1. Introduce & Subject
2. Existing CAPTCHAs and Problems
3. Approaches
   1. Variety Method

We suggest 4 methods to generate our own captcha. Noise addition, noise addition with other color, data removal and data reverse are them. Our goal is to find the method to generate captcha that human can recognize easily but computer cannot. Therefore, we are plan to compare the results of each method, how much human and computer can recognize. Details about each method are explained in section D.

* 1. Length

The Breaking-Captcha projects available on the Internet are mainly aimed at breaking Captcha presented on a specific website. In other words, even if only the number of captcha characters is changed, the breaking success rate of the project is greatly reduced. However, we aim to find a general anti-breaking captcha methodology that does not heuristically invalidate certain breaking captcha projects. While generating our captcha, therefore, we fix the length of original captcha letter in 4 with only alphabets.

* 1. Font size

The Breaking-Captcha projects available on the Internet are mainly aimed at breaking Captcha presented on a specific website. In other words, even if only the font of captcha is changed, the breaking success rate of the project is greatly reduced. However, we aim to find a general anti-breaking captcha methodology that does not heuristically invalidate certain breaking captcha projects. While generating our captcha, therefore, we fix the font size of original captcha letters in 55px.

* 1. Position

The Breaking-Captcha projects available on the Internet are mainly aimed at breaking Captcha presented on a specific website. In other words, even if only the size of the captcha image is changed, the breaking success rate of the project is greatly reduced. However, we aim to find a general anti-breaking captcha methodology that does not heuristically invalidate certain breaking captcha projects. While generating our captcha, therefore, we fix the position of original captcha letters in the center of the image and the size of the image into 200px by 150px.

1. 3 Methods

For color blind users, we decided on the most contrasting white and black as background color and letter color. And also in the Addition Color method, we have used only one additional color for the color fading users, and we selected between red and blue with a large contrast.

* 1. Addition

The most similar form to the existing captcha generating algorithm is to add noise to the original captcha image to confuse the machine. There are three stages depending on the degree of noise addition.

* 1. Addition (with color)

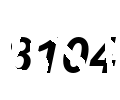
This is similar to the Addition method, but there is a distinction to add noise in a whole new color. This is based on an increase in people's recognition rate when the same amount of noise is added, but with a completely new color. There are three stages depending on the degree of noise addition.

* 1. Removal

This is the method based on space supplementation effects. Human can deduce data not given from peripheral data even if some data is not given.

When we create a captcha by removing it rather than by adding noise, there is a possibility that the direct features to predict the captcha may be removed. In this case, the computer cannot find the feature to recognize the captcha so the recognition rate will be reduced, however, the human can show almost the same recognition rate using the space supplement effect. There are three stages depending on the degree of data removal.

* 1. Reverse

This is also the method based on space supplementation too. It is possible for human to arbitrarily interpret data according to the tendency.

Therefore, we inverted the part of the data in a fixed shapes (circle). There are three stages depending on the degree of noise addition. Because the position of the captcha letters was fixed, it did not reveal the greatest advantages of this method. The advantage of this method was that it was difficult to tell which one was the actual captcha character or noise, but the computer was not confused with the assumption. There are three stages depending on the number of circles.

1. Experiment
   1. Dataset
      1. Overview

Dataset Name: Dataset\_InfoSecCaptcha\_200x150\_80000.zip (1.7GB)

Number of total samples: 400000

Number of methods: 5 (Original, Addition, Addition\_color, Removal, Compliment)

Level in each method: 3 (Low, Medium, High except the Original method)

Directory Structure:  
(Except the method Original) Generated Date - Method - Level - Captcha Images  
(Method Original) Generated Date - Method - Captcha Images

* + 1. Description

There are three levels of pattern domain images. These pattern domains are used when we generate captchas with the method Addition, Addition\_color and Removal. Actually, the pattern is extracted from the pattern domain, with the size 200px by 150px that matched with the captcha images.

In the method Addition and Addition\_color, the black pixels in the pattern image will be added as a noise. In the method Removal, in the contrast, pixels at the position that is black in the pattern will be removed.

The type of the captcha image is PNG, and the name of the captcha image is generated by following format, “{answer}\_{method}\_{level}\_{index}.png”

* 1. Against Human
     1. WEB Environment
     2. Result Table
  2. Against AI
     1. CNN Classifiers
     2. OCR
     3. Result Table

1. Result
   1. Equation to compute the goodness of each method

Score = (Accuracy of Human - Accuracy of Computer) \* Accuracy of Human

If the method that accuracy of human is higher than computer is better. And also, the situation that human cannot recognize the capture should not occur. Therefore, we both consider the metric accuracy of human and the difference between both accuracy.

* 1. Result Table

1. Conclusion
2. References
   1. Project GitHub Link  
      https://github.com/dshyun0226/InfoSecCaptcha