International Journal of Technical Research and Applications e-ISSN: 2320-8163,
www.ijtra.com Volume 4, Issue 2 (March-April, 2016), PP. 125-129

OCR WITH LANGUAGE TRANSLATOR

Happy Jain[1], Akshata Choudhari[2], Mohan Sharma[3], Jagdeep Yadav[4], Krunal J. Pimple[5]
[1,2,3,4] UG Students, [5]Professor
Department of Information Technology
Theem College Of Engineering, Boisar
Mumbai University,
Maharashtra-India
[1]happyjain02044@gmail.com

***Abstract*–part of our day to day life. With the revolution in mobile computing numerous extraordinary components were added to the field and the mobiles got smaller, faster and better as the decade Today mobile and mobile based applications have become a passed. It gave rise to the introduction of new mobile based operating systems frameworks where the software engineers where given open source working system named "ANDROID". (OCR)[1] is a powerful tool for bringing data from our simple lives into the inexorably digital world. This technology has long seen use in building digital libraries, recognizing text from natural scenes, understanding hand-written office forms etc.[2] Ocrhepls to take images from android mobile and translate it in any language that user wish to translate. Even user can take image from camera and translate the text or image. User can also take the text or copy text from the mobile or where ever it is saved, it also allows to copy the text to different apps(message ,text, email) or any other clipboard.**

***Keywords*-(OCR),Android, OCR engine, Bing Translator API, EI Plug in, Chopping, Android Tools, Android SDK, Android NDK, Tess-two, Eclipse, ADT plug in, El Google, ABBY mobile OCR engine.**

I.INTRODUCTION

Outline is the initial step into the improvement stage for any built item or framework. Outline is an imaginative procedure. A decent plan is the way to successful framework. The expression "configuration" is characterized as "the procedure of applying different methods and standards with the end goal of characterizing a procedure or a framework in adequate subtle element to allow its physical acknowledgment". It might be characterized as a procedure of applying different methods and standards with the end goal of characterizing a gadget, a procedure or a framework in adequate point of interest to allow its physical acknowledgment. Programming plan sits at the specialized part of the product building prepare and is connected paying little mind to the improvement worldview that is utilized. The framework plan adds to the design point of interest required to construct a framework or applications, research in optical character acknowledgment, an innovation very much created for filtered reports, is moving center to the acknowledgment of content implanted in advanced photos.

Android is an item stack for PDAs that fuses a working framework, middleware and key applications. The Android SDK gives the instruments and APIs important to start developing.[3]document to the programmers or database personnel. Framework plan experiences two periods of improvement: Logical and Physical Design. The traditional system is system is user dependent such as the user has to manually perform all the tasks needed by him. Only one application can be used at one time. User cannot keep log of the work done by him. User cannot maintain a common database for all applications. Maps have to be configured manually. Messages have to be read manually. Web handling has to be done by using a browser and the information needed has to be searched manually. No application in android market which convert Image into text and searches information about that text via Google Services and API's. Application framework enabling reuse and replacement of components. The project is based on the need to find particular place and its information through image. Dalvik virtual machine optimized for mobile devices. Application Works on Ginger Bread and higher version of Android Operating System. Application Is User Friendly. Design and Look of application is so good. It can support any format of image i.e. .Jpeg, .PNG etc to search related information about image.[4]

II.ANDROID

With the introduction of android the programmers were free to program freely and with the much awaited programming language as the programmers did not have to learn anything new. The android was a classic mixture of java and mobile computing. With this there came many applications which some were free and some were paid. But all the applications did not have the applications needed in combine.

The cutting edge open working frameworks are not on desktops or centralized servers but rather on the little cell phones individuals convey each day.The openness of these new situations prompts new applications and showcases and empowers more noteworthy incorporation. As the interest develops for cell telephone applications, research in optical

character acknowledgment, an innovation very much created for examined reports, is moving center to the acknowledgment of content implanted in advanced photos.

A. FEATURES

- Application structure empowering reuse and substitution of segments
- Dalvik virtual machine enhanced for cell phones
- Integrated program in view of the open source Web pack motor Optimized illustrations controlled by a custom 2D representation library; 3D design in light of the OpenGL ES 1.0 particular (equipment quickening discretionary)
- SQLite for organized information stockpiling
- Media support for regular sound, video, and still picture groups (MPEG4, H.264, MP3, AAC, AMR, JPG, PNG, GIF)
- GSM Telephony (equipment subordinate)
- Bluetooth, EDGE, 3G, and WiFi (equipment subordinate)
- Camera, GPS, compass, and accelerometer (equipment subordinate)
- Rich change environment including a contraption emulator, instruments for examining, memory and execution profiling, and a module for the Eclipse IDE

III.THEORY

OCR technology allows the conversion of scanned images of printed text into text or information that can be understood or edited using android mobile phones. OCR is abbreviated as Optical Character Recognisition, it translate the electronic or mechanic text and even translate the handwritten, typed text or printed into machine code. Open source OCR programming called Tesseract as a premise for Optical Recognition venture, which is considered as the most exact free OCR motor in presence. OCR innovation utilizes three stages Scanning securing of printed reports as optical pictures. Acknowledgment includes changing over these pictures to character streams speaking to letters of perceived words and the last component includes getting to or putting away the changed over content. Converted text is referred as extracted text. User captures the text from the image of mobile camera that is needed to be translated and optimised so that the OCR can be performed on it. Speech synthesizer is used to convert extracted text into the voice. Firstly it investigations content, changes content into pronounceable structure. Discourse synthesizer performs change of grapheme to phoneme shape and uses voice qualities of a man. The majority of the character acknowledgment frameworks will be perceived through the info picture with a scanner and PC programming. There is an issue in the extent of the PC and scanner, as PC and scanner requires expansive measure of space. With a specific end goal to conquer this issue of PC and scanner possessing an expansive space, optical character acknowledgment (OCR) framework in view of android telephone is proposed. Since the exhibitions of advanced mobile phone and PC are distinctive, the pace of character acknowledgment is moderate. In this paper, the character acknowledgment strategy is exhibited by utilizing OCR innovation and android telephone with higher quality camera.

A. Optical Character Recognition (OCR)

Optical Character Recognition (OCR) is broadly utilized innovation which changes over examined pictures of printed content, manually written content characters into machine encoded content data, for example, ASCII[5].Number plate acknowledgment is a type of programmed vehicle distinguishing proof. A number plate is the interesting distinguishing proof of vehicle. It is a picture handling innovation used to distinguish vehicles by their own particular number plates. Continuous number plate acknowledgment assumes an imperative part in keeping up law implementation and keeping up activity rules. Programmed number plate acknowledgment has three noteworthy parts: vehicle number plate extraction, character division and Optical Character Recognition (OCR). Number plate extraction is that stage where vehicle number plate is identified. The identified number plate is pre-prepared to uproot the commotion and after that the outcome is gone to the division part portion the separately characters from the extricated number plate. The divided characters are standardized and went to an OCR calculation. Finally the optical character data will be changed over into encoded content. The characters are perceived utilizing Template coordinating. The last yield must be through string of characters.[1]

B. Number Plate Extraction

Image that is captured from the mobile is in RGB format, it is than converted to gray scale image and than to binary image for regonization.

C. Character Segmentation

The character division part facilitate sections the character exclusively from the separated number plate. From info picture the primary procedure will be to trim out the number plate characters from beginning to the consummation point leaving all the additional wide spaces from top to beneath and from right to left as it may be. Characters are just as fit in the plate locale. For simple correlation of the information character with the character in the information base the outcome is standardized into the character set as the extent of the pictures in the database.

D. Optical Character Recognition

The optical character recognition is a recognition method in which the input is an image and the output is string of character. OCR is a process which separates the different characters from each other taken from an image. Template matching is one of the approaches of OCR. The cropped image

is compared with the template data stored in database. OCR automatically identifies and recognizes the characters without any indirect input. The characters on the number plate have uniform fonts then the OCR for number plate recognition is less complex as compared to other methods.

E. COMPONENTS OF OCR

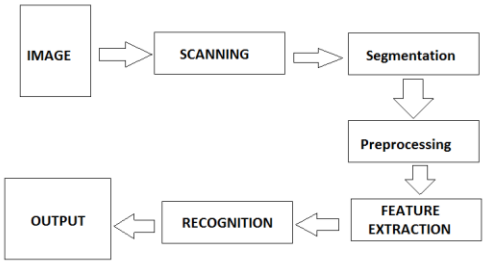

Fig 1:Components Of OCR

F. Hardware Module:

- P IV processor and higher
- 512 MB and higher
- Fast Internet Connection
- Smart phone with Android OS
- Internet Connection for PC

G. Software Module:

- Windows Xp or higher
- JDK 1.6 or higher
- Google map API.
- Android SDK tools.
- Android SDK platform tools.
- Android 3.2
- Indigo Eclipse.
- ADT-0.9.5

IV. ABBY MOBILE OCR ENGINE

The ABBYY Mobile OCR Engine depends on minimized code OCR innovation and is upgraded to work with gadgets that have little memory sizes – including cell phones, tablets and versatile scanners.[6] Features include:

- Upgraded memory administration. Another calculation for memory administration empowers the product to decide the precise memory size required to preparea picture. This takes out the need to distribute huge memory sections ahead of time, which can affect acknowledgment velocity and unwavering quality – guaranteeing productivity and quick execution speed.
- The Engine's code is exceptionally conservative. It possesses as maker as 8 MB of ROM and 10 MB of RAM contingent upon sought usefulness

Versatile programming improvement organizations and makers of cell phones and tablets can empower gadgets to peruse, perceive and prepare the content inside of photographs. Presently menus, signs, business cards, bills, notes and more can be captured and changed over into content for data into a wide assortment of applications.[7] Make an interpretation of the Recognized content into one of English,Spanish,Afrkaans,Catalan,Galacian,Esperanto,Occtian, Portuguese,French,Danish,Dutch and Aragonese.[5]

V. SNAPSHOT

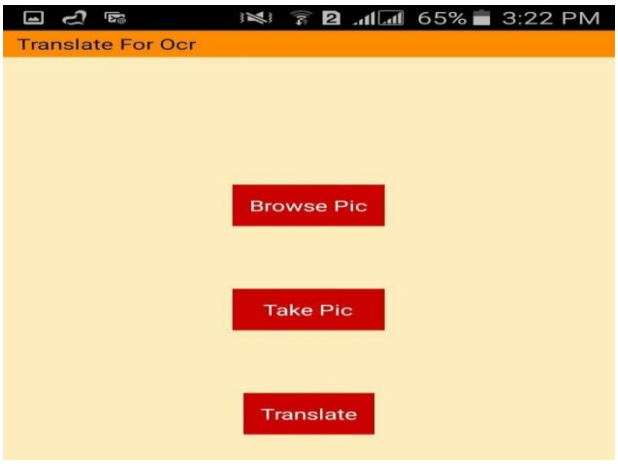

Fig 2: OCR Home Screen GUI

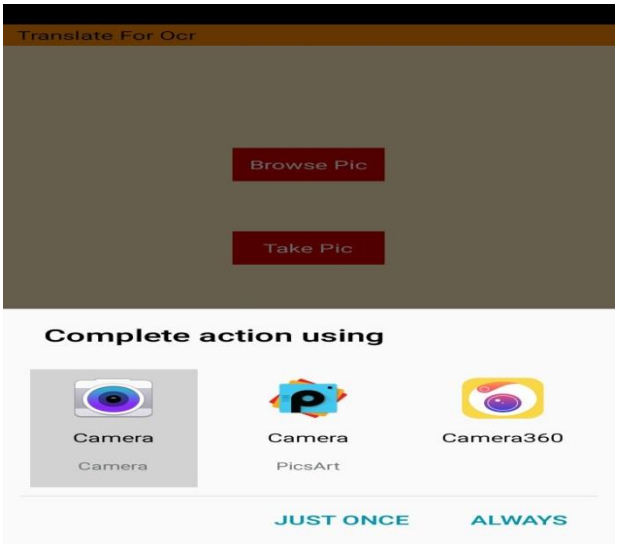

Fig 3: Capture Image Using Camera

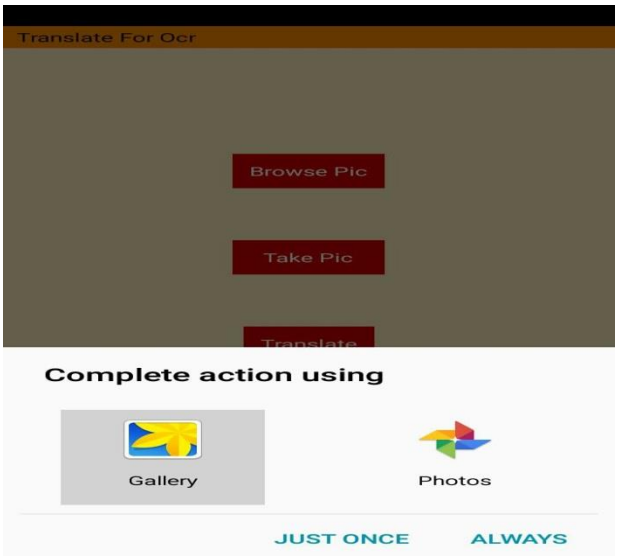

Fig 4: Browse Image From Storage

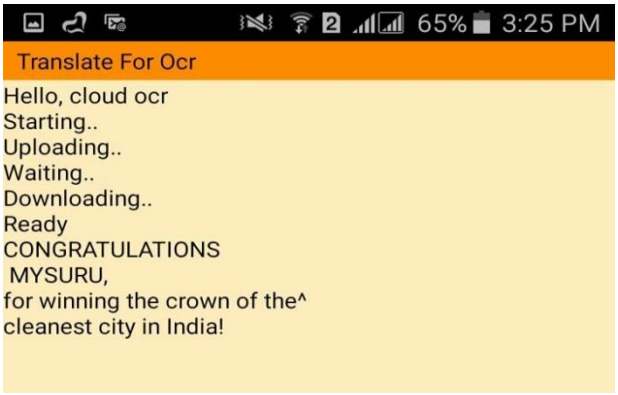

Fig 5: OCR Image Scanned

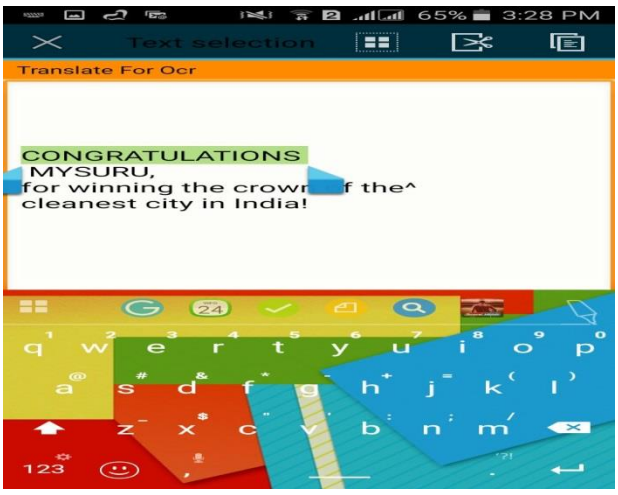

Fig 6: Copy the word to Translate

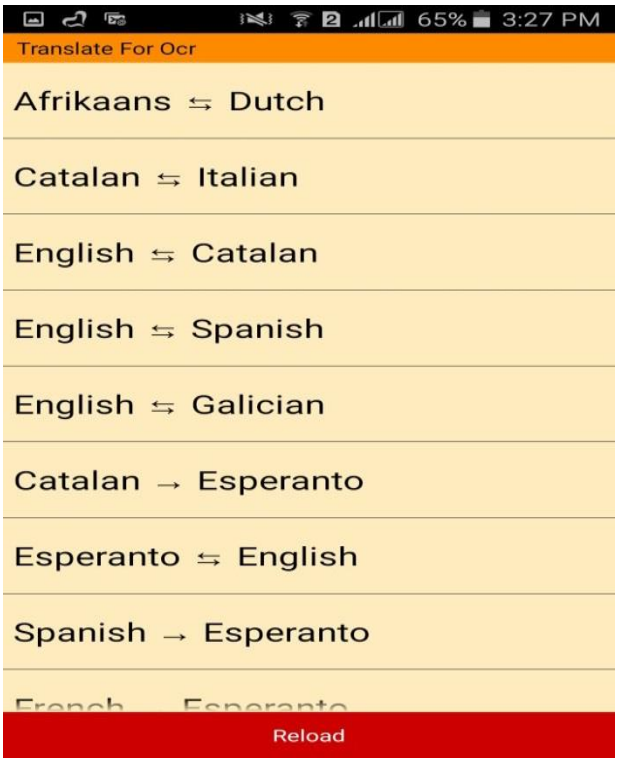

Fig 7: Choose the Language for translation

Fig 8: Download the package of Translation from English to Spanish

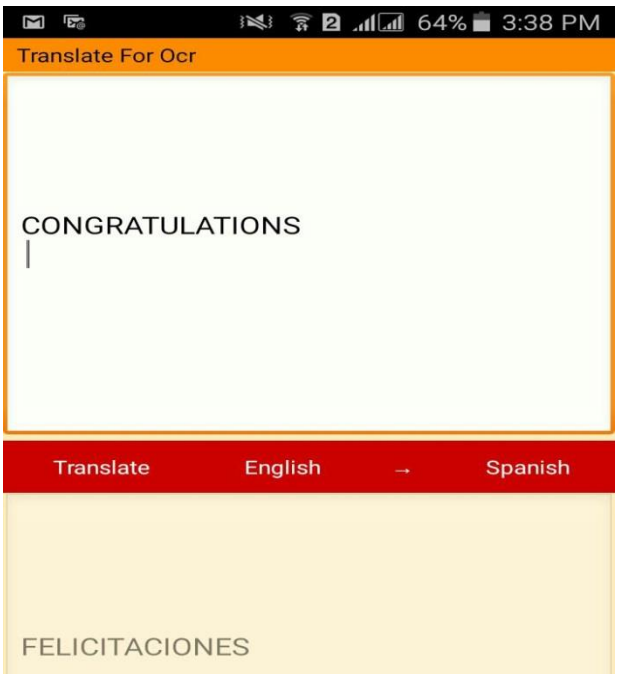

Fig 9: Translated output

VI. CONCLUSION

The Optical Character Recognition deals with recognition of optically processed characters. Reliably interpreting text from real-world photos is a challenging problem due to variations in environmental factors even it becomes easier using the best open source OCR engine. This paper introduce Android Mobile OCR Engine which provides 90% of accuracy in converting image to text.

REFERENCES

[1]Ravina Mithe, Supriya Indulkar, Nilam Divekar "Optical Character Recognition", International Journal Of Recent Technology And Engineering (Ljrte) Issn: 2277-3878, Volume-2, Issue-1, March 2013.

[2] Carlos A.B. Mello And Rafael D. Lins "Image Segmentation Of Historical Documents", Department Of Electronic And System-Ufpe- Brazil.

[3] Surjya Ray, Ilker Deminkol And Wendi Herinzelman "Atma:- Advertisement_BasedTdma Protocol For Bursty Traffic In Wireless Sensor Networks", Department Of Computer And Electrical Engineering University Of Rochester, Rochester, New York.

[4] Nitin Mishra & C.Patvardhan"Atma: Andriod Travel Mate Application", International Journal Of Computer Application(0975-8887) Volume 50-No.16, July 2012.

[5] Amit Dhanwani And Ruhi Bajaj. "World Tourister: An Andriod Application As Travel Mate", Proceeding Of 9th Irf International Conference, Pune, India,18th May 2014, Isbn:978-93-84209-20-9

[] Xilin Chen, Member, Ieee, Jie Yang, Member, Ieee, Jing Zhang And Alex Waibel, Member,Ieee
"Automatic Detection And Recognition Of Sign From Natural Scenes", Ieee Transactional Of Image Processing, Vol.13,No.1, January 2004.

[6] ABBY Mobile OCR Engine provided by Android SDKhttp://www.abbyy.com/mobile-ocr/android-ocr/

[7] ABBY Android Mobile OCR Engine key features http://www.abbyy.com/mobile-ocr/features/