



Recognition Number of The Vehicle Plate Using Otsu Method and K-Nearest Neighbour Classification

Maulidia R. Hidayah¹, Isa Akhlis², Endang Sugiharti³

^{1,2,3}Departement of Computer Science, Mathematics and Science Faculty, Universitas Negeri Semarang
Email: ¹maulidiarh@yahoo.com, ²132231405@mail.unnes.ac.id, ³endangsugiharti@mail.unnes.ac.id

Abstract

The current topic that is interesting as a solution of the impact of public service improvement toward vehicle is License Plate Recognition (LPR), but it still needs to develop the research of LPR method. Some of the previous researchs showed that K-Nearest Neighbour (KNN) succeed in car license plate recognition. The Objectives of this research was to determine the implementation and accuracy of Otsu Method toward license plate recognition. The method of this research was Otsu method to extract the characteristics and image of the plate into binary image and KNN as recognition classification method of each character. The development of the license plate recognition program by using Otsu method and classification of KNN is following the steps of pattern recognition, such as input and sensing, pre-processing, extraction feature Otsu method binary, segmentation, KNN classification method and post-processing by calculating the level of accuracy. The study showed that this program can recognize by 82% from 100 test plate with 93,75% of number recognition accuracy and 91,92% of letter recognition accuracy.

Keywords: Vehicle Plate, Image, KNN, LPR, Pattern Recognition

1. INTRODUCTION

The growth of vehicles in Indonesia is increasing significantly. BPS Semarang data [1], states that the growth of the average vehicle major cities in Indonesia about 8% per year with growth of road is only about 2-5% per year. Semarang city itself has a growth of private vehicles (cars and motorcycles) as much as 2%. Increased growth of such vehicles should also be balanced with efforts to improve services for vehicles such as the parking and toll systems.

Recently, the trends and topics that are interesting about efforts to improve service to the public transportation such as the parking lot and toll system is the recognition of a vehicle license plate numbers automatically or License Plate Recognition (LPR) [2] [3]. LPR widely used in the access to the parking lot, vehicle traffic control and security and surveillance vehicles [4]. LPR purpose is to improve the effectiveness of the parking system and toll roads. LPR utilizes input vehicle license plate as the vehicle's identity in the system.

LPR can be done by image processing on the plate captured using the camera. Parking system in Indonesia are still many who do not utilize the LPR because there are many obstacles in their utilization, especially in motorcycles. Research development still needs to be done in order LPR can be applied efficiently in Indonesia by searching for the best method to reduce the error recognition vehicle license plate numbers automatically. Research on the development of the LPR requires a combination of

image processing and artificial intelligence. LPR stages in the process are preprocessing, feature extraction and classification (grouping) [5].

Feature extraction in pattern recognition to facilitate the need for bineritation pixels. The feature extraction can be done by Otsu method. Otsu method is a method that can change the gray level image into a binary set based on the value threshold with digital image pixel color values. Thresholding Otsu is the development of a histogram that can provide a good segmentation results so that the binary image clean of *noise salt and papper* [6]. Previous studies [7], shows that the method of Otsu able to deliver a satisfactory binary image and is very helpful in the process of determining the characteristics of hand geometry. The approach taken by the Otsu method is the discriminant analysis is to determine the variables that can distinguish between two or more groups naturally without specifying a threshold value beforehand. The image characteristic determination result is then performed segmentation [8]. Results segmentation is further classified.

The method used in good LPR is a method of classification that remains consistent in a large amount of data. This type of classification can be used in a large amount of data is k-Nearest Neighbour (KNN). KNN algorithm classifies based on the shortest distance between the data evaluated by the closest point in the training data [9]. Function to find the closest point to the distance *Euclidean* formula. KNN Classified is based on an analogy comparing the similarity of the test data to training data. N training data given attribute, each record is a point in n-dimensional space. All training data stored in the pattern of n-dimensional space. When the training data that is not known, then KNN looking for a pattern space k closest training data with test data that is not known. Test data is then entered into the most common class (similarities) between the k nearest neighbors. Another advantage of the KNN algorithm is robust against noise training data plate number which does have a lot of noise.

Based on this background, then the discussion about the recognition of number plates of vehicles using Otsu method to convert the image into a binary image and KNN as the motorcycle classification at Parking Place of Faculty of Mathematics and Natural Sciences, State University of Semarang and then determine the level of accuracy.

This study is expected to develop previous research on the LPR so that it can be implemented in life and know the use of KNN as a method of classification on the recognition of a motorcycle license plate number.

2. METHODS

Chronology of the vehicle license plate number recognition system with Otsu method and KNN classification broadly divided into several parts as in Figure 1.

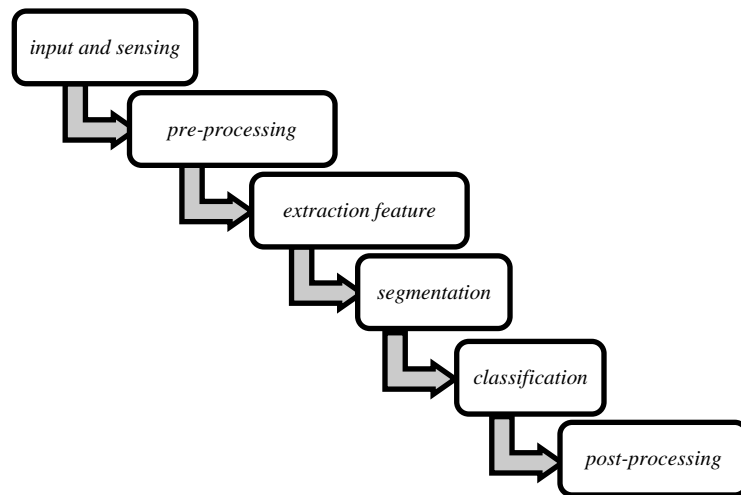


Figure 1. Pattern Recognition Process

2.1. Input and Sensing

process and sensing input is a process of data collection and data entry then data processing in accordance with the existing characteristics. The process started with a plate image acquisition with license plate image capture with a camera and then the process of cutting the top plate of the vehicle. Pieces plate of the vehicle is then input RGB image into the program with format **.jpg**. Image then resize to 500×2000 pixels, after the test data is ready, the program continues to the next process. Pieces plate input image is as shown in Figure 2.



Figure 2. Plat Input Image

2.2. Pre-processing

At this stage rebrand RGB to gray. The process of changing the RGB value of each character. Function to convert the RGB image into a grayscale on `rgb2gray` in Matlab is as follows.

$$\text{grayscale} = 0,289 \times R + 0,5870 \times G + 0,1140 \times B \quad (1)$$

Figure 3 is an RGB image display plate after becoming image grayscale and resize to 500×2000 pixels.



Figure 3. Plate Image Grayscale

2.3. Feature Extraction (Otsu Method)

Feature extraction is the process of finding a new representation of the image. Feature extraction process in this study with thresholding a single the method of Otsu. Otsu method to change the image with gray level pixel value between 0-255 binary image that is only worth 0 and 1. Thresholding in general must define a threshold value beforehand, but not by the method of Otsu. Otsu method has a function to determine a threshold value (T) as follows.

$$\sigma_B^2(T^*) = \max_{1 \leq T < L} \sigma_B^2(T) \quad (2)$$

defined

$$\sigma_B^2(T) = \frac{[\mu_T \omega(T) - \mu(T)]^2}{\omega(T)[1 - \omega(T)]} \quad (3)$$

The binary image on the plate vehicles forming each character into a matrix (pattern). Then the noise in the binary image is cleaned. Cleaning noise serves to facilitate the segmentation process. Noise or groups of pixels that are less than 11,000 pixels are then removed. Figure 4 is a view of the results of the Otsu bineritation been cleared of noise.



Figure 4. Image Binary Noise-Free

2.4. Segmentation (Segmentation Line and Character)

Segmentation is the process of separating one object with another object in an image. The segmentation process is done after the program to reduce noise in the image. Segmentation in this study is divided into two line segmentation and character segmentation. Line segmentation is performed to determine the position of a horizontal plate that has been used (scanline). Plat form is used only at the top there is only one line matrix recognizable.

Character segmentation is done to cut every character on the plate. Cutting character is based on a matrix formed on each character. Figure 5 shows the results of the segmentation of the characters in one of the test data.



Figure 5. Results of Character Segmentation

2.5. Classification (K-Nearest Neighbour Classification)

Classification process is a process of grouping test data into classes that have been determined by using a learning algorithm. KNN classification based on the value of neighborhood test data with training data.

KNN classification algorithm:

1. KNN classification process began by determining the value of k value neighbourhood amount of data, the value of k can be $k = 1, 2, \dots, n$. The recognition plate issued only a single result, this research uses the value $k = 1$.
2. Then calculate the value of neighborhood (distance) test data to each training data obtained based on *Euclidean* formula as follows.

$$d_i = \sqrt{\sum_{i=1}^p (x_2 - x_1)^2} \tag{4}$$

Calculate the distance here is the result of the calculation of the distance between the binary matrix test data with all training data matrix contained in the database with *Euclidean* the formula.

3. Sorting the results of neighborhood values.
4. Selecting the minimum value of neighborhood.
5. Check the minimum value of the class, then the test images is also included in the class.
6. Recognizable characters according to type of class.

Figure 6 shows the results of the recognition of number plates of vehicles as classified by KNN.



Figure 6. Results Number Plate

2.6. Recognition

Characters that have been identified, and then displays the results in the form of text. The results in text form is then saved automatically to *excel report.xls* file.

2.7. Post-Processing

Process Post-processing is a process of evaluating the level of succes in recognizing the characters. One way to evaluate the success rate is to determine the level of accuracy of the recognition of the license plate number of the vehicle. This level of accuracy is a representation of success in terms of percent. Function to calculate the accuracy rate is as follows.

$$accuracy\ rate(\%) = \frac{right\ data\ total}{data\ total} \times 100\% \quad (5)$$

The value of accuracy by calculating the value of the data that can be recognized divided by the total data is multiplied by 100%. Accuracy value calculation to take into account the recognition accuracy of letters, numbers and license plate number of test images.

3. RESULTS AND DISCUSSION

The results of the recognition of the vehicle license plate number is automatically saved on file an *excel* to do the next process is post-processing. The process for calculating the level of accuracy of the use of methods of Otsu and KNN classification on research LPR.

Based on the recognition result, the program can identify 82 of 100 plate entirety. Then the accuracy of the total image of the license plate of the test are as follows.

$$accuracy\ rate(\%) = \frac{right\ data\ total}{data\ total} \times 100\% = \frac{82}{100} \times 100\% = 82\%$$

The recognition of each character found that there are 25 errors recognize numbers and 26 error letters identify the out of a total 722 characters. Table 1 shows the results of calculating the level of accuracy based on image recognition results of test data to training data.

Table 1. Results of Calculation Accuracy Level

Characters	Numbers	Letter	Plate
Data Total	400	322	100
Error	25	26	18
Number of true	375	296	82
Accuracy Rate	93.75%	91.92%	82%

Program to recognize as much as 93.75% in number and 91, 92% can recognize the letters as well as an recognition to the license plate number by 82%. The use of this method is still not completely perfect.

An error on the recognize of the vehicle license plate number because some of the constraints and limitations that can lead to the recognize of the number plate becomes less perfect. Limitations may occur such as the recognition of **data98**, the recognition of data errors due to the position of the screws on the cover plate on the character of the plate so that the object has been recognized by the system into other characters. The result of the recognition of **data98** is KH4269MM. An error in recognizing a for their

object is not recognized (bolt) which change the shape of the object segmentation character. Shape of the object is unknown (bolt) which cover these characters is one of the missing feature. Missing feature is if there is an object unknown patterns that can be included in the classification, for example, the light intensity is too much on the image can be considered to be a new object. Missing feature, the object is not known not for cleaning due to noise. One solution that can be done with computation *Bayesian*[10]. Figure 7 shows the image of the original plate and after image processing.



Figure 7. Image Data98 (a) Original Image and (b) Binary Image

Recognition errors can also occur because of the form of font characters on the plate of motorcycles vehicles in Indonesia is still less standardized. Shape font at several different motorcycles when these motorcycles have an official license plate of traffic unit police. Training data available in the database can not recognize, such as the **data26** which has shape font less standard so that some characters can not be recognized properly. Character 0, D, 8 and Q indistinguishable eyes and to recognize the system, whereas the size of the test data in the database is large enough. K value being used increasingly large, it can be increasingly able to also test a large data size [10]. This study uses $k = 1$, making the results have been inconsistent because it is not balanced by the number of test data. Shape cutting less precise result that initially can be identified with 0 and D into a U. Figure 8 is the result of the recognition of vehicle registration number plates when the input image is skewed.

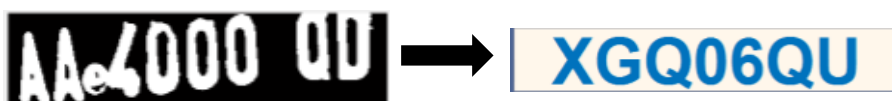


Figure 8. Results of Recognize **data26** Because Plat Leaning Position

Position affect vehicle license plate recognition results on **data3** very disappointing, there are only two characters that can be recognized correctly. Noise feature occurs in **data3** this because the characters figure 2 is not recognized as most forms of objects lost during the cleanup noise. Noise feature is the object which corrupt due to clearance noise statistical[10]. Cleaning noise at first to remove the bolts but most of the area code 2 to come clean because it is considered as part of the noise. Some of the other characters can not be recognized because of the position of the vehicle license plate is less precise make cutting plate becomes less precise. Cutting the plate is less precise form has led to a system error in the process of segmentation per character. **Data3** position actual plate is tilted and writing numbers too protrudes upward, then when it is done by cutting a rectangular plate the outer area to be in parts of the plate. The area outside the plate to form objects that have been difficult at the segmentation stage. Truncation errors for image segmentation process is not based on the pixel value

distribution, the distribution is based on the intensity of light is formed. Cutting at this time with the manual plate form a rectangle regardless of the form of the original object. Object segmentation can be automatically recognized can be done because of differences in the distribution of pixel [12]. One solution to recognize position offers may plate is done using masking techniques such as at research on facial expressions[13] that cut the shape object by object on 4 edge angle. Figure 9 shows a plate image **data3** after preprocessing, character 3, 6, V and C is regarded as one character so that the recognition of these data to be bad.



Figure 9. Binary Image Plate Tilting

Plate also affects the physical form of the recognition of this license plate numbers, as in **data45** and **data76**. Physical shape plate is not good as tangled make lighting disturbed image and shape objects that are not recognized. The lighting is very bright on a plate, less suitable when using this method because the method of this study is not the processing of the brightness intensity plate. This resulted in the character are exposed to light is too bright and too dark, the characters are not recognized properly. Example is **data45**, characters S and G becomes unrecognized due to no separation as the matrix is formed. Light and dark imagery have influenced the distribution of the histogram, changes in light and dark can be done by changing the brightness of the image. Brightly colored image histogram graph makes dominant shifted to the right while the dark image histogram on the dominance of the distribution on the left. If the spread is too bright or too dark then the distribution histogram into to one side, making the process of segmentation becomes difficult due to the high density of the histogram [14]. Number plate recognition results **data45** and **data76** be disappointing. Appearance on the binary image **data76** as shown in Figure 10.



Figure 10. Binary Image on **data76**

Recognition results show that the use of Otsu method makes it easy to program in bineritation image thus formed matrix pattern becomes clear and easy to segmentation. KNN method also makes the recognition of more flexible because it is based on the proximity of the existing training data. While the weakness of the study is the execution time and the limitations of the use of functions such as to recognize the position of the plate without previous cut, the shape of the plate, plate position and the position of the character on the motor vehicle license plate is still a lot different. The difficulty in this study also, the number of characters on the license plate are not clear and could be different across the plate with one another, resulting in an error on perkarakter segmentation.

Another problem that occurs in a private vehicle number plates in Indonesia are many variations of the shape and size of the plate. This is because a change in the design of vehicle number plates since april 2011, made many license plates are not standardized. Non-standard number plates which have affected the difficulties in the recognition of number plates of vehicles [15].

Conditions under which the arrest has affected the process of recognize of such a vehicle license plate number and position of the image capture of light intensity during an arrest. Position catching image be taken into account precisely because of the position of the catch plate also affects the image object detection, cleanup noise, segmentation and character recognition. Lighting time of arrest has affected the cleaning process noise and the next process because in this study does not change the light intensity of the original image.

4. CONCLUSION

Vehicle number plate recognition process using Otsu method in this study based pattern recognition process by utilizing the binary vector without the influence of the threshold value determination in advance by adjusting the distribution of the pixel values of the image so get a good result and a binary segmentation better. KNN classification in this study is quite good and proved a great recognize to the vehicle license plate number.

Results of recognition accuracy by using the vehicle registration number and classification KNN Otsu method that can recognize as many as 82% of the 100 plate test with 93.75% recognition accuracy of the numbers and letter recognition accuracy of 91.92%.

5. REFERENCES

- [1] Kariada, N. 2011. Tingkat Kualitas Udara di Jalan Protokol Kota Semarang. *SainteknoI*. Vol. 9(2): 111-120.
- [2] Xia, H., and D. Liao. 2011. The Study of License Plate Character Segmentation Algorithm based on Vertical Projection. *International Conference on Consumer Electronics, Communications and Network (CECNet)*. China, 2001: 4583-4586.
- [3] Anishiya, P. and P.S.M. Joans. 2011. Number Plate Recognition for Indian Cras Using Morphological Dilation and Erosion with the Aid Of Ocrs. 2011 *International Conference on Information and Network Technology*. Vol. 4: 115-119.
- [4] Wu, H.C., C.S. Tsai, and C.H. Lai. 2004. A License Plate Recognition System in E-Government. *Information Security*. Vol. 15(2): 199-210.
- [5] Khedidja, D. and M. Hayet. 2015. Printed Digits Recognition Using Multiple Multilayer Perceptron and Hu Moment. *Symposium on Complex System and Intelligent Computing (CompSIC)*. Algeria, March 2015.
- [6] Vala, H.J. and A. Baxi. 2013. A Review on Otsu Image Segmentation Algorithm. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*. Vol. 2(2): 387-389.

- [7] Putra, D. 2004. Binerisasi Citra Tangan dengan Metode Otsu. *Teknologi Elektro*. Vol. 3(2): 11-13.
- [8] Otsu, N. 1979. A Threshold Selection Method from Gray-level Histograms. *IEEE Transaction on Systems, Man and Cybernetics*. Vol. 9(1): 62-66.
- [9] Lili L., Y. Zhang, and Y. Zhao. 2008. K-Nearest Neighbours for Automated Classification of Celestial Objects. *Science in China Series G-Phys Mech Astron*. Vol. 5(7): 916-922.
- [10] Duda, R.O., P.E. Hart, and D.G. Stork. 1991. *Pattern Classification 2nd ed.* Wiley-Interscience, New York.
- [11] Han, J., M. Kamber and Jian P. 2012. *Data Mining Concepts and Techniques 3th ed.* Morgan Kaufmann Publishers, Waltham.
- [12] Gonzalez, R.C., R.E Woods, and S.L. Eddins. 2009. *Digital Image Processing Using Matlab*. Prentice Hall, New Jersey.
- [13] Abidin, Z. and A. Harjoko. 2012. A Neural Network based Facial Expression Recognition using Fisherface. *International Journal of Computer Applications*. Vol. 59(3): 30-34.
- [14] Munir, R. 2004. *Pengolahan Citra Digital dengan Pendekatan Algoritmik*. Penerbit Informatika, Bandung.
- [15] Amin, M.A. 2015. Penerapan Reduksi Region Palsu Berbasis *Mathematical Morphology* pada Algoritma *Adaboost* untuk Deteksi Plat Nomor Kendaraan Indonesia. *Journal of Intelligent System*. Vol. 1(1): 9-14.