Proceedings of the Sixth International Conference on Trends in Electronics and Informatics (ICOEI 2022)
IEEE Xplore Part Number: CFP22J32-ART; ISBN: 978-1-6654-8328-5

Object Detection and Recognition using Amazon Rekognition with Boto3

Vivek Sharma
Assistant Professor, GLA University, Mathura(UP), India
viveksharma.cea@gla.ac.in

2022 6th International Conference on Trends in Electronics and Informatics (ICOEI) | 978-1-6654-8328-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICOEI53556.2022.9776884

Abstract— **In Image Processing, the problem of Object Recognition is regarded as difficult. It's an important aspect of PC vision because so many PC vision programmes rely on it for effectiveness. To address this issue, a number of identifying terminology and frameworks have been discussed for quite some time. The purpose of this paper is to give a brief overview of a range of platforms in the field of computer vision and object recognition. For the most part this paper covers a relatively mature service from AWS(Amazon Web Services), Amazon Rekognition which is easy to integrate into other services like Lambda, providing huge volume of details in its analysis. The paper incorporates different viewpoints that have been utilized by various analysts for object spotting and recognition. Rekognition is able to recognise objects and scenarios within photographs in this article, returning a confidence measure ranging from 0 to 100% for each label returned. It accomplishes this by using AWS' Python client Boto3 and the detect labels method. This works by supplying an image, a limit on the number of labels to return, and a confidence level for filtering out ambiguous results.**

Keywords— Rekognition, Boto3, Amazon Web Services, AWS Lambda, AWS S3, AWS Elastic Search.

I. INTRODUCTION

Two pieces of article acknowledgment are used to classify objects. The basic goal is to categorise an article into one of a few predetermined groups[1]. The purpose of detection is to separate objects from their surroundings. Object recognition is elicited in a variety of ways. Normally, items must be identified over jumbled, boisterous foundations and different objects under diverse enlightenment and complex situations[2]. Appropriate highlight depiction is a significant advancement in an article identification system because it improves execution by separating the entity from the foundation or other questions in a variety of lightings and settings. Object recognition highlights are split into two categories: sparse and dense representations. Intrigue point identifiers are used to distinguish structures on the object, such as edges and masses, for poor component depictions[3]. Object identification is a challenging field in PC representation and example examination inquiry about region. There have been several strategies proposed and developed. Diverse strategies for identifying objects using various techniques, such as optical flow, point detectors, temporal differencing, frame differencing, and background subtraction, are currently available. This work have also evaluated the precision speed of various strategies and identified the benefits and drawbacks of each strategy. This work have also discussed the order tactics and component types of various object discovery strategies, such as edge-based element type and fix-based component type, and so on. This work look for a correlation between object ordering strategies and investigate the precision percentage and favourable conditions among them[4]. The prospect for this exploration region's future success is extremely promising. The scope of novel object identification and characterization strategies that could be developed is vast.

Rest of this paper is organized with the content of challenges in object recognition. Followed by it is already related work done under object recognition and then comparative study of all the terminology. By discussing the proposed work in continuation there is conclusion followed by acknowledgement.

II. CHALLENGES IN OBJECT RECOGNITION

1. Mirroring: The object recognition must be able to recognize mirror image.
2. Occlusion: The situation where object in a picture is not clearly visible is called as occlusion.
3. Rotation: The framework must be skilled in such a way that it handles the condition when the picture is in pivoted structure.

4. Lightning: During the course of the day, the lighting factor can affect. Similarly, the lighting in a photograph can be affected by the weather. For the same article, the lighting environment for in-entryway and open-air photographs can differ. The light of a photograph can be influenced by shadows. The framework must be able to recognise the item in the image, regardless of the fulmination [5].
5. Positioning: The system must handle such images uniformly in which template matching is used for positioning of the object in the image.
6. Scale: Variations in the item's size should not affect the object recognition framework's accuracy. These are only a handful of the challenges that can emerge during object recognition. By overcoming the aforementioned problems, an effective and powerful article recognition substructure can be created [6].

III. Related Work

Object detection and tracking are projected to play a big role in a lot of PC vision and pattern matching applications, like video arrangement, vehicle routing, observation, and self-ruling robot steering. Distinguishing objects and recognising designs at the periphery of a video sequence are all part of item identification[7]. At the following technique, whether in each edge or when the item first appears in the movie, an object detecting system is necessary. The most well-known method for object recognition is to use data in a single case. Although some object recognition algorithms use real-world data acquired from inspecting a succession of edges to reduce false positives and enhance accuracy [8]. There aren't many object recognition techniques that look like this[9].
Rajesh Tripathi et al[10] has given a review on deleted objects in visual surveillance. Ambika et al [11] has proposed an implementation of storage in Virtual Private Cloud using S3 on AWS platform. Saurabh Singhal [12] has reviewed various data deduplication techniques on cloud platform.

1. Optical Flow

The apparent speeds of items in an image are conveyed by optical flow. You may calculate the speeds of objects in a video by comparing optical streams between video outlines. When everything is said and done, drawing things that are closer to the camera will display more visible movement than detached articles moving at the same rate. Optical stream estimation is a technique used in computer vision to depict and measure the movement of objects in a video stream, which is commonly used in article locating and tracking systems.
It is classified into two sections that are Sparse Optical Flow and Dense Optical Flow. Sparse optical flow gives the stream vectors of some "intriguing highlights" (state hardly any pixels portraying the edges or corners of an item) inside the casing while Dense optical stream, which gives the stream vectors of the whole casing (all pixels) - up to one stream vector for each pixel. As you would've speculated, dense optical flow has higher exactness at the expense of being moderate/computationally costly.

2. Frame Differencing

This strategy is through the contrast between two back to back pictures to decide the nearness of moving items. The Frame contrast appears to be the simplest sort of background removal. In frame differencing, also known as temporal difference, the video outline at time t-1 is employed as the base model for the casing at time t. This method is touchy to clamor and varieties in brightening, and doesn't consider nearby consistency properties of the change veil. This technique additionally neglects to section the non-foundation objects on the off chance that they quit moving. Outline differencing is unlikely to distinguish the interior pixels of a huge, evenly coloured moving object since it only employs a single past casing. This is commonly referred to as the aperture issue. It offers a lot of versatility, but it's sometimes difficult to capture a complete blueprint of a moving thing, which might cause the unfilled marvel to show up, thus the identification of the moving item isn't exact.

3. Point Detectors

This model demonstrates a technique for identifying an object depending on the discovery of point correspondences between the reference and objective images. Regardless of scale or in-plane rotation, it can discern objects. It's also effective against a small number of out-of-plane pivots and impediments. This object identification approach works best for objects with non-rehashing surface examples that lead to one-of-a-kind element matches. This method isn't probably going to function admirably for consistently hued objects, or for objects containing rehashing designs. It's important to note that this method is intended to identify an object, such as the horse in the reference image, instead of any horse. It detects objects of a single category, like individuals or features, using vision. Cascade Object Detector

4. Temporal Differencing

The triple methodology starts with calculation of contrast pictures utilizing transient data. Contrast pictures are determined by subtracting two info outlines, at every pixel position. Rather than creating contrast pictures utilizing the conventional consistent casing distinction approach, we propose utilizing a fixed number of interchange outlines revolved around the present casing. This methodology helps in decreasing the computational unpredictability without settling on nature of the distinction pictures. After calculation of distinction pictures, a novel post-handling plan is utilized by using gamma amendment factor and Mahalanobis separation metric to decrease bogus positives and bogus negatives. Article division is at long last performed on the refined distinction picture by a neighborhood fluffy thresholding plan. This stays away from issues that are generally experienced in hard thresholding, particularly pixel misclassification, which is the most significant one. For powerful trial examination, recordings, CAVIAR, and datasets have been utilized. These chose recordings contain a wide assortment of regular difficulties looked during object discovery. A few models are the nearness of dynamic foundations, shadows, awful climate, and so forth. The outcomes set up the adequacy of the proposed plot over a portion of the current plans both subjectively and quantitatively as outlined in the test result area.

5. Background Subtraction

When employing static cameras to detect in motion objects, background removal is a common technique. The methodology uses the distinction between the current casing and a reference outline, sometimes referred to as a "foundation picture" or "foundation model," to detect the moving items. Foundation subtraction is commonly employed when the image in question is a section of a video stream. Foundation subtraction provides important instructions to a variety of PC vision applications, such as reconnaissance tracking and human posture assessment. Foundation subtraction is often based on a static foundation theory that isn't always appropriate in real-world settings. Reflections or vivified images on screens cause foundation shifts in indoor situations. Also, because of wind, downpour or brightening changes brought by climate, static foundations strategies experience issues with outside scenes.

IV. COMPARATIVE STUDY OF OBJECT RECOGNITION METHODS

A comparative study of different Object Recognition methods is listed in the **Table 1**.

Table 1: Comparative Study of Object Recognition Methods

Methods		Accuracy	Time Efficiency	Feedbacks	User Rate
Background Subtraction	Gaussian of Mixture	Moderate	Moderate	- Less memory storage	40%
	Approximate Median	Low to Moderate	Moderate	-Doesn't cope up with multimodal Background	
Optical Flow		Moderate	High	-No requirement of sub sampling of frames for the creation of an adequate background model	20%
				-Requires buffer for the computation with recent pixel values	
Frame Differencing		High	Low to Moderate	-Best and easy method for static background	30%
				-Background with no moving objects	
Temporal Differencing		Moderate	High	- Computationally less complex and good for dynamic changes	10%
				-More sensitive to threshold and leaves holes in the foreground section	

V. PROPOSED WORK

Extending from hand written number acknowledgment, to facial recognizable proof, to speech analysis, AI demonstrates significant over a bounty of fields. With a developing network joining Amazon Web Services, the interest for classification algorithms in the cloud keeps on developing.

Propelled in 2016, Amazon Rekognition **Figure 1** permits us to distinguish objects, look at compare faces, and moderate pictures and video for any unsafe content [13]. This is just a couple of the numerous

highlights it conveys. Requiring no training information or any information on our piece of the fundamental AI procedures, Rekognition effectively connects to AWS applications to convey detailed analysis [14].

Figure 1: Amazon Rekognition

Using an AWS Lambda function running Python 3.8, let's explore Rekognition's capabilities to demonstrate its impressive functionality out-the-box[15].

The first step is to make a list of every image in an S3 bucket. The Amazon S3 synchronous List API activity is typically used to accomplish this. Depending on the number of objects in the bucket, this could take a long time. The new Amazon S3 inventory tool provides a comma-separated values (.csv) level record yield of objects and their accompanying metadata on a daily or weekly basis. It can keep all of the items in an S3 bucket or all of the articles with a prefix (that is, objects that have names that start with a common string). This tool is used in the approach to get a number of objects stored in a bucket and examine only images provided in the png or jpeg formats that Rekognition supports.

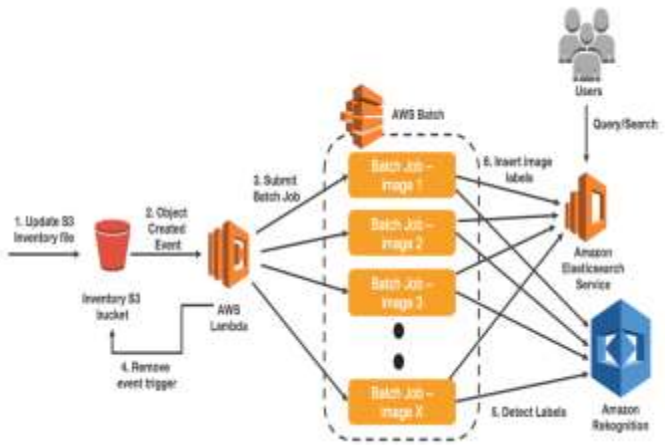

Figure 2: Image Backfill Workflow Component

Figure 2 shows the workflow diagram of Image Backfill components. This is how it works:

1. The Amazon S3 inventory tool creates a csv file that list out all of the objects in an images bucket and stores it to the inventory bucket in Amazon S3. The S3 service generates the inventory file on a daily basis.
2. The inventory bucket is configured to perform the Lambda function on any object with the.csv.gz file extension when a new version of the compressed S3 inventory.csv file is stored to the destination S3 inventory bucket, thus a Lambda function is launched.
3. For each image it discovers, the Lambda function analyses the contents of the csv file and creates a new AWS Batch Job with the image bucket and name it, then sends it to the AWS Batch queue. AWS Batch launches EC2 instances and runs batch jobs on those instances to execute batch jobs which have been submitted to the job queue.
4. To avoid the AWS Batch backfill workflow running several times, the Lambda function disables the S3 event trigger.
5. The AWS Batch tasks use an image bucket and named as input to check the Amazon ES domain index to see if the image has been processed before. If the image hasn't been processed yet, the AWS Batch Job uses the Amazon Rekognition to identify labels.
6. The AWS Batch jobs preserve the labels that Rekognition returns for the image in the Amazon ES domain index.

Detecting Objects and Scenes

Rekognition can distinguish items and scenes inside pictures, restoring a confidence metric somewhere in the range of 0 and 100% for each name (or tag) returned. Working with AWS' Python customer Boto3, we run the detect_labels activity to do this. Giving a picture, a cap on the number of labels to return, and a confidence threshold to sift through uncertain outcomes, Rekognition synchronously plays out this examination[16]. Pictures can either be given straightforwardly in bytes or referenced from an Amazon S3 bucket.

DetectLabels gives output as bounding boxes for occurrences of regular item marks in a variety of occurring objects. An Instance object comprises of a Bounding Box object, for the area of the mark on the picture[17]. It additionally incorporates the certainty for the recognition of bounding boxes. It likewise restores a various leveled scientific categorization of distinguished marks.

Syntax:

```
{
   "Image": {
      "Bytes": blob,
      "S3Object": {
         "Bucket": "string",
         "Name": "string",
         "Version": "string"
            }
         },
      "MaxLabels": integer,
       "MinConfidence": integer
   }
```

Figure 3 shows the objects and scene detection using Amazon Rekognition with confidence score.

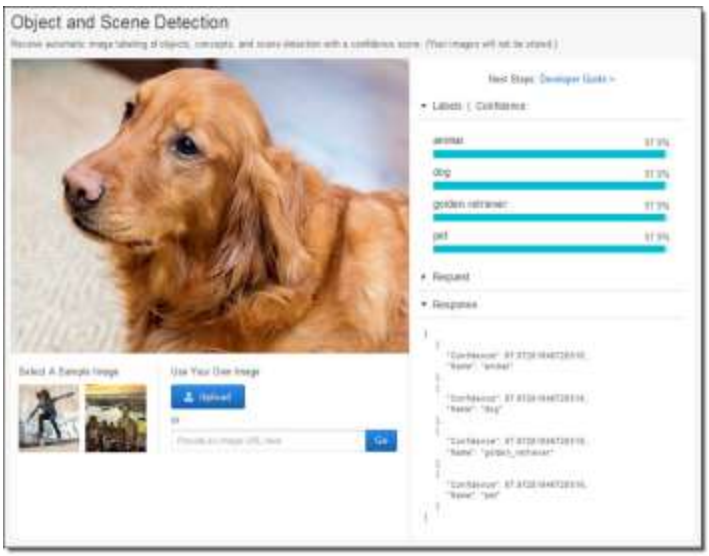

Figure 3: Output of Amazon Rekognition with Confidence Score

Figure 4: Output of Boto3 Image Recognition

Figure 4 is the final output of Boto3 Image recognition. No training was required to produce these labels. Rekognition runs this examination utilizing its own preparation information, which incorporates previous user data. This is allowed under the present information security strategy an arrangement which we may quit from by connecting with AWS Support.

VI. CONCLUSION

This research paper carries all the principal methods of object recognition and classification which have been inscribed in it. Object identification, highlight identification, and object classification are among them. For these levels, the most often utilized and widely understood tactics have been extensively defined. This study has already covered a variety of object identification approaches. A moderately develop service from AWS, Rekognition is easy to integrate into different services like Lambda. Use Rekognition and the Boto3 application to process subsequent photographs sent to a similar bucket if you have a bucket with multiple pictures that you need to evaluate to determine what items and situations they contain. The most significant part of an Object detection is the collection of the objects and forming groups of it as per the classification [18]. Among the numerous techniques for object grouping the majority of the researchers lean toward consistency based and shading based object classing [19]. Advancement in study may open the ways to discover proficient theorems and strategies to reduce the implementation and computational costs of identifying the item, as well as the time it takes to do so, for a variety of recordings with diverse features and to improve the exactness rate.

There are certain limitations with AWS recognition like it supports JPEG and PNG image formats only. Also, image size is restricted upto 15 MB when taken as S3 object and upto 5MB as raw image. The smallest object that could be recognized at least 5% the size of the of the image(in pixel).

REFERENCES

[1] Elgammal, A.,Duraiswami, R.,Harwood, D., Anddavis, L. 2002.Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. Proceedings of IEEE 90, 7, 11511163.

[2] Rupali S.Rakibe, Bharati D.Patil, Background Subtraction Algorithm Based Human Motion Detection,International Journal of Scientific and Research Publications, May 2013.

[3] Abhishek Kumar Chauhan, Prashant Krishan, Moving Object Tracking Using Gaussian Mixture Model And Op- tical Flow, International Journal of Advanced Research in Computer Science and Software Engineering, April 2013.

[4] Kinjal A Joshi, Darshak G. Thakore A Survey on Moving Object Detection and Tracking in Video Surveillance System International Journal of Soft Computing and En- gineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July, 2012.

[5] N. Paragios, and R. Deriche.. Geodesic active contours and level sets for the detection and tracking of moving objects. IEEE Trans. Patt. Analy. Mach. Intell. 22, 3, 266280, 2000.

[6] S. Zhu, and A. Yuille. Region competition: unifying snakes, region growing, and bayes/mdl for multiband image segmentation. IEEE Trans. Patt. Analy. Mach. Intell. 18, 9, 884900, 1996.

[7] Rupesh Kumar Rout A Survey on Object Detection and Tracking Algorithms Department of Computer Science and Engineering National Institute of Technology Rourkela Rourkela 769 008, India.

[8] W.T. Lee and H. T. Chen, "Histogram-based interest point detectors," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1590-1596.

[9] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. Acm Computing Surveys (CSUR), 38(4):13, 2006.

[10] Tripathi, R.K., Jalal, A.S. & Agrawal, S.C. Abandoned or removed object detection from visual surveillance: a review. Multimed Tools Appl 78, 7585–7620 (2019). https://doi.org/10.1007/s11042-018-6472-9

[11] A. Gupta, A. Mehta, L. Daver and P. Banga, "Implementation of Storage in Virtual Private Cloud using Simple Storage Service on AWS," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2020, pp. 213-217, doi: 10.1109/ICIMIA48430.2020.9074899.

[12] S. Singhal, P. Gupta, S. Singh, P. Agarwal and K. Kumar, "Data Duplication Removal Technology Using AWS Services," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), 2021, pp. 1-4, doi: 10.1109/ISCON52037.2021.9702380.

[13] Rupesh Kumar Rout A Survey on Object Detection and Tracking Algorithms Department of Computer Science and Engineering National Institute of Technology Rourkela Rourkela 769 008, India.

[14] Sen-Ching S. Cheung and Chandrika Kamath, Robust techniques for background subtraction in urban traffic video.

[15] K.Srinivasan G.Sainarayanan Improved Background Subtraction Techniques for Security in Video Applications.

[16] Himani S. Parekh1, Darshak G. Thakore 2, Udesang K. Jaliya 3 A Survey on Object Detection and Tracking Methods , International Journal of Innovative Research in Computer and Communication Engineering , Vol. 2, Issue 2, February 2014

[17] Himani S. Parekh1, Darshak G. Thakore 2, Udesang K. Jaliya 3 A Survey on Object Detection and Tracking Methods , International Journal of Innovative Research in Computer and Communication Engineering , Vol. 2, Issue 2, February 2014.

[18] Zhan Chaohui, Duan Xiaohui, Xu Shuoyu, Song Zheng, and Luo Min. An improved moving object detection algorithm based on frame difference and edge detection. In Image and Graphics, 2007 ICIG 2007. Fourth Interna- tional Conference on, pages 519523. IEEE, 2007.

[19] Changick Kim and Jenq-Neng Hwang. Fast and automatic video object segmentation and tracking for contentbased applications. Circuits and Systems for Video Technology, IEEE Transactions on, 12(2):122129, 2002.