Automatic Number Plate Recognition for Indonesian License Plate by Using K-Nearest Neighbor Algorithm

D Gunawan, W Rohimah, R F Rahmat
Department of Information Technology, Universitas Sumatera Utara, Medan, Indonesia
E-mail: danigunawan@usu.ac.id

Abstract. Transportation surveillance requires many human resources to cover all the road in a city. The human resources might be reduced by utilizing automation. For example, a task to recognize the license plate can be automated by utilizing CCTVs or surveillance cameras that usually installed in the crossroads. The automatic number plate recognition is required to record the license plate of the vehicle that violates the traffic rules. This number can be processed for future references. This research utilizes K-Nearest Neighbor (KNN) to recognize the license plate. The research is conducted to 125 license plate images which are divided into 100 training images and 25 testing images. The success rate of this research is 92.86%. The condition of the license plate and light intensity influence the recognition result.

1. Introduction
Automation in public transportation surveillance such as speeding violation or traffic light violation requires automatic number plate recognition. A surveillance camera has an important role in intelligent transportation system [1]. However, to observe the traffic in a city requires many human resources. The automatic number plate recognition might reduce the human resources to observe and record each vehicles license plate that runs in the road. A surveillance camera can provide the inputs for the automatic number plate recognition system. Previous research had implemented optical character recognition to recognize a license plate (or a number plate) [2]. In addition, the system was simulated by using Matlab. The violation is logged and sent to the user via SMS. Another research also uses optical character recognition to recognize a license plate [3]. However, instead of sending the result via SMS, this research compares the recognized license plate with the existing database. The research conducted by Bhavin et al [4] proposed an improvement to OCR-based number plate recognition by using a trained neural network. In 2015, a research conducted by Islam [5] proposed a method to recognize license plate by using structured elements. The method is developed based on morphological operations based on elements structure. The aim of this method is maximally excluded non-interested region and improving object area. This research yields a 92% success rate with the varying light condition. Automatic number plate recognition researches are also conducted for a specific region such as India and Sudan [6] [7]. A research to recognize number plate in Indonesia was conducted by using optical character recognition [8]. They improved the optical character recognition by construct higher resolution template, using more relevant font type, and removing
available noise. Our research utilizes K-Nearest Neighbor (KNN) instead of optical character recognition to recognize license plate. Because of the difference among license plate designs in many countries, this research limits the number plate recognition to Indonesian license plate only.

This paper organization is as follow: the first section is an introduction which describes the background of the research, related works, and research aim. The second section is the elaborate description about research method used in this research. The third section is the result of the research and followed by the research summary in the last section, which is the conclusion.

2. Research Method

There are several processing stages to detect number plate in the vehicle as illustrated in figure 1. The first stage is capturing the license plate images for dataset collection. These images will be used as the training and testing data. Second, the image pre-processing stage includes cropping, resizing and grayscaling. The third stage is the segmentation process includes edge detection and thresholding. The fourth stage is feature extraction which applies contour detection and Principal Component Analysis (PCA). In the next stage, all the images from the previous stage are segmented to obtain letters and numbers images. These images will be used as the input for training and testing data. The training and testing data are processed by using K-Nearest Neighbor (KNN) algorithm.

The detail of the processing stages is described as follow:

2.1. Dataset collection

License plate images in this research are collected directly from the parking area. These images will be used as the training and testing data. All the images are taken by using a digital camera which has been set up with 1080 x 810 pixels resolution. The captured images are saved in JPG format. The captured license plate has standard form and size, which is divided into three sections. The first section is the area code or regional code, which represents the place of the registered vehicle. This section can have one or two digits of the alphabet. The second section consists of 2-4 digits of the number. And the last section consists of 1-3 digits of the alphabet. This research used 125 images, divided into 100 training images and 25 testing images. All the images are taken in good quality. The example of testing data is shown in figure 2. Some of these images are taken indoor and outdoor. This research only takes the pictures of the license plate of the private vehicles. The private vehicles license plate has the black background and white foreground. In some cases, the owner of the vehicle modifies the license plate by adding sticker or writing.

2.2. Image pre-processing

This stage prepares the images before being the classification process is applied. The image pre-processing stage includes cropping, resizing and grayscaling. The detail of each process is explained below.

2.2.1. Cropping

The original images are taken from the close distance. However, the images are still not focus on the license plate (shown in figure 3a). Therefore, the images should be cropped to the license plate size manually. The result of the cropped image is shown in figure 3b.

2.2.2. Resizing

All the cropped images have different size because the image was taken within the various range of distance. However, the input of the images should be in the same size. Therefore, we should resize all the cropped images to the same size. This research resizes all the cropped license plate images to 216 x 106 pixel.
2.2.3. Grayscaling  Basically, the cropped license plate images have two dominant colors, such as black and white. The background of the license plate is black. On the other hand, the foreground is white. However, some vehicle owners put things such as stickers, pin or else which are stuck to the license plate and have various colors. Therefore, in order to perform segmentation, the image should be converted to grayscale. This research utilizes OpenCV library to complete this operation.

2.3. Segmentation  

The segmentation stage consists of three processes, such as edge detection, thresholding, and gaussian blur. The edge detection divides two homogenous areas which has different brightness. This research uses the canny algorithm to perform edge detection. This method is used because it has the ability to put and mark every edge according to convolutional parameter selection. In addition, it provides high flexibility to determine preferred edge thickness. Canny also supports edge detection with minimum error level. The result of the edge detection is shown in figure 4.

Moreover, thresholding is used to distinguish background and foreground (letters and numbers on license plate). This process converts gray-level images to bi-level images (binary images). The next process is removing the noise by using Gaussian. This process is required to avoid
detecting noise as the object. All these three processes are implemented by using the OpenCV library.
2.4. Feature Extraction
This research utilizes the Principal Component Analysis (PCA) to extract features from the images. PCA is a method to obtain important features from a set of data by performing decomposition to the data, therefore it yields uncorrelated coefficient. Feature extraction is done to reduce the data dimension by maintaining as much information as possible. PCA maintains that information from the original dataset. Its projection data has high variants determined by eigenvectors.

2.5. Classification
Extracted features from the images will be classified by using K-Nearest Neighbor (KNN) algorithm. This process will classify the object by determining the closest object based on the training data. The KNN algorithm is divided into two different processes, namely training and testing processes. The training process is the data validation stage. As shown in figure 5, each image of license plate will be divided into nine pieces. For each piece, we should put the corresponding value to the available text field below the image. The number of training data is 60% of the total dataset images. The testing process classifies the images based on the training data.

![Figure 5. Data input for training process](image)

3. Result and Discussion
This section is discussing the implementation result of the K-Nearest Neighbor algorithm in automatic plat number recognition. This research is tested on a computer which has the specification as follow:

(i) CPU: 1.80 GHZ
(ii) Harddisk: 500 GB
(iii) RAM: 2 GB
(iv) Operating system: Microsoft Windows 8.1 Pro 64-bit (6.3, build 9000)
(v) Library: OpenCV

The testing is conducted to 25 license plate images which will be recognized based on 100 training images. The result is shown in table 1. According to the result, we obtain 92.86% of success rate.

Table 1 no. 17 shows the license plate BK6892XZ is only 87.5% recognized by the system. The license plate is not fully recognized because the plate has a wavy surface. Therefore, because of its wavy surface, the light distracts the system to read the character correctly. The final result of the license plate recognition is BW6892XZ. The system recognizes BW instead of BK.
Table 1. Testing Result

| No | Image     | Result  | %   | No | Image     | Result  | %   |
|----|-----------|---------|-----|----|-----------|---------|-----|
| 1  | BK3856RAG | 100     | 14  | 14 | B45626ZAG | 89      |     |
| 2  | BK4235ADD | 100     | 15  | 15 | BB5038HI  | 87.5    |     |
| 3  | KK4152AGO | 89      | 16  | 16 | BK6750MZ  | 100     |     |
| 4  | BK4452AGD | 100     | 17  | 17 | B6 6892XZ | 87.5    |     |
| 5  | BK5274GG  | 100     | 18  | 18 | B94442AEW | 89      |     |
| 6  | KK6808AGF | 89      | 19  | 19 | BK3597AGN | 100     |     |
| 7  | BK6130CW  | 100     | 20  | 20 | B55550VAL | 89      |     |
| 8  | BK6063AE0 | 100     | 21  | 21 | B02763ACB | 89      |     |
| 9  | P1905PZB  | 87.5    | 22  | 22 | BK5700AGB | 100     |     |
| 10 | B6703WJF  | 100     | 23  | 23 | B26889 NS | 87.5    |     |
| 11 | B15362QI  | 87.5    | 24  | 24 | BZ3069 GP | 87.5    |     |
| 12 | B05966KF  | 62.5    | 25  | 25 | B26545ABW NS | 89 | |
| 13 | BK4560FF  | 75      | -   | -  | -         | -       | -   |

Table 1 no. 13, the success rate of the automatic number plate recognition is 75%. The system yields BK4560FF instead of BK4566MF because of the leak of light and brightness in the license plate. The light is the main problem of the automatic number plate recognition.

Table 1 no. 12 has a lower accuracy of 65.25%. This license plate is an example of broken license plate. The accident causes the size of the plate changed. We suspect the change in the plate dimension influence the plate number recognition. The ratio of the plate number will be different from the standard plate number. In addition, the accident brings up some curves and cause the reflection of the sunlight in the surface of the plate. This reflection still remains and cannot be eliminated in the pre-processing. The reflection will be segmented and considered as characters. Both reasons might decrease success rate.

According to the Table 1, most of the plate number recognition mistakes occur in the recognition of regional code such as BK, BL, BB, B, an so on. The mistakes might occur if one or more condition below presents. For example:
• The plate number condition is changed because of third-party modification or accident
• The plate number recognition requires enough light
• The characters are printed closely to each other
• Some characters are similar, for example, the character 6 is similar to G, 0 is similar to O
• The license plate position

4. Conclusion
This research aims to recognize license plate images automatically to fulfill the requirement for automation in transportation surveillance. This research obtains the license plate images by taking their pictures indoor and outdoor manually. The license plate usually has two dominant colors. The background of the plate is black, and the foreground is white. This kind of plate is used for private vehicles. This research utilizes K-Nearest Neighbor in order to recognize the numbers and characters in the plate. The result shows the success rate of number plate recognition is 92.86%. There are some obstacles to recognize the number plate such as license plate modification, the characters are printed closely to each other, similar characters to each other, and the license plate position. These issues should be covered in the next research.

References
[1] Qiu D, Lu J, Sun X and Jiang J 2010 Application of graphic minimal covering algorithm in the distribution of surveillance cameras in small and medium-sized city road networks 2010 International Conference On Computer Design and Applications vol 1 pp V1–200–V1–203
[2] Shreyas R, Kumar B V P, Adithya H B, Padmaja B and Sunil M P 2017 Dynamic traffic rule violation monitoring system using automatic number plate recognition with sms feedback 2017 2nd International Conference on Telecommunication and Networks (TEL-NET) pp 1–5
[3] Qadri M T and Asif M 2009 Automatic number plate recognition system for vehicle identification using optical character recognition 2009 International Conference on Education Technology and Computer pp 335–338 ISSN 2155-1812
[4] Kakani B V, Gandhi D and Jani S 2017 Improved ocr based automatic vehicle number plate recognition using features trained neural network 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT) pp 1–6
[5] Islam R, Sharif K F and Biswas S 2015 Automatic vehicle number plate recognition using structured elements 2015 IEEE Conference on Systems, Process and Control (ICSPC) pp 44–48
[6] Kulkarni P, Khatri A, Banga P and Shah K 2009 Automatic number plate recognition (anpr) system for indian conditions 2009 19th International Conference Radioelektronika pp 111–114
[7] Mubarak H, Ibrahim A O, Elwasila A and Bushra S 2017 Sudanese license plate identification using automatic number plate recognition 2017 Joint International Conference on Information and Communication Technologies for Education and Training and International Conference on Computing in Arabic (ICCA-TICET) pp 1–8
[8] Hidayatullah P, Syakrami N, Sukartini I and Muhlis W 2012 Optical character recognition improvement for license plate recognition in indonesia 2012 Sixth UKSim/AMSS European Symposium on Computer Modeling and Simulation pp 249–254