TEXT EXTRACTION FROM IMAGES USING CLOUD SERVICES

[1]Aditi Narkhede, [2]Manswi Meghe, [3]Ujwal Rahile, [4]Vaibhav Dhawale, [5]Rakhi Wajgi,[6]Ashish Bhagat
[1,2,3,4,5]*Department of ComputerTechnology, Yeshwantrao Chavan College of Engineering Nagpur, India*
[6]*Research Consultant, Jawaharlal Nehru Medical College, DMIMS, Sawangi (M), Wardha*

Abstract

Nowadays everyone has mobile phones and people usually capture images of textual documents, handwritten papers,banners, packed food, prescriptions,bills and anything which is of their work. Sometimes they need to work on same data which is present in images and they write/type it manually which requires a lot of time. Also, there is huge demand of storing the data or information available in paper documents format into a digital format so that the data is searchable and editable. Hence there is need of software which is able to extract text from captured images. So we are developing an android application which can extract text from uploaded images. The uploaded image is analyzed using AWS rekognition service of amazon web services which is implemented using various machine learning algorithm which convert the detected text into machine readable text.This AWS recognition is a cloud-based service which allow developers to directly use the machine learning algorithms without actually knowing about it which is a feature of platform as a service (PaaS) of Cloud Computing. With this service we can built an application using machine learning concepts without prior knowledge of it, but by simply accessing this service through cloud.The main advantage of this application is that it is free of cost as we are using all the services by remaining in a free-tier only. The objective of this paper is to compare different methods used for extracting text from images with the method we are using and describing our method of implementation for recognizing the text from image provided for better understanding of the reader by using particular sequence of different processes.

Key Words:AWS-Amazon Web Services, AWS Rekognition, Cloud, AWS Android SDK Identity Pool.

1. Introduction

Because of the growing demand for the software system to recognize text in computer system when information is scanned through paper documents or captured images as we know that we have number of newspapers and books which are in printed format related to different subjects. Generally, the images are divided into three types: document images, scene images and born-digital images. Document images are the image-format of the document. It is a way of transforming paper-based documents into image format for electric read. Such document images have good quality and the background is clean. Scene images contain the text, such as the advertising boards, banners, blackboards etc. which is captured naturally. As scene images are taken by the camera, scene text is embedded in the background as a part of the scene. Scene images are complex because the backgrounds are complex containing the text in different sizes, styles and alignments, and also the resolution of the images is low. In addition, scene text is usually affected by lighting conditions and perspective distortions. Born-digital images are generated by computer software and are saved as digital images. Compared with document images and scene images, there are more defects in born-digital images, such as more complex foreground/background, low resolution, compression loss, and severe edge softness[4].

We are trying to build a mobile application which can be used to get the text which is present in such images. Using this application user will be able to upload or directly capture the image from which text is to be extracted. For building this application we are using machine learning services provided by Amazon cloud, android studio framework for

building frontend of the application, and managing the backend using other amazon web services.

Cloud computing is the on-demand delivery of IT resources over the Internet with pay-as-you-go pricing. Instead of buying, owning, and maintaining physical data centres and servers, you can access technology services, such as computing power, storage, and databases, on an as-needed basis from a cloud provider. We are using AWS Rekognition service of Amazon Web Services for Extraction of text from images.

Amazon Rekognition makes it easy to add image and video analysis to applications. We just provide an image or video to the Amazon Rekognition API, and the service can identify objects, people, text, scenes, and activities. It can detect any inappropriate content as well. Amazon Rekognition also provides highly accurate facial analysis, face comparison, and face search capabilities. We can detect, analyse, and compare faces for a wide variety of use cases, including user verification, cataloguing, people counting, and public safety. Also, Text extraction of complex images can be applied to a variety of fields where the information needs to be analysed and understood. If the images can be automatic understood and indexed by computer, the efficiency of running digital libraries and video database system will be greatly improved[10].

Amazon Rekognition Text in Image enables to recognize and extract textual content from images. Text in Image supports most fonts, including highly stylized ones. It detects text and numbers in different orientations, such as those commonly found in banners and posters. When we upload the image it returns text present in it[11].Text extraction technology can be applied to detect scene text from images taken with laptops, phones and other equipment so as to be applied to maps, navigation, automatic translation, foreign-related tour guides for the blind Guide, walking robots and intelligent monitoring system. Vehicle license and scene subtitles have many features in common so text extraction can be used to supervise the traffic in real time. After text extraction from highway video flow the traffic situation can be overseen and vehicle licenses can be recognized easily from traffic accidents, which can improve the efficiency of the transport regulatory. Amazon Rekognition Text in Image enables to recognize and extract textual content from images [4].

Other than AWS Rekognition many different methods are used for same purpose some of those methods are discussed below.

2. Different methods and applications used for extracting text from images

(1)From Paper “Text Extraction from Images”

The Text Extraction system designed here first receives an input in the form of image which contains some textual information. The output of this system is in electronic format i.e. text information present in the image is stored in computer readable form. Further the system is divided into four parts: Pre-processing, System Training, Text Recognition and Post-processing. In the Pre-processing part activities like scanning documents and storing them as images are performed. System Training part is used to train the system for text recognition, it converts the printed documents into editable and searchable documents. The Text Recognition module is used for text recognition in the output image of pre-processing part and give output data in computer understandable form by using the Feature Extraction by using techniques like Principle Component Analysis (PCA), Linear Discriminate Analysis(LDA), Independent Component Analysis (ICA), Chain Code(CC), zoning, Gradient Based features, Histogram etc. and them performing classification using Artificial Neural Networks(ANN) and finally in the Post-processing part the output which is in the computer

Understandable form is stored into some proper format for further use [2].

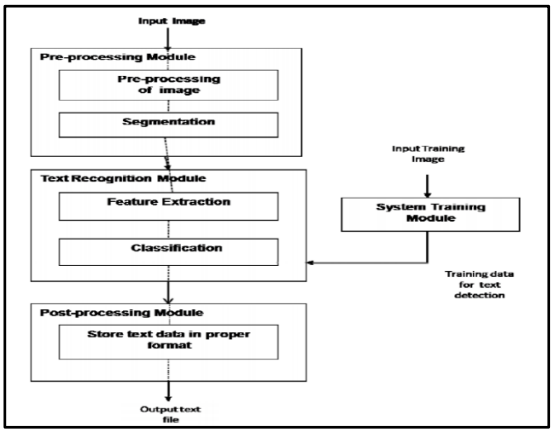

Fig.1

(2) From Paper “A Robust Algorithm for Text Extraction from Signage Images”

Here Maximally Stable Extremal Regions (MSER) detector is used to extract text areas because it shows better performance to lightning conditions. It is a blob detection method. It can extract texts from image without considering size and font. MSER can address the co-variant regions within image and distinguish them. MSER performs well

with a small computation cost and it can detect text while the image is in low quality. To enhance the MSER regions, novel enhancement technique is used by combining edge detection algorithm. Enhance edge image is then put to the stroke width detection algorithm. Based on stroke width information, connected components are obtained. Afterwards connected components are filtered to remove non text pixels. This algorithm shows best performance on light effected text images and noisy images[3].

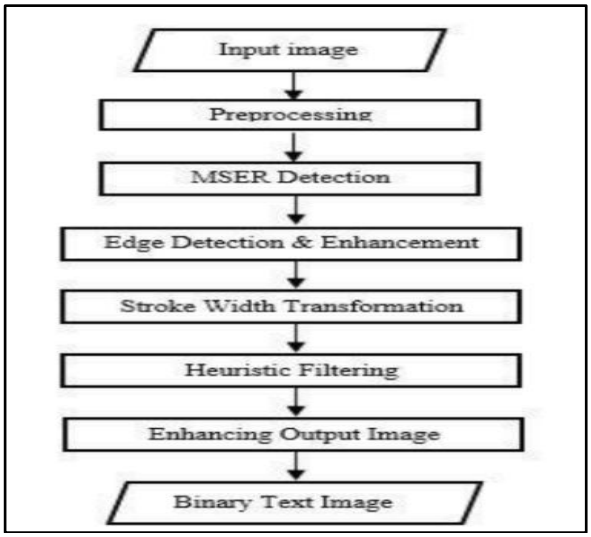

Fig.2

(3)From Paper "Research on the text detection and extraction from complex images"

Here, the research of complex images text extraction recently is reviewed, the complex image text extraction, image segmentation, object recognition is described and the research results are presented. Optical Character Recognition (OCR) is one of the text extraction techniques that mainly deal with document images. At this stage, the study includes preprocess (binarization, alignment detection, alignment correction, and character segmentation), character recognition, layout analysis, and graphic symbol recognition. The methods for detecting text can be categorized in three groups: the method based on connected component (CC) consisting of two steps: The first step is to draw CC from images using a specific method and the second step is to estimate whether the CC is text CC or not based on CC feature and CC relative feature, the method based on texture dealing with text regions as a special texture. The region is identified as text region or not according to the extracted relevant texture of the candidate regions. To overcome the disadvantages mentioned above, researcher present a hybrid approach, which takes the advantages of both texture-based and CC-based methods, to robustly detect and localize texts in natural scene images. In this method a text region detector is designed based on the texture to estimate the probabilities of the position and the scale of the text and then the candidate regions analysis and estimated to be a text region or not, and the method based on corner which is inspired by the observation that the character, especially in the text and the caption, usually contains multiple corner points. The method is to describe the text regions formed by the corner points using several discriminative features. The research on the method based on corners is still in the early stage. Compared with texture-based method, this method is faster but the performance is less satisfied[4].

(4)From Paper "Cloud based Text extraction using Google Cloud Vison for Visually Impaired applications"

Here an assistive device is implemented for visually impaired person such smart reader that is capable of capturing an image from a camera and extract the text from the captured image and further to convert the text to speech as voice - based output to assist the visually impaired people. The captured image is analyzed using Google Cloud Vision API Optical Character recognition (OCR). In order to extract text, image preprocessing methods to remove any noise or blur in the captured image is used so that the accuracy can be increased. Further, software-based text to speech to convert the text to speech as voice output is included. In this project Raspberry pi is used interfaced with Night vision camera and speaker to produce the converted text as voice. By applying Google Cloud vision OCR with preprocessing results in high accuracy level of the extracted text through image enhancement methods[5].

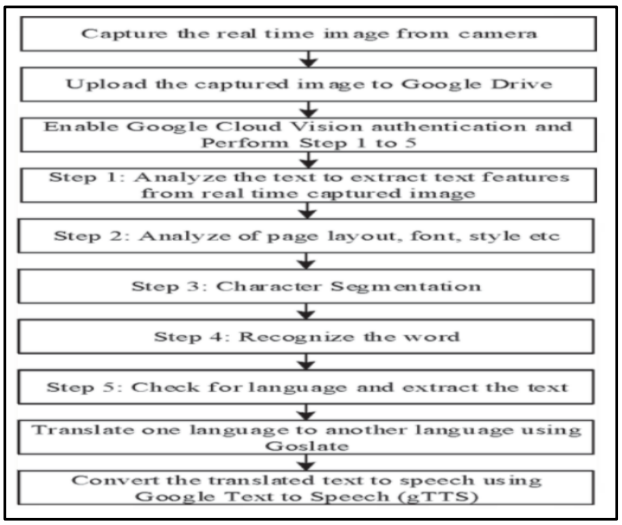

Fig.3

All the methods mentioned above somehow involve developer:

- In understandinghow actually the algorithm works.
- Developing a machine learning model
- In spending money on paid service
- And in understanding the complex processes involved in its working.

Talking about our implementation where we are using amazon rekognition which is a machine learning service provided by AWS, can be used in their free tier along with that developer do not have to worry bout the backend working of the service. No need of having prior knowledge about machine learning for implementing the concept of text extraction. Developer only needs to upload photo and then format the extracted text according to business requirement. Only thing that developers need to concentrate on is how to develop frontend and link it with the backend so that directly photo uploaded by user will reach to AWS rekognition. Also need to look after formatting of unformatted data obtained after extraction since output of recognition is JSON script. This is according to developers' point of view.

Whereas talking about users' point of view they only need to open app upload photo and get the text which is extracted from image. They don't have to worry about how actually application is working. This is what SAAS in cloud computing tells us about.

3. Our methodology

Our application is basically based on the Extraction of text from Images.We are developing an android application where the image can be captured or chosen from the gallery. We are developing android application using android studio as framework. Using this mobile application user will be able to retrieve the important text from images in user's required format. After capturing the image the android application will store the image in the S3 bucket one of the storage service provided by AWS, itis used to save the different type of data on Amazon Cloud.Now this S3 bucket is used by AWS Rekognition to get the image from which text is to be extracted.After Extracting text from image we get the output of extracted text in JSON format.Now we use Python script to arrange the extracted data in theproper format according to user's requirement.Then the formatted data will be sent back to the mobile application either in .CSV file (or .TXT file).

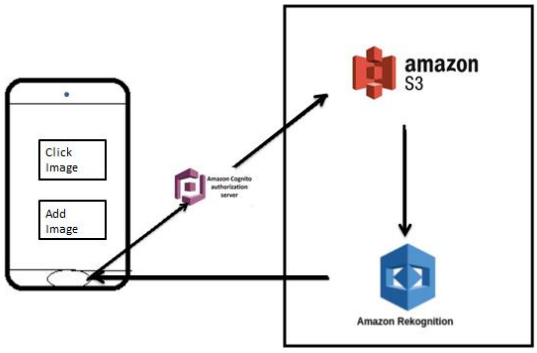

Fig.4 Architecture Diagram

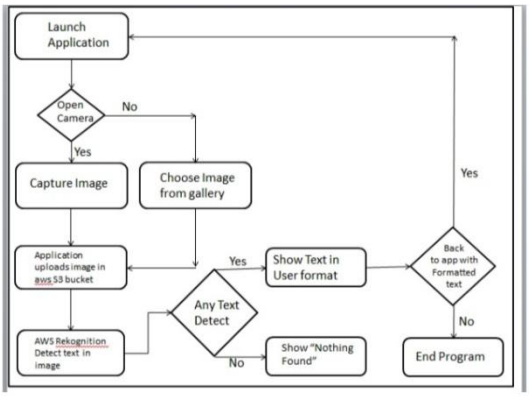

Fig.5 Flow Diagram

3.1Requirements for Methodology

3.1.1 Account of Amazon Web Service(AWS)

Amazon Web Services offers reliable, scalable, and inexpensive cloud computing services. Free to join, pay only for what you use. Itenables you to select the various operating system,programming, language,
web application platform,differentdatabase, and other services you need.The virtual environment is user friendly and easy to use for developing any application[6].

3.1.2Android SDK for Android Studio.

The Android Software Development Kit (SDK) is a crucial part of Android development for beginners to come to grips with. It's a selection of files bundled together that you will need to begin creating Android apps. It consists of tools like the virtual device manager (emulator) and ADB bridge, as well as a library of additional code for making Java programs work with the Android platform.The AWS Mobile SDK for Android helpsdevelopers develop mobile apps in Android Studio by providing simplified APIs for using AWS services, such as

Amazon Cognito and Amazon DynamoDB, Amazon S3 Bucket and various other services[7].

3.1.3 Identity pool and user pool in amazon Cognito

A user pool is a user directory in Amazon Cognito. With a user pool, your users can sign in to your web or mobile app through Amazon Cognito. Security features such as multi-factor authentication (MFA), checks for compromised credentials, account takeover protection, and phone and email verification.Amazon Cognitoidentity pools provide temporary AWS credentials for users who are guests (unauthenticated) and for users who have been authenticated and received a token. An identity pool is a store of user identity data specific to your account.User pools are for authentication (identify verification) but Identity pools are for authorization (access control).To connect with our android application we can use can find Identity pool ID for example:https://eu-south1.console.aws.amazon.com/cognito/home?region=eu-south-1 and create a Federated Identity[8].

3.1.4 Connection of s3 bucket to AWS rekognition

Amazon Rekognition Image can analyse images that are stored in an Amazon S3 bucket or images that are supplied as image bytes. It use the Detect Labels API operation to detect objects, concepts, and scenes in an image (JPEG or PNG) that's stored in an Amazon S3 bucket. You pass an image to an Amazon Rekognition Image API operation by using the Image input parameter. Within image, you specify the S3 object property to reference an image stored in an S3 bucket. Image bytes for images stored in Amazon S3 buckets don't need to be base64 encoded. One of the important thing, the region for the S3 bucket containing the image must match the region you use for Amazon Rekognition Image operations.

3.1.5 Image analysis

The captured images must be proper .The image should not be blurred. DetectText detects text in .jpeg or .png format images. You can use the Amazon Rekognition Image API to detect text in Image asynchronously with Start Text Detection and Get Text Detection. Both image text detection APIs support most fonts, including highly stylized ones. After detecting text, Amazon Rekognition creates a representation of detected words and lines of text, shows the relationship between them, and tells you where the text is on an image.The requirement for images are The blue boxes represent information about the detected text and the location of the text that's returned by the DetectText operation (on an image) or the GetTextDetection operation (on a single frame of a video). To be detected, text must be within +/- 90 degrees orientation of the horizontal axis[9].

Fig.6

Fig.7

Fig.8

Fig.9

In Fig.6 the container image is clean and appropriate to scan and Extract the image information.in Fig.7the image is proper but the brightness is high and lighten image that is why the image is not cleared. Fig.8 and Fig.9 are blur so the text is not extracted from image. If we compare the Fig.8 and Fig.9 ,Fig.9 is more blur than Fig.8 and thus no proper text is extracted. The extracted will be like,

If we giveFig.10 as an input

Fig.10

Aws recognition gives output in Jason script as shown below:

```
{
  "TextDetections": [
    {
      "DetectedText": "ALLU 521653 5",
      "Type": "LINE",
      "Id": 0,
      "Confidence": 99.60054779052734,
      "Geometry": {
        "BoundingBox": {
          "Width": 0.20173299312591553,
          "Height": 0.03854028508067131,
          "Left": 0.589670717716217,
          "Top": 0.2697930335998535
        },
        "Polygon": [
          {
            "X": 0.589670717716217,
            "Y": 0.2697930335998535
          },
          {
            "X": 0.7914037108421326,
            "Y": 0.26818278431892395
          },
          {
            "X": 0.7915767431259155,
            "Y": 0.30672305822372437
          },
          {
```


Fig.11

It is a huge code with complete dimensional details and text[12].

By using python we can get only required data as shown below:

```
Detected text:ALLU 521653 5
Detected text:22G1
Detected text:MAX.GR 67.200 LBS
Detected text:30 .480 KGS
Detected text:TARE .000 KGS
Detected text:28 4.410 480 KGS LBS
Detected text:NET 790 LBS
Detected text:9-
Detected text:9- CU.CAP. 33.2 CU.M
Detected text:1.172 CU.FT
Detected text:ALLU
Detected text:521653
Detected text:5
Detected text:22G1
Detected text:MAX.GR
Detected text:30
Detected text:.480
Detected text:KGS
Detected text:67.200
Detected text:LBS
Detected text:TARE
Detected text:.000
Detected text:KGS
Detected text:4.410
Detected text:LBS
Detected text:NET
Detected text:28
Detected text:480
Detected text:KGS
Detected text:790
Detected text:LBS
Detected text:CU.CAP.
Detected text:33.2
Detected text:CU.M
Detected text:9-
Detected text:9-
Detected text:1.172
Detected text:CU.FT
Text detected: 38
```

Fig.13

This detected text further will be sent to user as .csvOr .txt file. From there user can edit ,copy, search accordingly.

4.Conclusion

In this paper we discussed different methods used for implementing the task of extracting text from images and briefly described our methodology for same. Using AWS Rekognition for extracting text from image makes implementing machine learning concepts so easy for developers who have very less
knowledge about it. Also Cloud services makes it easy to access all the technology to implement the idea with minimum cost and without worrying about infrastructure and platform.

Acknowledgement

We would like to express our heartful gratitude to Mr. Sachin Jain for giving us the opportunity for developing such application for industry use and to Dr. R. D. Wajgi for their guidance and help us gain immensely enriching professional experience and sincere thanks for giving us valuable inputs and ideas. We thank Dr. G. M. Dhopavkar(HOD computer department) for her ongoing support and encouragement in every aspect.

References

[1]Automated Text Extraction from Images using OCR System
https://ieeexplore.ieee.org/document/8991281

[2]Text recognition from images
https://ieeexplore.ieee.org/document/7193210
P. M. Manwatkar and S. H. Yadav, "Text recognition from images," 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2015, pp. 1-6, doi: 10.1109/ICIIECS.2015.7193210.

[3]A robust algorithm for text extraction from images
https://ieeexplore.ieee.org/document/7760928
W. H. A. Nimasha, L. Ranathunge, B. R. Jayawickrama, K. L. Mahaliyanaarachchi and L. G. B. Subhagya, "A Robust Algorithm for Text Extraction from Signage Images," 2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer), Colombo, Sri Lanka, 2018, pp. 33-40, doi: 10.1109/ICTER.2018.8615528.

[4]Research on the Text Detection and Extraction from Complex Images
https://ieeexplore.ieee.org/document/6631707
J. Zhang, R. Cheng, K. Wang and H. Zhao, "Research on the Text Detection and Extraction from Complex Images," 2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies, Xi'an, 2013, pp. 708-713, doi: 10.1109/EIDWT.2013.122.

[5]Cloud based Text extraction using Google Cloud Vison for Visually Impaired applications
https://ieeexplore.ieee.org/document/9087314
D. Vaithiyanathan and M. Muniraj, "Cloud based Text extraction using Google Cloud Vison for Visually Impaired applications," 2019 11th International Conference on Advanced Computing (ICoAC), Chennai, India, 2019, pp. 90-96, doi: 10.1109/ICoAC48765.2019.246822.

[6]https://aws.amazon.com/

[7]https://en.wikipedia.org/wiki/Software_development_kit#:~:text=A%20software%20development%20kit%20(SDK,platform%20and%20operating%20system%20combination.

[8]https://docs.aws.amazon.com/cognito/latest/developerguide/identity-pools.html

[9]https://docs.aws.amazon.com/rekognition/latest/dg/text-detection.html

[10]https://docs.aws.amazon.com/rekognition/?id=docs_gateway

[11]https://docs.aws.amazon.com/rekognition/latest/dg/what-is.html

[12]https://ap-south-1.console.aws.amazon.com/rekognition/home?region=ap-south-1#/text-detection