ISSN 2278-3091
Volume 9, No.5, September - October 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse44952020.pdf
https://doi.org/10.30534/ijatcse/2020/44952020

Broken Character of Car Plate Recognition using Template Matching with Relocation Process

Syazreen Zulkifli[1], Suhaila Abd Halim[2]
[1]Center of Mathematics Studies, Faculty of Computer and Mathematical Sciences Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia, syazreenzulkifli@gmail.com
[2]Center of Mathematics Studies, Faculty of Computer and Mathematical Sciences Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia, suhaila889@uitm.edu.my

ABSTRACT
In the real situation, there are cases that the car plate consists of broken character. This case affects the process of recognition of the car plate especially when any dangerous situation happened such as accident and car robbery. Moreover, it caused difficulties to the Road Transportation Department (JPJ) and the police officer to run their work while preceding some law enforcement such as traffic law enforcement, AES and speed trap enforcement. Besides, the problems become severe when human is required to observe the car plate manually. Due to that, this research aims to propose and implement recognition method of broken character using template matching and measure the performance of the proposed process using correlation (CORR) and similarity measure (SSIM). In this study, 25 images of car plate consist broken character are used where 20 of them are from simulation process while the other 5 are real data. Next, the car plate area on the car images are detected and extracted. Then, each character is segmented and broken character is identified by *NewMeanThreshold*. After that, the character relocation is completed by making the comparison between the segmented and template character. Lastly, Template Matching is used for recognition process. The algorithm is evaluated using CORR and SSIM. Results of evaluation show the performance rate of 84% for car plate area detection, 100% for the character recognition by CORR and 76% for the character recognition by SSIM. As conclusion, the proposed processes are suggested to be an alternative for character recognition in car plate to assist in the recognition of the broken character.

Key words : broken character, connected component labeling, correlation, Structural Similarity Measure, Template Matching

1. INTRODUCTION

Nowadays, a lot of researches have been done in the detection and recognition of car plate using various methods. However, recognition of car plate problems such unclear image, non-standard font and character, or broken character remain a challenge. To overcome these problems, an extra study and work should be carried out for example providing the suitable and efficient algorithm to counter the problems.

Therefore, this study is focus on the non-split kind of broken character problem which means the shape of the broken character cannot be reconnect or reconstruct. The example of the split broken character is shown in Figure 1.

(a) Split Broken Character (b) Non-Split Broken Character

Figure 1: Example of Split and Non-Split Broken Character of N

In Figure 1(a), the letter 'N' is broken at the middle of the letter, this case is in the category where the broken letter is possible to reconnect or reconstruct in order to form the original shape of 'N' by using a method suggested by using graph combinatorics method [1], Fuzzy Logical approach [2] and modified vertical projection profile [3]. Furthermore, by observing the letter, the actual behavior of letter is already known and easily to be guess because of the little broken part. The letter 'N' in Figure 1(b) is partially broken on the right side of the letter and no possibilities to be reconstructed or reconnected. Moreover, the letter is kind of difficult to guess the actual behavior of the character because of the many possibilities that can the broken character be for example as 'M', 'K', 'R', and so on. Since that, the confidence rates of 5 highest possibilities of the recognized number plate regarding 5 possibilities that broken character can be are evaluated to provide the alternative results in the recognition process.

In the research of character segmentation in the application of documentation done by Droettboom [1], the concept of Connected Component Labelling (CCL) that also known as Connected Component Analysis (CCA) is used to segment the character. Besides that, CCA also be part of the Binary thresholding and morphological dilation processes for Natural language processing and image processing technology [4]. The input that based on voice output, whiteboard diagram and material in printed form were collected before the image processing application. In addition, the method is extended into Broken Connecting Character (BCC). The idea of the study comes when it is obvious that for a broken character on the old document, they are made up from too many connected components that makes the recognition becomes more difficult in the case of broken character. Therefore, BBC is used to reconnect the broken character to be a normal character to proceed in the recognition process. The process of segmenting the broken character, the broken character needs to be identified. Mean Based Thresholding in identification of broken character in degraded document is proposed to identify whether the segmented character is broken or not [5]. The method is used to investigate the width of each segmented character. Thus, the average or mean of the width of each character is calculated by dividing the summation of each character width and the number of segmented characters. Then each width of the segmented character is investigated by following the distribution.

For recognition, the examples of existing method nowadays are Template Matching, Support Vector Machine (SVM) and Artificial Neuron Network (ANN). Despite that, this study concentrates on using Template Matching to recognize the broken character.

The Template Matching takes place by matching up the extracted individual character with the standard character in the templates using CORR. The character with the maximum value of correlation function matched with the template is considered for CORR [6]. Then the computer recognized the character using the Template Matching algorithm. Thus, the process of recognition for the car plate should be a success if the recognized number plate is similar to the actual number plate as read by the human. Smearing algorithm for detection and segmentation is applied in addition to the application of template matching for character recognition [7]. The overall process produced 76.0% of system performance. Besides that, several factors that affected the system performance of misclassification also been measured. Choong, Keong and Cheah [8] proposed a system of simplified linear model where template matching is used for character recognition. The system demonstrated more than 90% success rate for the experimental result of each phases.

This study is aims to enhance the recognition algorithm using template matching, measure the performance of algorithm using two approaches of CORR and SSIM. This method offers an alternative method and observes the consistency of the algorithm by using both measurements. It is hypothesized that the enhanced recognition algorithm using Template Matching able to recognize broken character of the car plate correctly with high values of correlation and structural similarity measure.

2. TEMPLATE MATCHING

Template Matching uses the concept of CORR to find the rate of similarity of the comparison. In the meanwhile, the concept of Sum of Absolute Differences (SAD) in comparing the template and segmented character can also be used [9]. SAD values estimate the best position of the template within the input image by using (1).

$$d_1(I_J, T) = \sum_{i=1}^{n} \left| I_{i,j} - T_i \right| \qquad (1)$$

where, $d_1(I_J, T)$ is sum of the absolute difference, I denoted as an actual character in the template and T is a segmented character. In the meanwhile, *i* is number of current segmented character and *j* is the number of a current actual character in the template.

Other than that, [10] investigated the recognition of the car plate also uses Template Matching in the character recognition process. But, instead of using correlation, [1] used Normalize Cross Correlation (NormCross) to compare the similarity of template images and input images. The equation used as in (2).

$$s(I, Tn) = \frac{\sum_{i=0}^{w} \sum_{j=0}^{h} (I(i,j) - |I|)(Tn(i,j) - |Tn|)}{\sqrt{\sum_{i=0}^{w} \sum_{j=0}^{h} (I(i,j) - |I|)^2 (Tn(i,j) - |Tn|)^2}} \qquad (2)$$

where $s(I, Tn)$ is NormCross values, which I denoted as the image of the segmented character and Tn is the image of an actual character in the template respective to the number *n*. Then, *w* and *h* denoted as the total number of segmented characters and actual characters respectively. i is a current number of segmented characters while j is the current number of the actual character. The $I(i, j)$ and $Tn(i, j)$ are the image of the segmented character and actual character respective to the number i and j .

3. METHODOLOGY

Figure 2 shows the flowchart for the processes involved in recognizing broken character from car plate image.

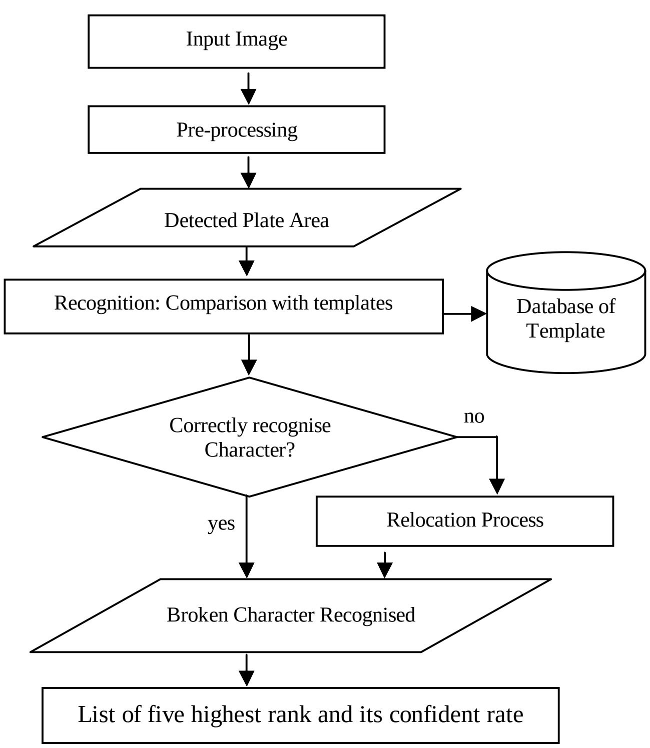


Figure 2: Flowchart of Broken Character Recognition

There are two types of data used is this study that are simulated data and real data. Simulated data is obtained by editing process for the character on the car plate of the original image to create the broken character on the car plate. Besides, the simulated forms of broken character are taken variously to study the performance of the proposed algorithm for various kinds of data. In the meanwhile, real data is the image that already consists of broken character on its car plate originally, the real broken car plate character are taken into consideration as a data collection in order to determine application of the proposed method on real data.

Then, the pre-processing process in detecting and extracting the area of the car plate is applied. There are 4 important steps which are conversion into grayscale images, thresholding process, calculate and selecting maximum area of white region, and decision of succession for car plate extraction.

The process starts with pre-processing, character extraction by using morphological process and CCL, character identification and manual relocation of segmented broken character. Each of the character need to be identified to determine which one of the segmented character is the broken one using (3).

$$NewMeanThreshold = 0.85 \times \frac{\sum_{i=1}^{n} A_i}{n} \tag{3}$$

where, *A* is the width of each segmented character while *n* is the total number of segmented characters. Then the threshold following the distribution in (4) as a decision to identify the broken character was broken character *b* of perfect character *p*.

$$BCI = \begin{cases} p, & if \quad w \geq mT \\ b, & if \quad w \leq mT \end{cases} \tag{4}$$

where, *BCI* is broken character identification, *p* is a perfect character, b is a broken character, while, w is the width of each segmented character and mT is the *NewMeanThreshold* values as in (3).

In the next step, the process of recognition is performed to display all the possibilities of the original character of the broken one using template matching. Template matching is used as it is responsible for object recognition by comparing the acquired image with the image in template [11]. Two different measurements in Template Matching are used which are CORR and SSIM. The CORR is as in (5).

$$r = \frac{1}{n-1} \sum \left(\frac{X - x}{s_x}\right)\left(\frac{Y - y}{s_y}\right) \tag{5}$$

where, r is correlation coefficient of (X_1, Y_1), (X_2, Y_2)…… (X_n, Y_n) observations, n is number of observation, x is mean of X_1, X_2……… . *X* and *y* are mean of $Y_1, Y_2 \ldots Y_n$. S_x is the total observation of *x* and s_y is total observation of *y*.

The SSIM can be expressed in following (6).

$$\rho(x, y) = \frac{(2\mu_x \mu_y + c1)(2\sigma_{xy} + c2)}{(\mu_x^2 + \mu_y^2 + c1)(\sigma_x^2 + \sigma_y^2 + c2)} \tag{6}$$

where, $\rho(x, y)$ is a similarity value between two images *x* and y , In the meanwhile μ_x is mean of template character and μ_y is mean of segmented character. Then, σ_{xy} is co-variance of template and segmented character.

Once the data is failed to recognize, the process goes back to previous process of character segmentation and broken character identification is required. This because of the failure factor of recognizing the character may cause during this process. The Performance Rate (PR) is calculated using (7).

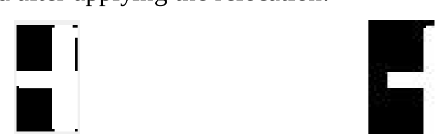
$$PR = \frac{n_{sb}}{N} \tag{7}$$

where n_{sb} is number of car plate that successfully recognized and N is a total number of car plate executed.

Template matching method is used in this study to recognize each of the character. Two different measurements which are CORR and SSIM are applied to provide the alternative of measurement in this study. Along the study, the processes of recognizing all the perfect characters are done successfully. However, there is problem in recognizing the broken character accurately for certain data.

In order to overcome the problem, this study introduced an extra step called as Relocation procedure. This step aids the process for the broken character recognition to produce more accurate result compare than not applying the relocation procedure. Relocation procedure is the process of manually crop the broken character to ensure the size of the box are comparable as close as its perfect character so that it be more relevance to compare with the template.

Figure 3 shows the example result of the broken character before and after applying the relocation.



(a) Segmented character before Relocation of Broken 'H'   (b) Segmented character after Relocation of Broken 'H'

Figure 3: Results of the Broken Character Recognition

Before relocation process, the method detects the broken character as J, but after the relocation process, the method correctly detects the character as H.

At the end, the five highest rank and its confident rate (CNFDT) in percentage (%) for both CORR and SSIM are calculated and a list of 5 possible broken characters with maximum confident rate is listed. The CNFDT% for CORR and SSIM are calculated using (8) and (9) respectively.

$$CNFDT\% = \frac{\sum_{i=1}^{n} r_i}{n} \tag{8}$$

$$CNFDT\% = \frac{\sum_{i=1}^{n} [p(x,y)]_i}{n} \tag{9}$$

where $CNFDT\ \%$ is the confidence rate in percentage, i is the number of segmented character and n is total number of the segmented character.

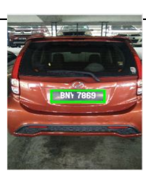
4. RESULTS

This section shows the findings that consist the results of car plate area detection and broken character identification, perfect and broken character recognition and the performance evaluation. This section also indicates that the objectives of the study are achived by implementing the proposed algorithm and evaluating the performance of the study.

The result of the car plate area detection and broken character recognition are divided into two sections which are for the simulated data and real data.

Table 1 shows the results of 5 samples out of 20 simulated data for car plate area detection and broken character identification.

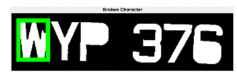
Table 1: Car Plate Area Detection and Detected Broken Character on Samples of Simulated Data

Data	Detected Car Plate Area	Broken Character Identification
S1		
S2		
S3		
S4		
S5		

By observing Table 1, the car plate area of the simulated data is successfully detected as shown in second column of the table. In addition, the location of broken character is also detected as shown in the third column.

Table 2 shows the result of car plate area detection and broken character identification by executing on the real data.

Table 2: Car Plate Area Detection and the Detected Broken Character on Real Data

Data	Detected Car Plate Area	Broken Character Identification
R1		
R2		
R3		
R4		
R5		

For the plate area detection, manual cropping might be required if the method unable to detect the car plate area automatically due to the cause of the undesired disturbance factor such as building, tree and available of other car in the image. Besides that, low intensity of the light is also be the factor that make the car plate area difficult to be detected automatically.

After the car plate area is detected, the broken character for the car plate is detected and recognized. The location of the broken character for the real data is successfully detected correctly.

Since there is possibility the character might be recognized wrongly when compared with the database templates, hence it is suggested to list 5 maximum CORR and SSIM for the recognized broken character.

Table 3 shows the results of character recognition and performance evaluation of 5 samples from simulated data using CORR and SSIM.

The result shows the recognized plate number, the sum of the CORR and SSIM values for each character and the rate of confidence based on the CORR and SSIM values.

Table 3: Character Recognition and Performance Evaluation on Simulated Data

Data	Actual Plate Number	Rank	Recognized Plate Number					
			CORR			SSIM		
			Detected Plate Nmber	SUM	CNFDT %	Detected Plate Nmber	SUM	CNFDT %
S1	 (PNE 7567)	1	PNE 7567	5.039	71.98	PNL 7567	3.322	47.46
		2	PNF 7567	5.021	71.73	PNP 7567	3.319	47.42
		3	PNL 7567	5.002	71.45	PNB 7567	3.282	46.89
		4	PNC 7567	4.944	70.63	PND 7567	3.273	46.76
		5	PNP 7567	4.826	67.19	PNM 7567	3.246	46.37
S2	 (PNE 7567)	1	PNE 7567	4.850	69.28	7NE 7567	2.989	42.70
		2	9NE 7567	4.841	69.15	2NE 7567	2.988	42.70
		3	PNE 7567	4.833	69.05	PNE 7567	2.968	42.40
		4	7NE 7567	4.775	68.21	FNE 7567	2.952	42.17
		5	3NE 7567	4.752	67.88	PNE 7567	2.939	41.98
S3	 (PNE 7567)	1	PNE 7567	4.914	70.20	PNE 7567	3.113	44.47
		2	PNE 7567	4.896	69.95	PNE 7567	3.064	43.78
		3	PNE 75Q7	4.824	66.40	PNE 7507	3.043	43.48
		4	PNE 7567	4.800	68.56	PNE 75O7	3.022	43.17
		5	PNE 75L7	4.798	68.54	PNE 75O7	3.022	43.17
S4	 (BNY 7869)	1	BNY 7869	5.769	82.41	BNY 7869	4.857	69.39
		2	BNY X869	5.557	79.38	BNY Z869	4.587	65.53
		3	BNY Z869	5.544	79.20	BNY X869	4.573	65.33
		4	BNY Y869	5.496	78.51	BNY Y869	4.542	64.88
		5	BNY A869	5.380	76.86	BNY T869	4.511	64.44
S5	 (BNY 7869)	1	BNY 7869	5.613	80.19	BNY 7869	4.432	63.32
		2	BNV 7869	5.561	79.44	BNT 7869	4.336	61.95
		3	BNT 7869	5.366	76.65	BNV 7869	4.309	61.56
		4	BNX 7869	5.279	75.41	BN1 7869	4.195	59.93
		5	BN9 7869	5.275	74.00	BNX 7869	4.179	59.70

From Table 3, the highlighted row shows the highest rank for each simulated data that able to recognize correctly. Majority of the results shows at first rank able to determine the succession of the system toward the respective data. It can be observed that, the first rank using CORR provides the correct recognition as compared with the actual plate number for these 5 samples of the simulated data. While by using SSIM, inconsistent results obtained from the first rank with the highest CNFDT% such as Data S1 and S2. For Data S1, none of the listed 5 rank of recognition results gives the correct recognition. While Data S2, the correct recognition is at the third rank. This is due to the shape of the broken character that closely similar with the wrong template in the database. This shows inconsistency of the results using SSIM. Furthermore, the confidence rate by using CORR is always greater than SSIM for all the simulated data. This shows that, the CORR provide much confident measurement as compared with SSIM. But SSIM can be an alternative for the measurement with some adjustment to improve its accuracy.

Table 4: Character Recognition and Performance Evaluation on Real Data

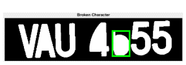
Data	**Data & Actual Plate Number**	**Rank**	**Recognized Plate Number**					
			CORRELATION			**SSIM**		
			Detected Plate Number	**SUM**	**CNFDT%**	**Detected Plate Number**	**SUM**	**CNFDNT%**
R1	 (WB6253S)	1	WB 6253 S	3.400	70.19	WB 625J S	3.637	51.96
		2	WB 625J S	3.264	69.24	WB 6253 S	3.584	51.20
		3	WB 6256 S	3.257	68.54	WB 625S S	3.510	50.14
		4	WB 625S S	3.220	68.32	WB 625U S	3.490	49.86
		5	WB 625U S	3.201	67.57	WB 625A S	3.479	49.69
R2	 (WYP 376)	1	WYP 376	4.080	68.00	WYP 376	2.566	42.77
		2	DYP 376	3.875	64.58	UYP 376	2.378	39.63
		3	AYP 376	3.859	64.32	QYP 376	2.333	38.88
		4	QYP 376	3.838	63.96	8YP 376	2.327	38.78
		5	QYP 376	3.834	63.90	AYP 376	2.324	38.74
R3	 (WAJA 1110)	1	WAJA 1110	6.285	78.56	WAJA 1110	4.381	54.76
		2	WAJA 1190	6.217	77.72	WAJA 1190	4.259	53.23
		3	WAJA 1190	6.135	76.69	WAJA 11P0	4.216	52.70
		4	WAJA 11P0	6.107	76.34	WAJA 1170	4.186	52.33
		5	WAJA 1190	6.074	75.92	WAJA 11Q0	4.182	52.28
R4	 (JJY 653)	1	JJY 653	4.624	77.06	JJY 653	3.436	57.27
		2	UJY 653	4.385	73.09	5JY 653	3.191	53.18
		3	3JY 653	4.337	72.28	UJY 653	3.169	52.82
		4	5JY 653	4.295	71.58	9JY 653	3.135	52.24
		5	9JY 653	4.263	71.05	LJY 653	3.128	52.13
R5	 (VAU 4655)	1	VAU 4655	5.232	74.74	VAU 4655	3.678	52.55
		2	VAU 4655	5.061	72.30	VAU 4655	3.602	51.46
		3	VAU 4655	5.042	72.03	VAU 4855	3.552	50.74
		4	VAU 4455	4.987	71.24	VAU 4655	3.546	50.66
		5	VAU 4455	4.982	71.17	VAU 4055	3.546	50.66

Table 4 shows the character recognition using CORR and SSIM for real data. The highlighted row shows the rank for each data that able to be recognized correctly. Majority of the results shows at first rank able to determine the succession of the system toward the respective data. It can be observed that the first rank using CORR provides the correct recognition as compared with the actual plate number. While by using SSIM, the results give the correct recognition at the first rank except for Data R1.

In order to evaluate performance of the recognition process with three aspects of PR are calculated based on the succession of the car plate area detection, succession of the character recognition by CORR and succession rate for character recognition by SSIM. Table 5 shows the results of the PR toward the three mentioned aspects.

Table 5: The PR of three Criteria

Criteria	**Car plate area detection**	**Character recognition**	
		CORR	**SSIM**
Total number of data	25	25	25
Succesfull image detect/recognise automatically	21	25	19
PR	84%	100%	76%

For detection of the car plate area, 21 out of 25 data can be detected automatically while the other 4 are detected by using manual detection. For the recognition result, all of 25 data are successfully recognized by using the CORR while 19 out of 25 data are successfully recognized by SSIM.

5. CONCLUSION

It can be concluded that the recognition of the broken character on the car plate is completed by several processes involved that are data collection, pre-processing for car plate area detection, character segmentation, and character recognition by comparing with database of template by template matching. Relocation process is required if the character is unable to be recognized correctly. By implementing the proposed processes, the problem of recognition regarding to available data can be overcome by the process of character relocation before the recognition process. All the objectives set up in this study are achieved that are the proposed processes are successfully implemented in simulated data and real data. Besides that, the broken character is successfully recognized using CORR and SSIM. For each of the possibilities of the car plate number, the confidence rate are evaluated, and the result shows that the recognition by using CORR is better than using SSIM measurement because of the confidence rate for CORR is higher than SSIM for each data. In addition, the performance rate for the whole processes also evaluated and the result of three different categories of succession rate are also shown. In the result, the succession rate for the car plate area detection is 84%. In the meanwhile, for the recognition process, the succession rate for CORR is 100% while 76% for SSIM.

ACKNOWLEDGEMENT

The author would like to express our gratitude to Faculty Computer and Mathematical Sciences, Universiti Teknologi MARA Shah Alam for the support and opportunity given in completing this research.

REFERENCES

1. M. Droettboom. **Correcting Broken Character in the Recognition of Historical Printed Document**, 2003 Joint Conference on Digital Libraries, pp. 364-366. Houston, 2003. https://doi.org/10.1109/JCDL.2003.1204889
2. V.Novak, P.Hurtik, and H. Habiballa. **Recognition of Distorted Characters Printed on Metal using Fuzzy Logic Methods**. 2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), 2013. https://doi.org/10.1016/j.jal.2014.11.003
3. K. Kaur, and A. K. Bathla. **A Review on Segmentation of Touching and Broken Characters for Handwritten Gurmukhi Script**. *International Journal of Computer Applications,* vol. 120, no. 18, 2015. https://doi.org/10.5120/21326-4279
4. T.C. Sandanayake, A.M. Bandara. **Automated Classroom Lecture Note Generation Using Natural Language Processing and Image Processing Techniques**. *International Journal of Advanced Trends in Computer Science and Engineering, vol. 8, no. 5, pp. 1920-1926, 2019*. https://doi.org/10.30534/ijatcse/2019/16852019
5. K. Ramalingam, and R. Bhojan. **Identification of Broken Characters in Degraded Documents**. *International Journal of Intelligent Engineering & System*, vol. 11. 2017. https://doi.org/10.22266/ijies2018.0630.14
6. N. N. Ansari, A. K. Singh, and M. Student. **License number plate recognition using Template Matching**. *International Journal of Computer Trends and Technology (IJCTT)*, vol. 35, pp. 175-178, 2016. https://doi.org/10.14445/22312803/IJCTT-V35P133
7. M. Abdullah, M. A. H. Bakhtan and S. A. Mokhtar. **Number Plate Recognition of Malaysia Vehicles using Smearing Algorithm**, *Science International (Lahore)*, vol. 29, no. 4, pp. 823-827, 2017.
8. Y. J. Choong, L. K. Keong and T. C. Cheah, **License Plate Number Detection and Recognition using Simplified Linear-Model**, *Journal of Critical Reviews*, vol. 7, no. 3, pp. 55-60, 2020. https://doi: 10.31838/jcr.07.03.09
9. P. Gupta, G. Purohit, and M. Rathore, **Number Plate extraction using Template Matching Technique**. *International Journal of Computer Applications*, vol. 88, no. 3, 2014. https://doi.org/10.5120/15336-3670
10. W. Devapriya, C. N. K. Babu, and T. Srihani, **Indian License Plate Detection and Recognition using Morphological Operation and Template Matching**, *World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering*, vol. 9, no. 4, 2015. https://doi.org/10.5281/zenodo.1338967
11. I. T. Plata, A. C. Taracatac, E. B. Panganiban, **Development and Testing of Embedded System for Smart Detection and Recognition of Witches' Broom Disease on Cassava Plants using Enhanced Viola-Jones and Template Matching Algorithm**, *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 5, pp. 2613-2621, 2019. https://doi.org/10.30534/ijatcse/2019/113852019