

Extraction of Text and Document Image For Vehicle Number Plate Using Image Processing.

Navdeep Kaur¹, Er. Sumit Chopra²

¹*KC College of Engineering and Technology, Nawanshahr, Punjab, India.*

¹*(navmahey@gmail.com)*

²*(sumitonhunt@gmail.com)*

Abstract— Vehicle number plate acknowledgment is the most fascinating and testing research theme from recent years. It is demonstrated that the number plates are diverse shape and measure and furthermore have distinctive shading in various nations. In India the most widely recognized vehicle number plate utilized yellow or white as foundation and dark utilized as closer view shading. In this paper we proposed a framework to restriction of number plate chiefly for the vehicles in West Bengal (India) and portioned the numbers as to distinguish each number independently. This exhibits a methodology dependent on basic and effective morphological task and sobel edge discovery technique. We likewise present a basic way to deal with fragmented every one of the letters and numbers utilized in the number plate. Subsequent to decreasing clamor from the info image we attempt to improve the difference of the binarized image utilizing histogram adjustment. We for the most part focus on two stages; one is to find the number plate and second is to section all the number and letters to distinguish each number independently.

Keywords—*Image Processing, Text Extraction, Character recognition.*

I. INTRODUCTION

These days all over digitization innovation is utilized. Text Acknowledgment normally truncated to OCR technique, includes a framework intended to decipher images of handwritten, printed text into computer editable text or to make an interpretation of images of characters, numbers into a standard encoding plan speaking to them. OCR (Optical Character recognition) started as a field of research in computerized reasoning and computational vision. Text Acknowledgment utilized in authority assignment in which the huge information need to type like post workplaces, banks, universities and so on, all things considered, applications where we need together some data from text composed image. Individuals wish to check in an archive and have the text of that record accessible in a .txt or .docx arrange. The point of Optical Character Acknowledgment is to group optical examples (regularly contained in a computerized image)

relating to alphanumeric or different characters. The procedure of OCR includes a few stages including division, highlight extraction, and arrangement. On a fundamental level, any standard OCR programming would now be able to be utilized to perceive the text in the portioned casings. Notwithstanding, a hard take a gander at the properties of the hopeful character locals in the sectioned casings or image uncovers that most OCR programming bundles will have critical trouble to perceive the text. Archive images are not quite the same as characteristic images since they contain primarily message with a couple of designs and images. Because of the plain low-goals of images of those caught utilizing handheld gadgets, it is difficult to remove the total format structure (intelligent or physical) of the reports and much more dreadful to apply standard OCR frameworks. Thus, a shallow portrayal of the low-goals caught report images is proposed. If there should be an occurrence of unique electronic records in the archive, the extraction of a similar mark is clear; the PDF or PowerPoint type of the first electronic reports is changed over into a moderately high-goals image (TIFF, JPEG, and so on.) on which the mark is figured. At last, the caught report's mark is contrasted with all the first electronic records' marks so as to discover a match.

II. RELATED WORK

Preprocessing is the initial phase in the handling of filtered image. The examined image is checked for clamor, skew, incline and so on. There are conceivable output of image getting skewed with either left/right introduction or with commotion, for example, Gaussian. Here the image is first believer into grayscale and after that into parallel. Thus we get image which is reasonable for further handling.

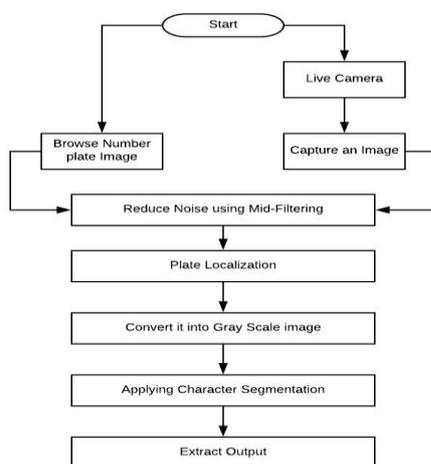
After pre-preparing, the clamor free image is passed to the division stage, where the image is disintegrated into individual characters. The binaries image is checked for bury line spaces. In the occasion that bury line spaces are recognized, the picture is divided into sets of areas over the interline opening. The lines in the sections are filtered for level space convergence regarding the

foundation. Histogram of the image is utilized to distinguish the width of the level lines. At that point the lines are examined vertical space crossing point. Histograms are used to recognize the width of the characters. By then the words are weakened into characters using character width count.

Highlight extraction pursues the division period of OCR where the individual image glyph is considered and separated for highlights. Initial a character glyph is characterized by the accompanying qualities like stature of the character, width of the character. Grouping is finished utilizing the highlights extricated in the past advance, which relates to each character glyph. These highlights are breaking down utilizing the arrangement of standards and named as having a place with various classes. This grouping is summed up to such an extent that it works for single textual style type. The tallness of the character and the width of the character, different separation measurements are picked as the contender for characterization when strife happens. So also, the grouping rules are composed for different characters. This strategy is a nonexclusive one since it removes the state of the characters and need not be prepared. At the point when another glyph is given to this classifier square it extricates the highlights and thinks about the highlights according to the guidelines and after that perceives the character and marks it.

III. ALGORITHM

Number plate is an example with high varieties of differentiation. In the event that the number plate is fundamentally the same as foundation it is hard to recognize the area. Splendor and complexity is changes as light fall changes to it. The morphological tasks are utilized to extricate the complexity highlight inside the plate. The work is separated into a few sections:



Figure,1. (Flow chart of Number plate extraction)

- Input raw image
- Image binarization.
- Reduce noise using mid-filtering method
- Enhance contrast using histogram equalizer.
- Plate localization
- Character segmentation

IV. TEXT EXTRACTION

Text extraction and acknowledgment process contains five stages in particular text discovery, text confinement, text following, division or banalization, and character acknowledgment.

Text Detection: This stage takes image or video outlines input and chooses it contains text or not. It likewise distinguishes the text areas in image.

Text Localization: Text confinement blends the text areas to plan the text items and characterize the tight boundaries over the text articles.

Text Tracking: This stage is associated with video data in a manner of speaking. For the lucidity reason, content embedded in the video appears more than thirty progressive housings. Content after stage abuses this transient occasions of a comparable content thing in various consecutive edges. It might be used to revise the delayed consequences of content ID and limitation sort out. It is similarly used to quicken the content extraction process by not having any kind of effect the binarization and affirmation dare to each distinguished thing.

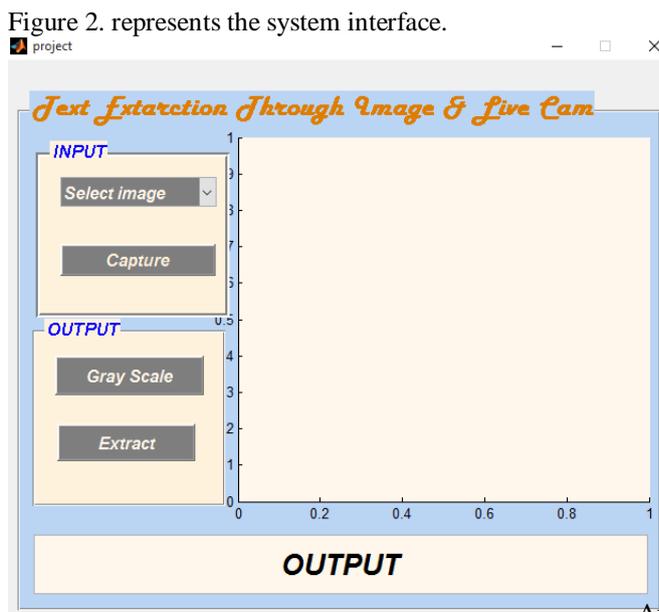
Text Binarization: This progression is used to fragment the text article from the foundation in the limited text items. The yield of text binarization is the paired image, where text pixels and foundation pixels show up in two diverse double dimensions.

Character Recognition: The last module of text extraction process is the character recognition. This module matches the shape of character with the shape of character or numbers images that is located in backend database.

V. RESULTS AND DISCUSSIONS

Text extraction or number plate extraction is the widely used system nowadays. We improve the output quality of this project and also introduce live image capturing function in our project. The output of this system is now up to more than 85%. As per our base paper the accuracy of output is just 65-70%. Sometimes the error has been occurred like system confuses when tow similar texts show like Z or 2. Because the shape of the character and word similar.

Figure 2. represents the system interface.



Figure,2. (Number Plate extraction GUI Interface).

Figure 2, shows the basic interface of our project. After that simply click on select image and click on browse and open the image in interface. We also use live cam option to take any image. Figure 3, shows the interface after selecting an image.



Figure,3. (Interface after select an image.)

After select the image convert it into gray scale image as per figure 4. Gray scale image is very useful to extract text from it.



Figure,4. (Gray Scale Image)

After conversion into gray scale simply click on extract button to view an output. The extracted text will be shown on the edit text block and as well as display text on notepad file.

VI. CONCLUSION

In this paper we proposed calculation for taking care of the issue of disconnected character acknowledgment. We had given the contribution to the type of images. The calculation is preparing on the preparation information that is first present in the database. We have done preprocessing and division and distinguish the line.

The paper exhibits a concise overview of the applications in different fields alongside experimentation into few chose fields. The proposed strategy is amazingly productive to extricate a wide range of bimodal images including obscure and light. The paper will go abase a decent writing study for scientists beginning to work in the field of optical character recognition (OCR).

VII. REFERENCES

- [1] Line Eikvil, "Optical Character Recognition", NorskRegnesentral, Oslo, Norway, Rep. 876, 1993.
- [2] M Usman Raza, et al., "Text Extraction Using Artificial Neural Networks", in Networked Computing and Advanced Information Management (NCM) 7th International Conference, Gyeongju, North Gyeongsang, 2011, pp. 134-137
- [3] Bertolami, Roman; Zimmermann, Matthias and Bunke, Horst, 'Rejection strategies for offline handwritten text line recognition', ACM Portal, Vol. 27, Issue. 16, December 2006
- [4] C.P. Sumathi, T. Santhanam, G.Gayathri Devi, "A Survey On Various Approaches Of text Extraction In Images", International Journal of

- Computer Science &Engineering Survey (IJCSES). Vol.3, August 2012, Page no. 27-42.
- [5] Datong Chen, Juergen Luetin, Kim Shearer, “A Survey of Text Detection and Recognition in Images andVideos”. Institute Dalle Molle d’Intelligence ArtificiellePerceptive Research Report, August 2000, Page no. 00-38.
- [6] Xu-Cheng Yin, Xuwang Yin, Kaizhu Huang, and Hong-Wei Hao, “Robust Text Detection in Natural Scene Images”, IEEE transaction on Pattern Analysis AndMachine Intelligence, 2013, Vol. 36, Page no. 970 – 983.
- [7] K. Wang, B. Babenko, and S. Belongie, “End-to-end scene text recognition”. International conference on computer vision ICCV 2011, vol. 10, Page no.1457 – 1464.
- [8] Xiaobing Wang, Yonghang Song, Yuanlin Zhang,“Natural scene text detection in multi-channel connected component segmentation”, 12th International conf. onDocument Analysis and Recognition, pp. 1375-1379, 2013.
- [9] Shehzad Muhammad Hanif, Lionel Prevost, “TextDetection and Localization in Complex Scene Images using Constrained AdaBoost Algorithm”,
- [10] 10thInternational Conference on Document Analysis and Recognition, pp.1-9, 2009 Lukas Neumann, Jiri Matas, “Real -Time Scene Text Localization and Recognition”, IEEE Conf. on Computer Vision and Pattern Recognition, 2012.

