. A) Image class preparation [8] For each object class we prepared a group of characteristics from sample images where the objects are segmented manually. Every single entry is defined by the set di = (hj, ej, PJ where 'hi' is the histogram of the image, 'ei' is the edge matrix of the image and Pi is the mean variation of the contrast inside the edge boundary of the object. B) Segmentation of the identical type of object from background [8] Here the author considered two images for segmentation one having the item used within the class preparation and another having n number of overlapping objects. We used Maximum Ownership Labeling to mark the required object and we stored the weight matrix difference as a parameter for the graph partitioning. For a pixel x that k

maximizes the ownership of F(x) within the MoG model M, the segment label c(x) = k. That is where MoG is the Mixture of Gaussians Model and is defined by F(x) is the feature vector.

As per the Author K Goyal, K Agarwal, R Kumar "Face detection and tracking" [10] Face recognition takes a picture from a video or a camera as input and outputs the recognized image topic. Facial features may include regions within the face, variations in the face structure, face cuts and angles that have been formatted and styled. Face extraction includes collecting the features from camera. Face detection involves removal of the background and focusing on the foreground by eliminating any other elements except the face region, however the system still contains some drawbacks because it cannot detect the head count which can be present due to overlapping of faces or improper recognition of two faces having similar facial features.

1.2 LIMITATIONS

The limitations of the system are that using Computer Vision and OpenCV we cannot go for very small products, means the small products cannot be detected and processed. This is not very cost effective since there are many small products which most of the customers will be looking for but keeping such products on shelf will result in a lot of ambiguities. Even packed items such as rice, dals etc. are not that efficient to keep in the stores. How will the children entering the shop and purchasing the items will be tracked? As children does not have smart phone, items picked up by the children will be charged in whose account. What if the customer takes the fresh juice bottle from the shelf and replaces it with an empty bottle? Or else consumes there itself and keeps the empty bottle back on the shelf. There are many observing cameras all around the stores and shelves, but much cheaper solution would be a weight sensor under each shelf. All the items on a particular shelf section have the identical weight. When the customer consumes the juice bottle there it and keeps the empty bottle back, will still be billed as weight sensor and cameras will identify the misbehavior of the customer.

2. PROPOSED MODEL

We are proposing a computer vision-based system for tracking and tracing objects (humans, products) with the integration of digital image processing. There are two main situation which is to be considered which using this system they are as follows:

1) Who has taken the item? Bluetooth beacons can identify whose mobile device is nearest the shelf. A dense beacon network can be accurate to within less than 0.5 meters.
2) The application in the shopper's mobile device will be communicating with the store's beacon network.
3) What item was taken? Shelf cameras will detect when an item has been removed or added and what that item looked like. This data feeds into an AI System.
4) AI system will analyze the vision, weight and data of stock location to recognize the items added or removed efficiently.
5) Weight sensors on the shelf are used for the detection of weights of the removed or added items. This data is fed as input to AI system.

In our proposed system, we plan to build a product tracking and tracing application using OpenCV to keep a track of the products purchased in the shop right from the rack of product to the customers bag so that no errors are made and there are no charges for products not brought home.

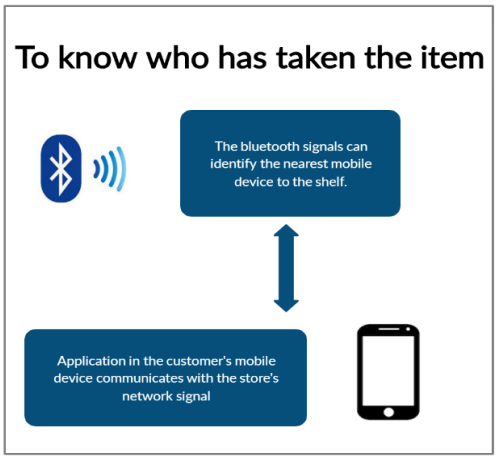
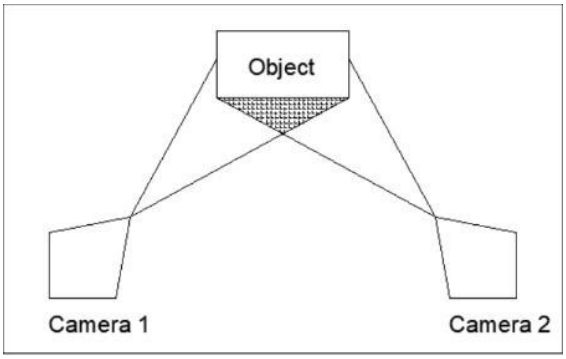


Fig -1: Explains who has taken the item

Fig -2: Explains what item was taken

3. STEREO VISION CAMERAS

The stereovision camera estimates the 3D coordinates of a picture or its features. It analyzes the foreground and background information in the picture and extracts objects with the help of a computer. It does not even require human supervision, and with such good efficiency closest to the human perception. Stereo Vision basically recreates the human vision system using multiple 2D views of the same scene to Compute the 3D depth information about the scenario. This depth information can be used to track moving objects in space, in our project are the moving human beings in the retail store who will add or remove items from their cart. This depth information can then be processed through the code to extract features and identify the human being and add items in his virtual cart for the billing amount.



4. CHALLENGES AND FUTURE SCOPE

We have several challenges that need to be resolved while building this system. Keeping track of small objects is a challenge in this system, also we cannot go for packing of items like rice, dal, etc. Biggest limitation we face is the list of items we can provide in such environment. The threat we can see for this system is that once it gets hacked people will not use it further because of its credibility also financial lock it can incur is tremendous. The video surveillance should be very efficient since there will be number of customers moving dynamically in real time with or without the objects picked. The observed problem is that the image that is considered might be obtained from any source and the image may be captured in conditions having no constraints. The human face has a high degree of variability in its appearance like, the face(s) in the image may vary in parameters like size, location, pose, quality, distance, alignment, lighting conditions, overlapping with many other faces or objects, facial expression etc. These kinds of unconstrained conditions make the accuracy of face detection a challenging and tedious task.

The system will speed up the customers' purchase process, especially for the customers who have less time to shop. Nowadays in a fast-paced society, it goes hand in hand with people being busy. The cameras, location in store still not known, will scan the barcode of the product, so that it acknowledges that you still have the item, and not mistaken you for someone else putting the identical item back. These cameras can detect a bar code correctly even if it is mostly obscured, so speedily putting an item in your bag won't cut it.

5. CONCLUSIONS

The system plans to totally open its checkout-free grocery shopping and provides the best experience for the customers while shopping. This strategy is an evolution from the self-checkout version that is utilized in many grocery chains today. Advanced technology and computer integrated inventory management systems permit the customer to take off the products from shelves, put them in their carts, and leave the retail shop without going through a checkout line. Besides the convenience to the customer, it also evidently plans to profit by reducing the checkout clerks. The value of implementation, maintenance, and sustaining the system may offset or exceed cost-savings via reduction of checkout clerks.

REFERENCES

[1] Nivedha N, Prof. Ganesh Pai, Dr. M Sharmila Kumari "Recent Trends in Face Detection Algorithm".

[2] Chaitra T. K, M.C.Chandrashekhar, Dr. M. Z. Kurian "Face Detection and Feature Extraction".

[3] K. Wankhede. Dr. B Wukkadada, V Nadar "Just Walk Out Technology and its challenges".

[4] G.Bakkiyaraj, V.Dinesh. "The Process of Face Detection and Recognition in Multi-Videos".

[5] P Munro, A Gerdelan "Stereo Vision Computer Depth Perception".

[6] Duy Nguyen Phuong "A hybrid active vision system for real time application running object recognition". 23 May 2019.

[7] Heeseok Oh, Sanghoon Lee." Stereoscopic 3D Visual Discomfort Prediction: A Dynamic Accommodation and Vergence Interaction Model". 7 Dec 2015.

[8] Debalina Barik , Manik Mondal "Object detection for computer vision using image segmentation". 29 July 2010.

[9] Chaohui Lü, Xi Wang, Yinghua Shen. "A stereo vision measurement system based on OpenCV". 20 Feb 2014.

[10] K Goyal, K Agarwal, R Kumar "Face detection and tracking" 18 Dec 2017.