Automatic Number Plate Recognition Based on Connected Components Analysis Technique

Sarmad Majeed Malik, and Rehan Hafiz

***Abstract*—**Automatic Number Plate Recognition (ANPR) is a technique employed in the recognition of license plates of vehicles. The objective is to design an efficient algorithm which will help in the identification of vehicles. ANPR has many uses ranging from automated Toll Plaza to security systems. The image processing technique may even be extended to robotics with the robot extracting the license plate. Different algorithms have been put forward but each has its own limitations. The goal is to design an efficient algorithm which has high efficiency. A camera first detects and captures the snap of the vehicle and then the algorithm searches for the license plate. The algorithm is divided into three parts: License plate region extraction, segmenting the characters and comparing the characters for recognition. The system is implemented on MATLAB. The results show that the algorithm successfully extracts and recognizes the number plate and has high performance.

***Keywords*—**Correlation, Optical Character Recognition (OCR), Template matching, Segmentation.

I. Introduction

AUTOMATIC Number Plate Recognition (ANPR) has gained much popularity in recent years. It is a technique aimed at the extraction and recognition of license plates using image processing techniques. The need of more secure, automated security systems, toll plaza, over speed detection systems have greatly elevated the need for an automated system which can detect the vehicles. The increase in the number of vehicles also requires a vehicle identification system. ANPR is the solution to these problems.

Connected components are becoming increasingly popular in image processing techniques. These are essentially a group of pixels in contact with one another. Any object in an image is made up of pixels. The most common image model is RGB model which consists of red, green and blue color components. The objects in the image may be identified and separated from others by means of connected components.

Various algorithms have been proposed for license plate detection. Each algorithm utilizes some features of the license plate. Smearing Algorithm for license plate extraction is proposed in [1]. The proposed algorithm works on yellow colored plates. Optical Character Recognition is used in the last step to recognize the characters. Morphological image processing techniques are utilized in [2]. Edge detection along with centre of gravity calculation of the license plate is used by calculating the length and width of the edges. Hidden Markov Model is used in plate extraction which eliminates the similarity between image samples. Smearing algorithm with horizontal and vertical scans for license plate region detection is proposed in [3]. A threshold of 10 is set to convert RGB to binary and apply dilation technique. Template matching technique is used to detect the license number. Sliding Concentric Window technique along with HSI color model and position histogram to detect license plates is proposed in [4]. Sliding Concentric Window technique works by finding the edges contained in the license plate. Edge detection technique used here is similar to [2]. HSI color model is used for detection in varying light conditions is used in [5]. This model utilizes hue and intensity of colors. Histogram is used to detect and verify the right region which contains the license plate.

The proposed ANPR algorithm is used to recognize Punjab license plates, currently being implemented in Pakistan, which are green and white. In the first part, the license plate region is extracted. This is the most important step since the accuracy of the whole system depends on it. The second part involves segmentation in which the characters are separated for identification. The last step involves Optical Character Recognition (OCR) in which the extracted characters are compared against the stored templates for recognition.

The paper is organized as follows: Section II presents the process of license plate recognition, Section III presents the simulation results and Section IV presents the conclusion and future work.

II. Process of License Plate Recognition

The license plate recognition system works as follows: a signal is sent to the camera which acquires the picture of the vehicle. A 2 mega pixels USB camera is used for this purpose. The pictures are taken from 2 meter distance. The picture is saved on the system. The path to the picture is defined in the MATLAB algorithm. As soon as the picture is saved, MATLAB starts the number plate extraction algorithm. The algorithm is implemented in MATLAB 2009. It is divided into three parts:

- Extract the license plate region.

Sarmad Majeed Malik is doing Masters in Electrical Power Systems and Automation from North China Electric Power University, Beijing, China (Phone:+8613264400924; e-mail: sarmadmalik90@gmail.com).

Rehan Hafiz is with National University of Sciences and Technology (NUST), Islamabad, Pakistan. (Phone:+923315189556; e-mail: rehan.hafiz@seecs.edu.pk).

Segmentation

- OCR.

A scheme comprising of various stages involved in the process is presented below in Fig. 1.

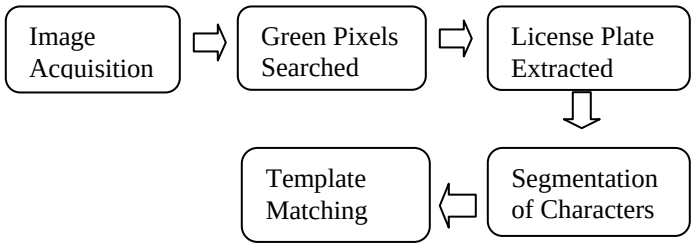

Fig. 1 Various stages of license plate extraction

A flow chart of the algorithm is presented in Fig. 2. In the first part, the license plate region is extracted. This is the most important step involved and it determines the accuracy of the whole algorithm. If this step extracts the license plate region correctly, then the remaining algorithm will search and identify the license plate correctly. An incorrect license plate region extraction leads to incorrect results. For this step, the common features of the license plate are utilized. Since the Punjab license plate is green color followed by white color, so the approach adopted is to search for green colored regions in the picture. This is the starting point of the algorithm. The RGB color model is used. The picture is converted to this format. Each pixel in this model is a combination of three colors, namely red, green and blue. The perfect green color has only green component and no red or blue components. But the green color of Punjab license plate is not perfect green but a shade of green color which also has red and blue components. A condition is set which utilizes the ratio between the three colors. This is shown in (1).

$$\frac{x}{y} \; and \; \frac{x}{z} > 1 \tag{1}$$

where x is the green color weight, y is the red color weight and z is the blue color weight.

RGB model consists of red, green and blue color and each of these colors is one byte long. This implies that a total number of 2^8=256 combinations may be used. The weight of the color is this value which lies between 0 and 255. Shades of different colors are possible by mixing these three colors in variable quantities. For green color recognition, the ratio of green color to red color and the ratio of green color to blue color should be greater than one and this correctly identified the green color regions. But the picture may also contain other green regions which may be smaller or larger than the target region and still satisfy this condition. These regions were eliminated by connected component analysis technique. Since connected components are a set of pixels in contact with one another, each region may be identified by labeling the connected components. Labeling assigns each region with a number which helps in their identification. The connected components were labeled and their coordinates found. The areas of these components were calculated and ratio technique utilized. The length to height ratio of the green region of actual license plate was measured and recorded. This ratio remains the same no matter what the distance. The length to height ratio of each of the connected components was recorded and compared against the set value. This correctly extracted the green region of the license plate.

To identify the full license plate region, the concept of ratio was used again. Actual dimensions of the license plate revealed that the ratio of lengths of green to white color in the license plate was 1:1.6. No matter what the distance is from which the picture is taken, this ratio will remain the same. So the length of the license plate was calculated and multiplied by 1.6 to get the entire license plate region. This correctly extracted the license plate region and is shown in (2).

$$A:B = 1:1.6 \tag{2}$$

where A is the length of green region and B is the length of white region.

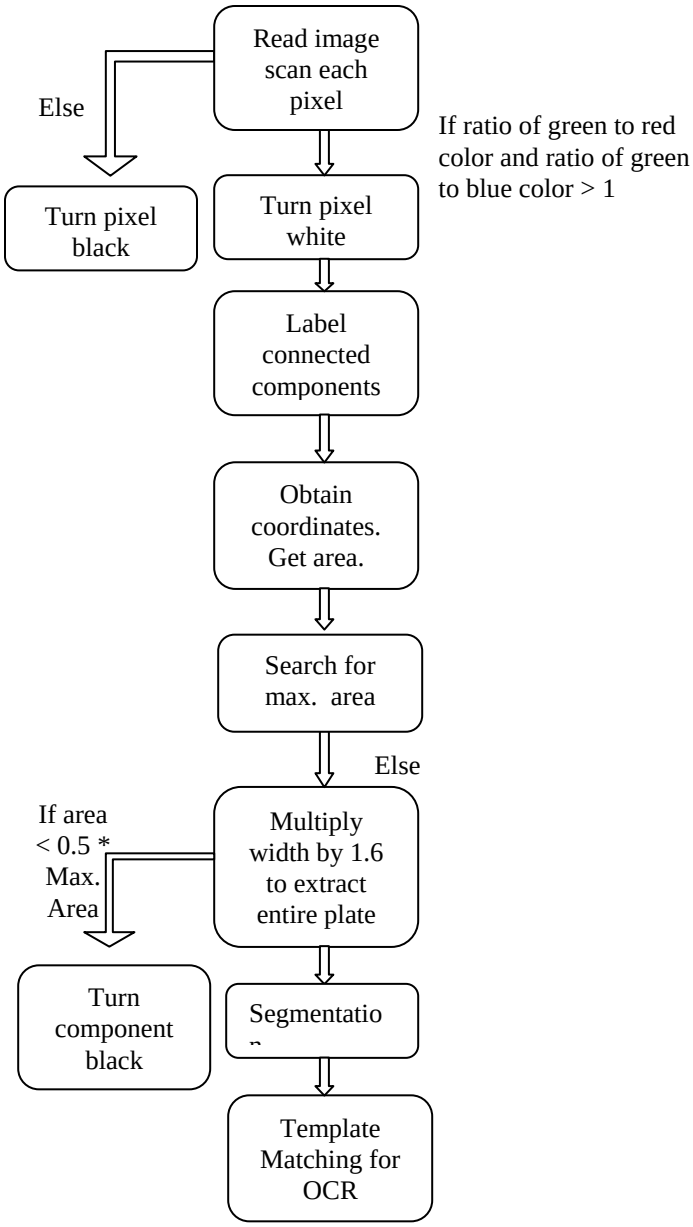

Fig. 2 Flow chart of algorithm

The second part required segmentation of characters in which the characters had to be separated from each other for recognition by the OCR. In this part, the concept of connected components was utilized since each character on license plate was just a connected component with a group of pixels in contact with each other. The image was changed from RGB to binary, connected components were labeled and their coordinates extracted. Using these coordinates, each connected component in the binary image was extracted. This concluded the segmentation part.

The final part in the algorithm was the Optical Character Recognition in which the extracted characters needed to be recognized. This was done using the template matching technique. The templates of characters were stored on the computer and each extracted character from the image was compared with these templates one by one. Correlation was used to search for maximum match. The result was a correctly extracted license plate.

III. Simulation Results

The results produced from the implementation of the algorithm are presented in this section. The image is acquired using a 2 mega pixels USB camera. The picture captured is in RGB format and is shown in Fig. 3. Fig. 4 shows the connected components located on the picture. These are shown in white color with black background. Fig. 5 shows the green color region of the license plate extracted which is later used to extract the entire license plate shown in Fig. 6. Only the desired region is shown with the rest of the picture shown as black.

The concept of connected components is applied to extract license plate. The characters on the license plate are nearly of the same size and smaller objects are eliminated by choosing the condition if the area of the connected component is less than the maximum area found, then that component is eliminated. This helps to refine the picture so that only the region which has the license plate is used. The actual ratio of length and height of license plate is also utilized in matching with the correct area. The condition is found to work to perfection and the results are shown in Fig. 7.

In the final step, the results of segmentation are subject to template matching. The templates are stored on the computer and correlation is used to check for maximum match. The figures which have the maximum match give the highest correlation value and this value is utilized to determine the characters. The results are shown in Fig. 8.

The algorithm has been implemented on 50 pictures and gives correct result on 45 with the accuracy of 90%. Varying light conditions play a major role. Different lighting conditions vary the intensity of the green color being searched which, in turn, influences the overall working of the algorithm. For the purpose of this algorithm, normal lighting conditions are used. Damaged and disoriented license plates are other factors which contribute significantly to license number extraction.

Fig. 3 Captured images using USB camera

Fig. 4 Corrected components shown in white

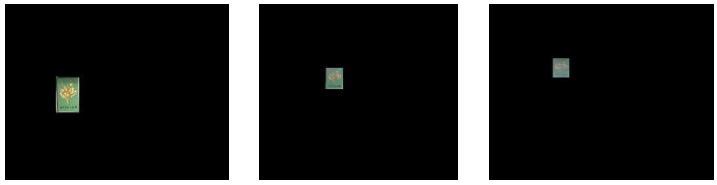

Fig. 5 Detected green region

Fig. 6 Entire license plate extracted

Fig. 7 Segmentation of characters

Fig. 8 Results after Optical Character Recognition (OCR)

IV. Conclusion and Future Work

This paper presents an algorithm for the license plate detection of vehicles. Various image processing techniques are used in implementing the algorithm in MATLAB 2009b. MATLAB 2009b is used since the concept of connected components is used and previous versions of MATLAB don't support this function. Real time images are taken using a USB camera and algorithm applied to these images. The results show that the algorithm successfully detects the Punjab license plates of vehicles and can be implemented in various applications.

Some modifications are still possible even though the algorithm works fine. MATLAB takes a lot of processing time if a high resolution camera is used because it is slow. This is due to the fact that MATLAB is designed on java programming. A solution to this problem is presented in the algorithm by cropping the processed image with the goal of decreasing processing data for MATLAB and making it faster. OpenCV is another solution to this problem since it is faster than MATLAB. High resolution cameras may be used while implementing the code in OpenCV. Still images are used in the implementation. To get images of moving vehicles, a high FPS (frames per second) camera may be used. Further clarity in image may be obtained by using anti-blurring functions to reduce the blur caused by motion of vehicle.

Since the algorithm utilizes the concept of template matching, alignment is an important factor. Small misalignments may produce different results. Affine transformation may be used to prevent this problem. The algorithm is designed to operate on standard Punjab license plates. For designing the algorithm to extract different license plates, the same process may be used but the initial condition of green color search should be changed. For example, if the algorithm is to be designed for China license plate extraction, then the initial condition should be to search for blue colored regions instead of green ones. The remaining algorithm should be the same.

REFERENCES

[1] Muhammad Tahir Qadri and Muhammad Asif, "Automatic number plate recognition system for vehicle identification using optical character recognition," *International Conference on Education Technology and Computer*, pp. 335-338, April 2009.

[2] S. Adebayo Daramola, E. Adetiba, A.U. Adoghe, J.A. Badejo, I.A. Samuel and T. Fagorusi, "Automatic vehicle identification system using license plate," *International Journal of Engineering Science and Technology (IJEST)*, vol. 3, no. 2, pp. 1712-1719, February 2011.

[3] Serkan Ozbay and Ergun Ercelebi, "Automatic vehicle identification by plate recognition," *World Academy of Science, Engineering and Technology 9*, pp. 222-225, January 2005.

[4] Kaushik Deb, Hyun Uk Chae and Kang Hyun Jo, "Vehicle license plate detection method based on sliding concentric windows and histogram," *Journal of Computers*, vol. 4, no. 8, pp. 771-777, August 2009. http://dx.doi.org/10.4304/jcp.4.8.771-777

[5] Yassin M.Y.Hasan and Lina J.Karam, "Morphological text extraction from images," IEEE Transactions on Image Processing, vol. 9, no. 11, pp. 1978-1983, November 2000. http://dx.doi.org/10.1109/83.877220

[6] Prathamesh Kulkarni, Ashish Khatri, Prateek Banga and Kushal Shah, "Automatic Number Plate Recognition (ANPR) system for Indian conditions," 19th International Conference on Radioelektronika, pp. 111-114, 2009.

[7] Xiaojun Zhai, Faycal Bensaali and Reza Sotudeh, "Real-time optical character recognition on field programmable gate array for automatic number plate recognition system," Circuits, Devices and Systems (IET), vol. 7, pp. 337-344, 2013. http://dx.doi.org/10.1049/iet-cds.2012.0339