Abstract. The development of the number of online shopping consumers in Indonesia is increasing. In 2016, research from e-Marketer estimated that it would reach 8.6 million people who shop through the internet. This figure increased from 7.9 million people in the previous year. The increasing number of people who know the internet along with the birth of generation Z (Gen Z) in the digital era makes changes in shopping habits that previously conventionally turned into online. Online transaction is an automatic ticket number that is used to differentiate transactions for each buyer. For example in the purchase transaction of goods in the online shop, each purchase transaction is distinguished through an additional three digits behind the nominal to facilitate payment into the seller's account. The problem faced by the owner of the online shop is when BRI internet banking login to see account mutations, user authentication by entering the Captcha code. This process is repeated because there is a duration limitation for access to the system. It is necessary to check the account mutation automatically on a computer system that is able to read and recognize the character of the Captcha image. In this study, we propose the Template Matching method in recognizing characters in the captcha image. The dataset contains 300 images taken from the BRI internet banking website. The first step is to convert the extension from .png to .jpg. After the image is converted to extension .jpg the next step is pre-processing to correct the image of the noise that exists. Pre-processing result images that have been fixed are quality, then the separation of objects and background using Otsu. Then the process of labelling and segmentation. The characters are then processed using the Template Matching method so that they can be identified as numbers. The results of the experiments that have been carried out, found that the method we propose is able to recognize characters in the image with an accuracy rate of 90%. Errors in character recognition are due to characters that intersect, so when the results are processed the results do not match the ground truth.

1. Introduction
The development of the number of online consumers in Indonesia is increasing. In 2016, research from e-Marketer estimated that the number would reach 8.6 million people who shop through the internet. This figure increased from 7.9 million people in the previous year. With the increasing number of people who know the internet as the birth of generation Z (Gen Z) that was born in the digital era makes the habit of shopping for goods and services that before conventionally will turn into
online[1]. In online shopping transactions there are automatic tickets that will be used to differentiate transactions that have entered. For example in the purchase transaction of goods on the website tokopedia.com, each purchase transaction will be distinguished through an extra amount of nominal price called an automatic ticket (Code Unique) to help transfer payments to the seller's account. The large number of transactions that occur makes it difficult for online shop owners to check and find out the balance mutations that have been entered or not because they have to look back on transactions one by one which results in the length of service of a transaction and does not rule out the possibility of errors in the checking process. Another problem faced by online shopping owners is when Bank BRI internet banking login when checking whether the transfer from the buyer has entered the account is user authentication by entering the CAPTCHA code[2]. So, from these problems, it is necessary to check mutations automatically on a computer system that is able to read and recognize the character patterns of the CAPTCHA image. One popular way to recognize characters in various types of documents is Optical Character Recognition (OCR) with the Template Matching method which is expected to read and re-recognize the forms or patterns of CAPTCHA character images correctly. Character recognition of plate number of vehicle number, using a total of 30 sample data with a total of 238 characters with a success rate of 80.25%[3]. Template Matching is a technique in digital image processing to find small parts of images that match the image template[4]. Template matching method is one of the methods used to explain how our brains recognize forms or patterns. In this case, it can be a solution for online shop users to recognize this Captcha image character which aims to provide convenience in checking balance mutations automatically.

2. Related Work

Security of authentication and transactions on internet banking in Indonesia there are those who use CAPTCHA imagery as a confirmation tool that the user who is actively logging in or transaction is not a robot / hacking machine[5]. A survey about CAPTCHA security was carried out by Soni, et al. [6]found that there were several types of CAPTCHA that could be used for security systems and how to recognize characters from these types were also different. Optical Character Recognition (OCR) character recognition method is needed in CAPTCHA images to be able to log in to internet banking automatically[7]. Generally, there are two approaches to conducting OCR, an image processing approach with template matching techniques and with artificial intelligence techniques. Template Matching Correlation has been tested on CAPTCHA images with a variety of noise variations, type and font size, and degree of slope but not tested on CAPTCHA images that have coinciding text[9]. Template matching has also been tested on car number plate image, with 30 sample data and 238 characters with a success rate of 80.25%, but there is no text on the image of the car license plate. With an artificial intelligence approach, the research conducted by Ye Wang and Mi Lu novel adaptive algorithm managed to recognize characters in CAPTCHA images with an average value of 70.78% accuracy[10]. Sliding window based on neural networks is also carried out by Hussain, et al with a success rate of 95.5% in character recognition[8].

3. Methodology

The proposed system process begins with the selection of captcha image datasets. The dataset is retrieved from Bank BRI's internet banking website, then the image is converted to an extension from .png to .jpg. After the image is converted to extension .jpg the next step is pre-processing to correct the image of the noise that exists. The pre-processing results that have been repaired are of quality, then positioned between the object and the background using the Otsu thresholding method. Objects that have been separated from the background are segmented and subsequently propagated per character. The characters that have been obtained are then processed using the Template Matching method so that it can be identified as a number that matches the template image that has been prepared in advance. And the last is to calculate the accuracy of character recognition results. Figure 1 shows how this system will work.
3.1. Pre-Processing
The pre-processing process is needed to improve the quality of the image being analyzed. The image is pre-processed in the form of 1 channel only (red / green / blue). In this study, the selected channel is a green channel, which is then improved. Figure 2 shows the difference in image quality of each channel and pre-processing image.

![Figure 2](image)

**Figure 2.** Quality of color difference every channel: (a) Red, (b) Green, (c) Blue

![Figure 3](image)

**Figure 3.** Threshold Background segmentation and object:
(a) T=0.01, (b) T=0.05, (c) T=0.1
3.2. Segmentation Object and Background
The next process in this study is the separation of the number and background character objects that are in the CAPTCHA. In accordance with the characteristics of the previously mentioned image that the image has a colour gradation background from white to black. So to separate the background and object to be recognized is to use the Otsu thresholding method. This method can distinguish objects and backgrounds, so that it will be obvious that the object is black and the background is white. Following are examples of the results of the separation of objects and background. The optimal T value can be determined by try and error and is presented in Figure 3.

3.3. Segmentation
After getting the point of interest in the previous process, the next step is segmentation per character of each number. Number character segmentation with labelling on cropping results with the size of each character number measuring 14 x 30 pixels, is presented in Figure 4.

3.4. Character Recognition
After getting the object of each number (each captcha image has 4 numbers), the next step is to recognize the number starting from 0 to 9. The method used to recognize characters is Template Matching. The number of characters that have been obtained is then correlated with a database of small and large size numbers of 20 images. The character image of cropping results which has the greatest correlation value with the existing number database is a representation of the actual number.

4. Result
The number of CAPTCHA images tested as many as 50 taken from BRI internet banking page. The purpose of this test is to find out the success and failure of the system to measure accuracy, and later conclusions can be drawn according to observations from the test. Example 5 the results of the first and second trial tests are presented in Tables 1 and 2. The matching of values using a statistical approach and for the character identification process using a template matching algorithm where the match value is taken from the number of pixels that are close to or even the same for each character in the CAPTCHA image. The value of the pixel count is obtained using uninterrupted binary pixel image analysis. The value of the number of pixel characters will process with the Template Matching algorithm. Next will be known the number of pixel characters that match the number of pixels of the template dataset character. Accuracy results from match detection test data with template datasets totaling 100 test data. Accuracy evaluation is done by calculating the confusion matrix model can be seen in table 2. The high accuracy of character recognition in CAPTCHA imagery due to the first experiment is at least characters in overlapping images, whereas in the second experiment with CAPTCHA images that occur overlap each character, making accuracy decreases. This is shown in Figure 4.

5. Conclusion
Character recognition in CAPTCHA images of BRI bank internet banking identification has a success rate of 93.5%. The mistake of recognizing a lot of characters is caused by characters that behave in the image so that in the process of segmentation and cropping occurs objects that overlap. This results
in the process of Matching Templates being less good because the area that is not an object is considered an object so that it can affect the value of the correlation between objects with database templates that result in error detection of characters. In subsequent studies, it is expected that there will be improvements in the segmentation and cropping processes so that objects that do not overlap, or with artificial intelligence approaches so that the system is more robust against CAPTCHA image variations.

Figure 5 Binary Image and Segmentation per character (a) Binary Image (b) First Character Image, there is an additional white area on the right so that it is detected as number 9 (c) second character image, there are more white areas on the left, but numbers are still detected 0 (d) Third character image, there is an additional white area from the fourth pixel image fragment (e) fourth character image, there is an additional white area from the third-pixel image fragment.

Table 1. First Testing Result

| No | Image | Biggest pixel Detected | All Character Detected | Recognize | True | False |
|----|-------|------------------------|------------------------|-----------|------|-------|
| 1  | Y     | Y                      | 7639                   | 4         | 0    |       |
| 2  | Y     | Y                      | 3710                   | 4         | 0    |       |
| 3  | Y     | Y                      | 1476                   | 4         | 0    |       |
| 4  | Y     | Y                      | 6243                   | 4         | 0    |       |
| 5  | Ya    | Ya                     | 3852                   | 4         | 0    |       |
Table 2. Second Testing Result

| No | Image | Biggest pixel Detected | All Character Detected | Recognize | True | False |
|----|-------|-------------------------|------------------------|-----------|------|-------|
| 1  | ![Image](image1.png) | Y | Y | 7500 | 3 | 1 |
| 2  | ![Image](image2.png) | Y | Y | 3 97 | 3 | 1 |
| 3  | ![Image](image3.png) | Y | Y | 3 07 | 3 | 1 |
| 4  | ![Image](image4.png) | Y | Y | 306 | 3 | 1 |
| 5  | ![Image](image5.png) | Y | Y | 9094 | 3 | 1 |

6. References

[1] Riswandi B A 2005 *Aspek hukum internet banking* (Jakarta: PT. Raja Grafindo Persada)
[2] Yan J and El-Ahmad A S 2009 *IEEE Security & Privacy* 7 22–8
[3] Kurniawan B S, Sentinuwo S R and Lantang O A 2016 *Journal Teknik Informatika* 8 7-12
[4] Zou H, Zhang B, Tao Z and Wang X 2016 *Journal of Physics Conference Series* 680 art ID 012001
[5] Pitchay S A, Nabihah N, Suhaimi M and Saudi M M 2017 *Proceedings of the World Congress on Engineering and Computer Science* (San Fransisco: International Association of Engineers)
[6] Soni S and Bonde P 2017 *Int. J. Comput. Appl* 68 44-7
[7] Hamad K A and Kaya M 2016 *International Journal of Applied Mathematics, Electronics and Computers* 4 244-49
[8] Hussain R, Gao H, Shaikh R A and Soomro S P 2016 *IEEE International Conference Communication Software and Networks* (Beijing: IEEE Conference)
[9] Sakkatos P, Theerayut W, Nuttapol V and Surapong P 2014 *International Conference Information and Communication Technology, Electronic and Electrical Engineering* (Chiang Rai: IEEE Conference)
[10] Wang Y and Lu M 2016 *Proceeding IEEE International Conference Industrial Technology* (Taipei: IEEE Conference)