Image recognition system

Swati Dewangan
dewangan.swati98@gmail.com
Bhilai Institute of Technology, Raipur, Chhattisgarh

Sneha Dey
sneha.dey349@gmail.com
Bhilai Institute of Technology, Raipur, Chhattisgarh

Megha Sahu
meghasahu77@gmail.com
Bhilai Institute of Technology, Raipur, Chhattisgarh

Seira Tak
seira.tak@bitraipur.ac.in
Bhilai Institute of Technology, Raipur, Chhattisgarh

ABSTRACT

Image recognition has been an active and challenging area of research. In recent years image recognition has received veritable attention from the research communities but it still remained very challenging in real applications. It plays a very important role in today's world. Though we might have seen various applications and websites of text recognition, label detection and face detection they are all separate applications this review paper tells us how to combine them together so that user don't need to download numerous software but just a simple image recognition system that uses Google cloud vision API. In image recognition, we have combined 3 things together that is text, label, and face. It should be noted that the combination of these three things together is nowhere found and might be difficult to evolve. This review paper will focus on scattering and separating the image by defining each element of the image by recognizing the text, detecting its label and detecting the face that can make the human work's easy. By focusing on the history and evolution of the text, label and face recognition system we combine them together so that it can be used for security purposes, organization, computer software and it can also be used in the government field. This review paper will represent the limitations and superiorities of different techniques which are used for image recognition system.

Keywords— *Google cloud vision API, PHP, HTML, Text recognition, Label detection, Face recognition*

1. INTRODUCTION
Image recognition is that the ability of a laptop to receive and interpret pictures which may be inputted from multiple sources like photographs and bit screen devices. Recognition of handwritten, machine characters, label, face and emotions of the face are an emerging area of research and finds extensive applications in banks, offices and industries etc.

There are 3 types of recognition:
1.1 Text recognition
Such software's are useful when we want to convert our Hard copies into soft copies. Such software's reduces almost 90% of the conversion work while still some verification is always required [1]. Image recognition, involves computer software designed to translate images of typewritten text (usually captured by a scanner), digits and hand written characters and digits into machine-editable text. The image recognition began as a field of research in artificial intelligence and machine vision. Recognition of text and identification of characters in scene images is a challenging visual recognition problem. As in much of computer vision, the challenges posed by the complexity of these images have been combated with hand designed features [2], [3], [4].

1.2 Label detection
The main advances in label detection were achieved thanks to improvements in object representations and machine learning models**.** Specially, this paper takes label detection as a multi-label super pixel labeling problem by minimizing an energy function [5]. In label detection the objects are labeled by their name like- In any image if there are number of objects like cat, table, books, etc. then image recognition system will detect and list out the all object's name present in the image.

1.3 Face Recognition
Face recognition is an important topic in the fields of computer vision and artificial intelligence owing to its significant academic and commercial potential. Facial emotions and expressions of human beings are the important factors in the communication that help us understand the state of mind of others. In general, people infer the emotional states of other people, such as joy, sadness,

surprise and anger, using facial expressions [6]. It can also recognize the multiple faces present in the image. Even this software also tell us that the person in the image has wear something on head or not, and it also give information whether the image is blurred (quality).

2. METHODOLOGY

Image recognition system falls under the AI and machine learning model, in which we have used Google cloud vision API for these three recognition facilities. Once users uses the image recognition system firstly they need to select a image from the system and there will be independency use image in any kind of format (like JPEG/JPG, TIF, PNG, GIF etc.), but there is a limitation on the uploading time i.e. if any image takes more than 30 seconds to upload on the software then after 30 seconds it will show the fatal error on the page. Images that upload on the image recognition software are stored in the cloud storage by vision API. After that the main frame of the software are categorized in to 3 different tabs, that are text detection, label detection and face detection.

In the text detection tab, it will show all the texts and digits which are present in the image and it also detect the hand written characters in image. The main feature of this software is that, it can detect the smallest text which cannot be seen with the naked eye. There is another feature i.e. a button on the bottom that is" SELECT ALL" button. This button helps us to copy the text which comes from the image without using the copy/paste option on another place on the screen. But, it cannot recognize the text and characters which is not in the Google font library, cursive or calligraphic writings or in another language (like Hindi, Urdu, and Chinese etc.) except English.

In label detection tab, after the process it will show the name of all the objects present in the image. The main property of this recognition process is, it will also show the smallest things present in the image. If in the image there is more than one person available in the uploaded image then it will define the relationship between two (or more than two) people (relationship like – friendship, love, fight etc.). Above the object, this software can also detect the profession of person which is present in the image.

In face recognition tab, mainly it recognizes the emotions of the person who is in the image. It shows the possibilities of the feelings of the person through image recognition. In this we categorized all the possibilities in 4 words that are: Very_Likely > Likely > Unlikely > Very_Unlikely. In the form of the mathematical percentage form very likely is equal to 100 percentage, likely is equal to 66.67 percentage, unlikely is equal to 33.33 and very unlikely is equal to 0 percentage. This shows all the expressions in these 4 categories. If in any image, there is a girl who's smiling then our project will show the joy is Very_Likely which means she is very happy. In face detection there are specified categories of the feelings of the human like joy, sorrow, angry, surprise. It can also recognize multiple faces at a time and show each face emotions separately (face 1, face2, face 3….face n). If in any image a person is wearing a cap or wig on head then this software will tell us that a person who is in the image has worn the cap (or something) on head that is, head wear: cap. It also shows us the quality of the picture that how much the image is blurred or clear.

2.1 API working process

To access Google APIs using the official client SDKs, create a service object based on the API's discovery document, which describes the API to the SDK. It will need to fetch it from the Vision API's discovery service. Through the image recognition system an image is uploaded to the main frame of the recognition system then it is automatically uploaded to the Cloud Storage with text, object and face in any image (text that appears in the image itself) [7]. A Cloud Function is triggered, which uses the Vision API to extract the text, object, face and queues the text to be translated into the configured results. For each queued translation, a Cloud Function is triggered which uses the Translation API to translate the text and queue it to be saved to Cloud Storage. For each translated text, a Cloud Function is triggered which saves the translated text to Cloud Storage [7].

For face and label detection, the device takes soft copy of image and extract result by using Google cloud platform. The coding of API is done on the python language by the Google services. To use this API we have to use another language for connectivity purpose like Android, PHP, and Python etc. to set a base or a platform to run API.
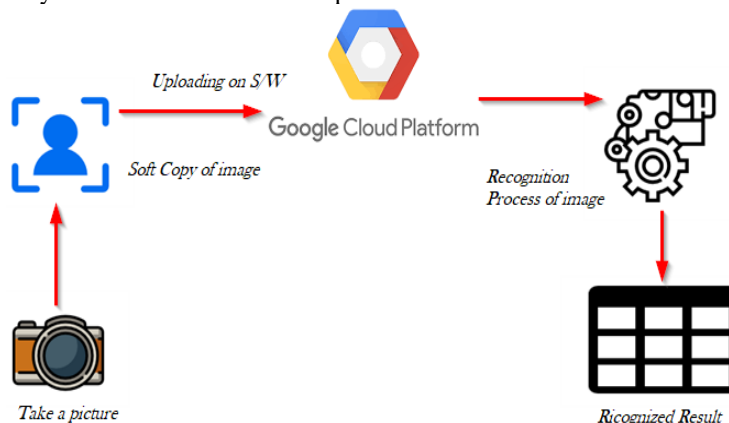
Fig. 1: Working process of the Google cloud vision API for image recognition [7]

2.2 Front end and back end

For the front end we use HTML5 language. HTML is the most user friendly language for front end works and this is mainly used for GUI (Graphical User Interface). For back end of the image recognition system, we have used PHP language. As a local server, we have used XAMPP apache server, and connection of network is necessary to run API with the XAMPP.

3. MODEL DESIGN

Using AI (Artificial Intelligence) and Machine learning through Google Could vision API, we have designed a very useful and updated software. Suggestions for acquiring user needs is to apply disciplines such as user friendly, time saving as well as looking at past successful design. Also, constant testing and improvement of the design by appealing to the actual users were considered to be crucial. User centered design was the focus because the general design principles may not work on small screen size. To comprehend the user philosophically, and using various methods and activities to find out what is significant for the users. Some of the methods included are creating user framework, online inquiry and user self-reporting techniques. The most effective way was to get user response on design through a prototype. There were few design instruction that stood out. The authors argued that well-designed outline are a powerful feature that can overcome restrictions of small screen size [8].

There are some screenshots of the Image Recognition System to have an overview and clear view of the software.
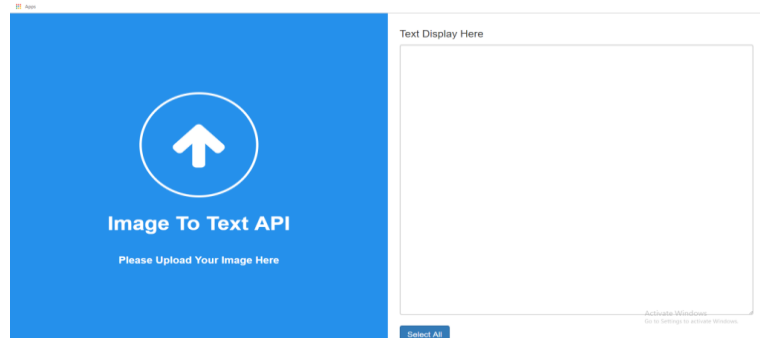
3.1 Main screen

Fig. 2: First page of the image recognition system
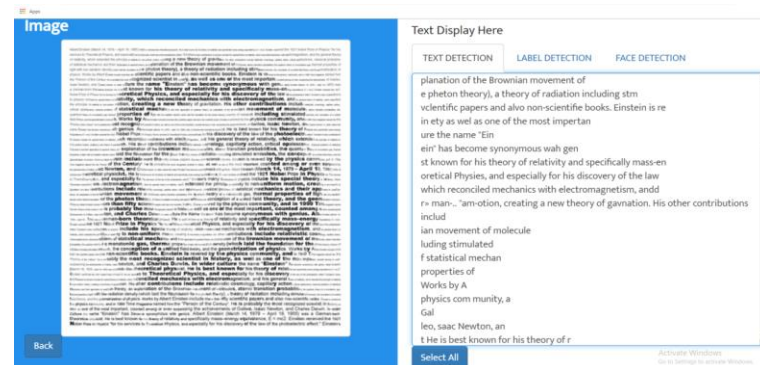
3.2 Text recognition

Fig. 3: Recognition of small text which is not seen by the naked eye
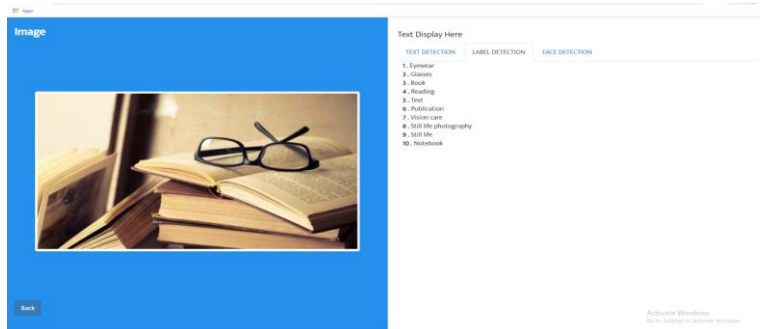
3.3 Label recognition

Fig. 4: Detection of the all object which is in the image
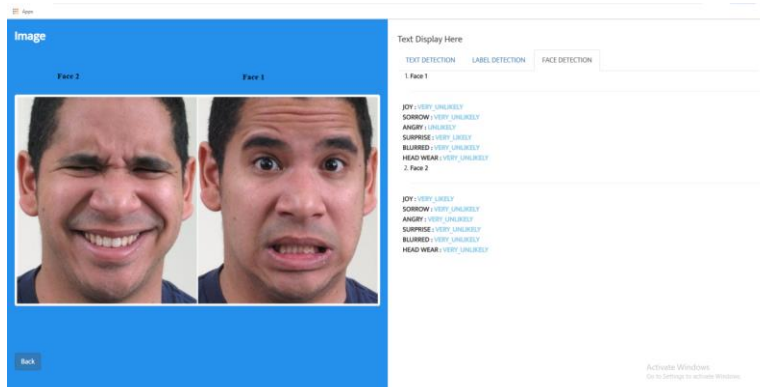
3.4 Face Recognition

Fig. 5: Multiple face recognition

4. CONCLUSION

There are various number of Recognition Systems but no one can recognize these three recognition (text, label and face) in a system simultaneously. The main role of Image Recognition system is that it is serves all things in a single system and even it has various features for enhancing this system.

It totally depends on the user needs and requirements in different fields like banks, industries, Government, security purposes.

5. REFERENCES

[1] Simmon Tanner, "Deciding whether Optical Character Recognition is Feasible".
[2] T. E. de Campos, B. R. Babu, and M. Varma, "Character recognition in natural images," in Proceedings of the International Conference on Computer Vision Theory and Applications, Lisbon, Portugal, February 2009.
[3] M. Yokobayashi and T. Wakahara, "Binarization and recognition of degraded characters employing a most disconnection axis in color area and handgun correlation," in International Conference on Pattern Recognition, vol. 2, 2006, pp. 885– 888.
[4] J. J. Weinman, "Typographical options for scene text recognition," in Proc. IAPR International Conference on Pattern Recognition, Aug. 2010, pp. 3987–3990.
[5] [5] Junjie Yan; Yinan Yu; Xiangyu Zhu; Zhen Lei; Stan Z. Li, "Object detection by labeling superpixels".
[6] Byoung Chul Ko, "A Brief Review of Facial Emotion Recognition Based on Visual Information", Department of Computer Engineering, Keimyung University, Daegu 42601, Korea; Published: 30 January 2018.
[7] Google https://cloud.google.com/vision/docs/quickstart-client-libraries ; 2009
[8] Jones, M. and marsden, G. 2006. Mobile interaction design. John Wiley & Sons, Ltd.