Stereoscopic Text-based CAPTCHA on Head-Mounted Displays

Tadaaki Hosaka and Shinnosuke Furuya
School of Management, Tokyo University of Science, Chiyoda-city, Tokyo, Japan

Keywords: CAPTCHA, Image Authentication, 3D Vision, Stereo Matching, Virtual Reality.

Abstract: Text-based CAPTCHAs (completely automated public Turing test to tell computers and humans apart) are widely used to prevent unauthorized access by bots. However, there have been advancements in image segmentation and character recognition techniques, which can be used for bot access; therefore, distorted characters that are difficult even for humans to recognize are often utilized. Thus, a new text-based CAPTCHA technology with anti-segmentation properties is required. In this study, we propose CAPTCHA that uses stereoscopy based on binocular disparity. Generating a character area and its background with the identical color patterns, it becomes impossible to extract the character regions if the left and right images are analyzed separately, which is a huge advantage of our method. However, character regions can be extracted by using disparity estimation or subtraction processing using both images; thus, to prevent such attacks, we intentionally add noise to the image. The parameters characterizing the amount of added noise are adjusted based on experiments with subjects wearing a head-mounted display to realize stereo vision. With optimal parameters, the recognition rate reaches 0.84; moreover, sufficient robustness against bot attacks is achieved.

1 INTRODUCTION

Various types of CAPTCHA (completely automated public Turing test to tell computers and humans apart) have been proposed (Roshanbin and Miller, 2013; Hasan, 2016), all of which present a task, which is easy for humans but highly difficult for machines to perform, to the user requesting authentication. In typical text-based CAPTCHAs, some distorted letters or digits are displayed, which must be input into the response field by the user.

However, there have been advances in image segmentation and character recognition techniques to break CAPTCHAs (Bursztein et al., 2014; Gao et al., 2016; Chen et al., 2017); therefore, it was necessary to present characters with a large degree of shape distortion. Consequently, such authentication sometimes became difficult even for humans to pass. Furthermore, recent developments in machine learning are remarkable and most CAPTCHAs can be broken if segmentation is correctly performed (George et al., 2017; Ye et al., 2018); thus, it is important to realize anti-segmentation. Therefore, new text-based CAPTCHA that does not overly deform characters and has anti-segmentation properties is required.

In this study, we propose text-based CAPTCHA that uses binocular-vision-based stereoscopy, which has not been previously considered in this research field, for technologies that use three-dimensional (3D) applications. The proposed method uses stereo images that contain a character in front of a wall-like background. By using the same color pattern for the character and background, segmentation of either left or right image into these two regions becomes impossible, which is a major advantage of our method over conventional CAPTCHA methods. In plain stereo images that enable stereoscopic viewing based on binocular disparity, character regions can be extracted using disparity estimation or background subtraction; the proposed method intentionally adds noise to incapacitate such image processing. The validity of our proposed method is confirmed by experiments with subjects wearing a head-mounted display (HMD) to realize stereo vision.

2 PREVIOUS RESEARCH ON TEXT-BASED CAPTCHA

Extensive research has been conducted on text-based CAPTCHA. Large IT enterprises such as Microsoft (Chellapilla et al., 2005) and Google (Kluever and Zanibbi, 2009) have actively conducted research on CAPTCHAs. In particular, reCAPTCHA, currently provided by Google, has been introduced

Hosaka, T. and Furuya, S.
Stereoscopic Text-based CAPTCHA on Head-Mounted Displays.
DOI: 10.5220/0008893407670774
In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 5: VISAPP*, pages 767-774
ISBN: 978-989-758-402-2; ISSN: 2184-4321

on many websites. Although the latest version of reCAPTCHA, i.e., reCAPTCHA v3, evaluates the user's behavior on the website as a score and determines whether the user is a bot, traditional text-based CAPTCHAs are still used on numerous websites.

Most traditional text-based CAPTCHA methods enhance robustness against bot attacks by using techniques such as character distortion, rotation, twisting, and overlap (Roshanbin and Miller, 2013; Hasan, 2016). However, as characters are increasingly transformed, it becomes more difficult for humans to recognize them. To solve this problem, three-dimensional (3D) text-based CAPTCHAs have been proposed (Macias and Izquierdo, 2009; Imsamai and Phimoltares, 2010; Ye et al., 2014). In these methods, characters as 3D objects are deformed by operations such as rotation and projected onto a two-dimensional (2D) plane corresponding to the user's monitor. The validity of these methods is based on the fact that humans can perceive 3D space even in a 2D image.

Research on methods to break CAPTCHAs has also been actively conducted (Bursztein et al., 2014; Gao et al., 2016; Chen et al., 2017). Typical breaking techniques for text-based CAPTCHA consist of three parts: preprocessing, segmentation, and recognition. Preprocessing is performed to facilitate the subsequent segmentation or recognition process; typical methods include binarization, noise removal, and thinning. Segmentation aims to detect the boundaries of the presented characters by measuring the size of the connected regions, performing vertical projection, and so on. Representative techniques of the recognition process include template matching, clustering, and machine learning such as support vector machines (Starostenko et al., 2015) and deep learning (Stark et al., 2015).

With regards to machine learning, recent developments in the techniques have been remarkable. George et al. proposed the recursive cortical network (RCN), inspired by the human brain having the ability to learn and generalize from a few examples (George et al., 2017). RCN is a hierarchical model that represents the contours and surfaces of characters, and allows character recognition without requiring much learning data. Their results showed that the recognition rate for reCAPTCHA database is 0.943 for letters and 0.666 for words. Ye et al. proposed a method to break CAPTCHAs using generative adversarial network (GAN) which is a method of generating data necessary to train deep neural networks (Ye et al., 2018). Their method firstly trains two competing networks corresponding to a generator and discriminator. The generator tries to create a CAPTCHA, which is visually similar to the target CAPTCHA. The discriminator then tries to determine if the synthesized CAPTCHAs are genuine. After the learning of the two networks has progressed sufficiently, the generator can synthesize a CAPTCHA indistinguishable from real ones and the deep neural networks can be trained using these images. Chen et al. proposed a method to further improve the recognition accuracy of deep convolutional neural network for confusion classes (Chen et al., 2019).

The development of such techniques to break CAPTCHAs defeats most methods when character and background segmentation is potentially possible. 3D CAPTCHAs can also be broken with high probability along with traditional text-based CAPTCHAs (Ye et al., 2014; Bursztein et al., 2014; Gao et al., 2016; Chen et al., 2017). Therefore, it is important to realize new text-based CAPTCHA that does not overly deform the characters and has anti-segmentation properties.

In this study, we propose text-based CAPTCHA that uses binocular-vision-based stereoscopy. By using the same color pattern for the character and background, it becomes impossible to segment either the left or right image into character and background regions. In our study, HMDs are used to make it easier for the subject to view the image stereoscopically, which may seem to narrow the applicability of the proposed method. However, most conventional text-based CAPTCHAs are intended for use on personal computers with a large screen and a keyboard and are not always useful in VR or MR environments where users have an HMD and/or handheld controllers (Bhat et al., 2015; Singh and Singh, 2017). It is worthwhile to propose a secure authentication method that can be used even when the user wears devices specialized for VR or MR environments. As an example, Yang et al. proposed a CAPTCHA method by letting the user play a game on a handset that had only a small screen without a keyboard (Yang et al., 2015). By investigating the behavior during the game, it was possible to determine whether the user is a human or a bot. Similar to their proposed CAPTCHA that takes advantage of the characteristics of a handset, we realize new CAPTCHA by effectively using stereovision provided by an HMD.

3 PROPOSED CAPTCHA BASED ON BINOCULAR VISION

To realize a new kind of text-based CAPTCHA, we utilize stereo images in which the characters have the same color pattern as that of the background. Furthermore, noise is intentionally added to the stereo im-

Figure 1: Example of background image. This image consists of only eight colors (each RGB component is either 0 or 255). In this research, this image is utilized as the background for both the left and right images, creating a wall at infinite distance.

ages to enhance robustness against supposed attacks by bots.

The procedure of the proposed stereo image generation is as follows:

1. Generate background
2. Draw characters in front of the background
3. Add noise for robustness against stereo matching
4. Add noise for robustness against background subtraction.

The details of each process are described below.

3.1 Generating Background

We first generate two identical images of size 285 × 241. For each pixel, each RGB component is randomly determined to be either 0 or 255. These images are composed of eight colors, as shown in figure 1. These two images are arranged horizontally as a stereo image pair, creating a wall as background at infinite distance. Even though the background has zero disparity in this research, the following argument holds even for general cases of non-zero disparity.

3.2 Drawing Characters in Front of Background

To place a character in front of the background, we add a letter with disparity k to the stereo image pair. RGB combinations of each pixel in the character region are randomly selected to have one of the eight colors used for the background. Therefore, it is in principle impossible to recognize the character by viewing a left or right image alone. The character can be recognized only as a perception of depth using binocular vision, as shown in figure 2.

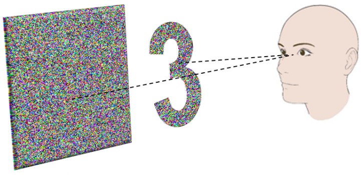


Figure 2: Diagram of how humans can see embedded character. In stereo images, the character region has disparity k and the background has zero disparity. By looking at the left and right images with the left and right eyes, respectively, the character appears to float above the background to a human even though it cannot be recognized when viewing the left or right image alone.

It is in principle possible to present multiple characters for the CAPTCHA. However, in this research, stereovision using an HMD is considered and only one character is displayed because the size of the monitor on an HMD is limited. By performing multiple tests, identification with an accuracy same as when presenting multiple characters simultaneously can be realized.

3.3 Adding Noise for Robustness against Stereo Matching

Although it is not possible to segment each image into the character region and background, several possible attacks on the proposed method are considered.

One possible attack on the above method is to extract the character region using disparity estimation via stereo matching. Therefore, to make disparity estimation difficult, every pixel in each row of the right image is shifted by $+1$ or -1 in the x direction with probability $a/2$. With this operation, for a pixel whose RGB values become indefinite at the left or right end in each row, we randomly define the color again. However, this operation is ineffective against $m \times 1$ $(m \in \mathcal{N})$-sized block matching. Therefore, a similar operation is applied to the y-axis; every pixel in each column of the right image is shifted by $+1$ or -1 in the y direction with probability $b/2$. As the parameters a and b increase, the robustness to stereo matching increases but stereoscopic viewing becomes more difficult for humans. It is thus necessary to optimize these parameters.

3.4 Adding Noise for Robustness against Background Subtraction

The other possible attack on the proposed method is to extract the character region by shifting the left or right image by the amount of the background disparity (zero in the present case, but a non-zero value in general) and taking the difference of the two images. An example of a difference image for a stereo image pair in which the character "3" is embedded is shown in figure 3(a). This subtraction operation can extract the character part as the region with a non-zero difference (depicted as white pixels).

To prevent this, one of the RGB values of each pixel in the stereo image pair is redetermined as 0 or 255 with probability c. The difference image for the stereo image pair after this process is shown in figure 3(b). This process generally increases the number of pixels with differences and is thus not useful for hiding the character region, which indicates that erosion and dilation processing can be used to extract the character region from the background.

Therefore, we add further noise to the stereo image pair to disturb the character region of the difference image. To achieve this, the RGB values of each pixel in the character region of the left image are copied to the identical coordinates in the right image with probability d. The aim of this operation is to intentionally set the difference value in the character region of the left image to zero. An example of a difference image after this process is shown in figure 3(c), which indicates that this process can work well.

3.5 Parameter Adjustments

The four parameters a, b, c, and d need to be adjusted so that stereoscopic viewing by a human is possible and character recognition by a machine cannot be easily performed. Furthermore, the four parameters need to be properly balanced because if even one of them is too large, binocular vision by a human becomes highly difficult. In contrast, if one of the parameters is close to 0 with the other parameters kept not too large, extraction of the character regions by image processing becomes easy. Although it is necessary to moderately increase all parameters, it is difficult to theoretically obtain the optimal values. The optimal combination of parameters was thus determined using experiments with subjects.

In this study, experiments were conducted in a situation where the subject could view the image stereoscopically by wearing an HMD. The proposed method is also applicable to other devices that enable stereoscopic vision, which include polarized or active shutter glasses (used in cinemas, etc.), lenticular lenses or parallax barriers (used in naked-eye 3D systems), and traditional anaglyph glasses. In addition, with a little training, some people can see the right image with their right eye and left image with their left eye without using a special device. For such people, the proposed method can be used as an alternative to typical 2D CAPTCHAs.

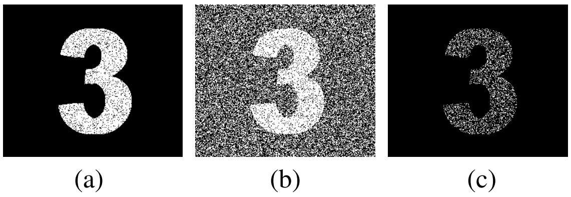


Figure 3: Character extraction by background subtraction. White pixels represent non-zero difference and black pixels represent zero difference. (a) Difference image for a stereo image pair not subjected to any processing; the character region can be easily extracted. (b) Difference image for the stereo image pair processed with $c = 0.5$ (other parameters set to zero); although the background is contaminated by noise, simple erosion and dilation operations can be used to extract the character region. (c) Difference image for the stereo image pair processed with $d = 0.7$ (other parameters set to zero); noise is added to the character region. Adjusting both c and d is necessary for enhancing robustness against background subtraction.

4 EVALUATION EXPERIMENTS

4.1 Experimental Settings

Evaluation experiments were conducted with a total of 20 participants. Stereo image pairs corresponding to the seven sets of parameters shown in table 1 were presented and viewed through an HMD designed for VR headsets. As the value of each parameter increases, the difficulty level of stereoscopic vision rises and robustness against stereo matching and background subtraction is enhanced. Therefore, images with parameter set 1 were the easiest to perceive and least resistant to attacks, and images with parameter set 7 were the most difficult to perceive and most resistant to attacks. Since parameter b more strongly and negatively influences the stereoscopic view than parameter a in most cases, the value of parameter b is set smaller than that of a.

In each stereo image pair, a digit from 0 to 9 was embedded using the method described in the previous section with disparity $k = 10$. Examples of stereo image pairs for some parameter sets are shown in figure 4. Subjects viewed these images through an HMD having a divider at the center (figure 5) and were asked to state the number they perceived by stereopsis. Subjects first optimized their viewing setup in terms of visibility of targets by adjusting the focal length and the distance between lenses using plain stereo images corresponding to the parameter set $(a,b,c,d) = (0,0,0,0)$. With the optimized setup,

Table 1: Parameter sets in our experiments.

No.	a	b	c	d
1	0	0	0	0
2	0.1	0.1	0.1	0.1
3	0.3	0.1	0.1	0.1
4	0.3	0.15	0.15	0.15
5	0.3	0.2	0.15	0.15
6	0.3	0.2	0.3	0.2
7	0.3	0.3	0.3	0.3

the subject sequentially viewed images with various parameters.

Some people could not stereoscopically view the numbers even under parameter set 1; they were excluded from the evaluation because the proposed CAPTCHA is expected to be utilized in VR or MR applications and it is reasonable to assume that such users have at least plain stereoscopic vision.

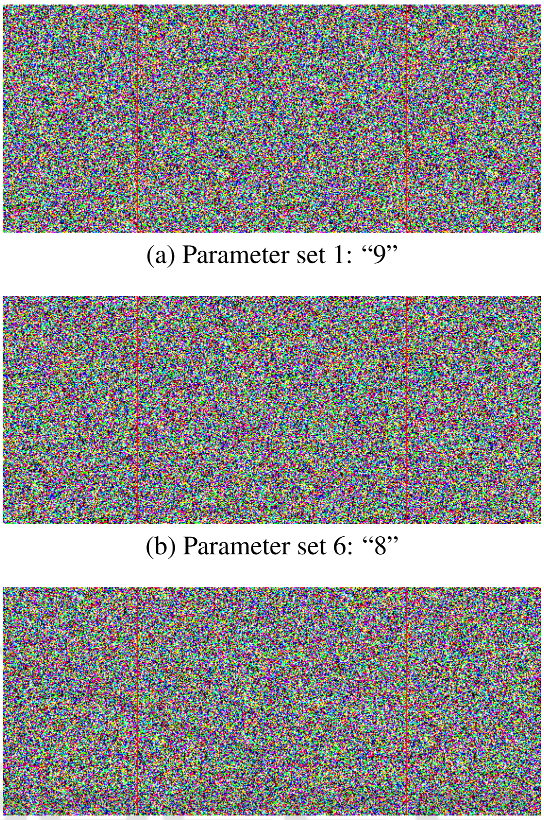
For each subject, two tests each for parameter sets 1-5 and four tests each for parameter sets 6 and 7 were conducted. For each test, correct/incorrect, response time, and confidence of answer were recorded. The degree of confidence was self-assessed by subjects according to the following criteria:

3: Subjects can clearly perceive the whole character as if it were floating on the background.

2: Subjects can recognize the character without relying on estimation, but the clarity of the stereoscopic vision is inferior to that at level 3.

1: Subjects can estimate the character from some perceivable information such as piecewise contours.

The key point of levels 2 and 3 is that the subject does not rely on estimation. Subjects learned the degree of level 3 by stereoscopically viewing the image generated with parameter set 0; this degree was used as a reference for level 2. When the subject gave an incorrect digit or could not even estimate the digit, the degree of confidence was not recorded.



(a) Parameter set 1: “9”

(b) Parameter set 6: “8”

(c) Parameter set 7: “2”

Figure 4: Examples of stereo image pairs. The left and right images were concatenated for subjects to view through an HMD designed for VR headsets. By adjusting their viewing setup so that the vertical red lines drawn at the center of the left and right images overlap, the subjects could easily perform binocular vision with the parallel-eyed stereo method. The character (digit) embedded in each stereo image is given in the caption.

Figure 5: Head-mounted displays utilized in our experiments. (Right) This HMD device has a divider between two lenses, so subjects could easily perform the parallel-eyed stereo method.

4.2 Experimental Results for Human Recognition

The accuracy rates for the seven parameter sets are shown in figure 6. For parameter sets 1-4, which make it relatively easy to recognize the character, the correct answer rate was 1. In practical applications, considering that another image can be presented if an incorrect answer is given, an accuracy rate of 0.8 or more, as obtained for parameter sets 5 and 6, seems sufficient. For parameter set 7, with $(a,b,c,d) = (0.3,0.3,0.3,0.3)$, the correct answer rate sharply decreased to 0.45. Each parameter value should thus be less than 0.3.

Figure 7 shows the average response time for subjects who responded correctly. To eliminate the influence of extremely rapidly or slowly responding subjects, the top and bottom 20% of response times were

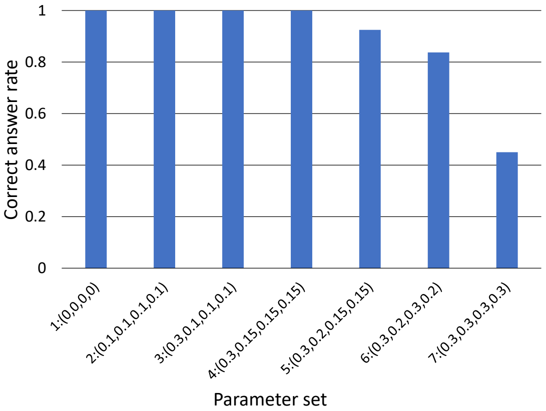


Figure 6: Correct answer rate (recognition rate). For parameter sets 1-4, every subject gave the correct answer for all tests. The recognition rate slightly decreased for parameter sets 5 and 6, and dropped sharply (under 0.5) for parameter set 7.

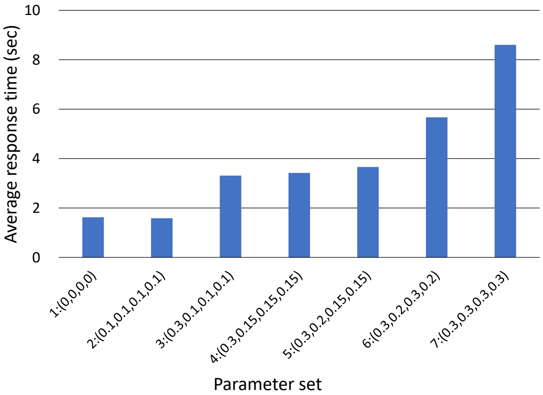


Figure 7: Average response time. Some subjects could immediately give an answer and some took more than one minute to answer. To mitigate the influence of these extreme cases, the top and bottom 20% of response times were excluded from the average calculation of each parameter set. As the difficulty of recognition increases, response time increases.

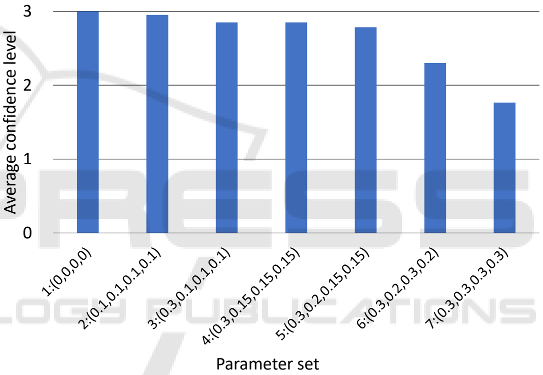


Figure 8: Average confidence degree for answers. For parameter sets 1-5, a high degree of confidence was observed. The confidence level dropped sharply, but remained above 2, for parameter set 6, indicating that recognition was made without estimation.

omitted from the average calculation. The experimental results show that as the values of parameters a, b, c, and d increase, the response time tends to increase. In future versions of the proposed method, presenting multiple characters is conceivable, and thus the recognition speed per character should be high. From this viewpoint, the stereo image pairs generated with parameter set 7 are too stressful for users in practical situations.

Figure 8 shows the average confidence level for subjects who responded correctly. The confidence degree for parameter sets 1-6 exceeds level 2 (characters are perceptible without estimation). For parameter set 7, the confidence level is lower than 2, confirming that this set is unsuitable for practical applications.

The above results indicate that parameter sets 1-6 are appropriate from the viewpoint of human recognition performance.

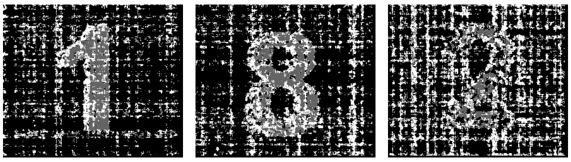
4.3 Robustness against Bot Attacks

Robustness against possible attacks using stereo matching and background subtraction was investigated.

Disparity maps for parameter sets 5-7 obtained using stereo matching with 5 × 5 blocks are shown in figure 9. To generate a binary image representing the foreground and background, pixels with an estimated disparity of 0 or ±1 are black and pixels with an estimated disparity of more than 1 are white in consideration of the fact that the correct disparity of the background is zero and that a shift of one pixel may occur in relation to parameter a. To extract the character region from this binary image, we performed erosion and dilation operations, in which the order and number of erosion and dilation processes were manually tuned based on the grid search method. The following two types of results are shown in figure 10:

- (Left) As many pixels as possible belonging to the character region are retained as white pixels.
- (Right) As many pixels as possible belonging to the background region are retained as black pixels.

These results qualitatively show that parameter sets 6 and 7 have sufficient robustness against character region extraction; when we try to retain the character region as the foreground, approximately the same or more mispredicted white pixels appear in the background and when we attempt to eliminate these background noise, the character part cannot be recognized.



(a) Param. set 5 (b) Param. set 6 (c) Param. set 7

Figure 9: Disparity maps estimated using stereo matching with 5×5 blocks. The true disparity values are 10 in the character region and 0 in the background region. Estimated disparity values are amplified by a factor of 10, and their absolute values are expressed as brightness in these maps. As the values of parameters increase, more noisy disparity maps are obtained.

Robustness can be increased in future work by adaptively varying the disparity of the background.

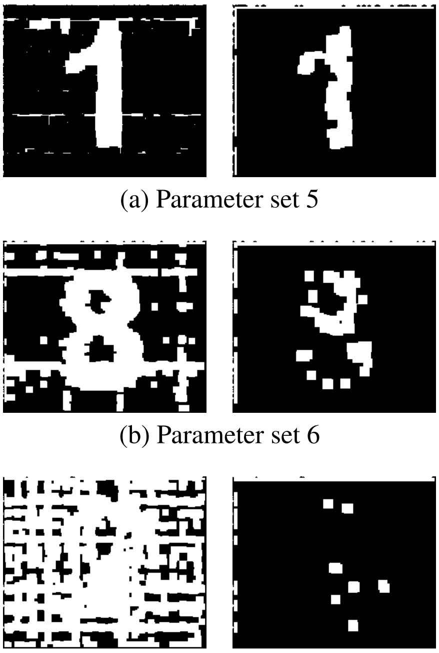
Figure 11 shows the binarized subtraction images for parameter sets 5-7, where pixels with identical RGB values for the left and right images are black, and pixels that are different in any one RGB component are white. We attempted to extract the character region by performing erosion and dilation operations, as done above. The results are shown in figure 12. The images generated with parameter sets 6 and 7 have significant tolerance against attacks, which may be further enhanced by making the character and background structures more complicated. For parameter set 5, it is speculated that recognizing the character by machines is possible.

The above results indicate that parameter sets 6 and 7 are appropriate from the viewpoint of robustness against bots attacks. Furthermore, together with the results of human recognition performance described in the previous subsection, we conclude that parameter values of approximately $(a,b,c,d) = (0.3,0.2,0.3,0.2)$ are suitable for practical use.

5 CONCLUSIONS

This paper proposed a text-based CAPTCHA that uses stereopsis based on binocular vision. To enhance robustness against supposed bot attacks using stereo matching and image subtraction, we manipulated the pixel values, which correspond to artificial noise addition. In evaluation experiments with subjects wearing an HMD, users achieved a recognition rate of more than 0.8 without resorting to speculation for images that were resistant to attacks. It is speculated that by making the scene depth structure more complicated, the proposed stereo images can become even more robust against bot attacks.

Future work will include verification of our method with incorporation of traditional text-based CAPTCHA techniques, such as multiple letters, multiple strings, overlap, distortion, and adhesion. As another direction, robustness against breaking techniques based on sophisticated machine learning such

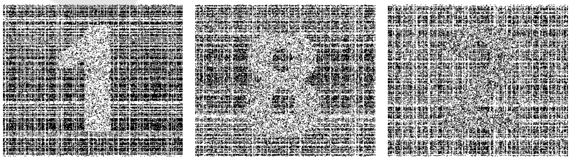


(a) Parameter set 5

(b) Parameter set 6

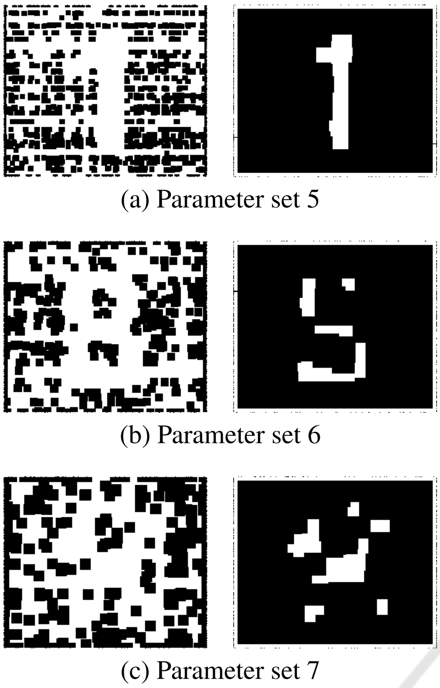
(c) Parameter set 7

Figure 10: Extraction of the character region from binarized disparity maps using erosion and dilation operations. (Left) The order and number of erosion and dilation processes were adjusted so that many pixels in the actual character region were maintained as the foreground. (Right) These parameters were adjusted so that as many mispredicted white pixels as possible in the actual background region was removed. It seems possible for machines to recognize the character with parameter set 5 because most of the predicted foreground (white) pixels are concentrated in the true character region. With parameter sets 6 and 7, it seems difficult for machines to recognize the character because collecting white pixels only in the character region seems difficult.



(a) Param. set 5 (b) Param. set 6 (c) Param. set 7

Figure 11: Binarized subtraction images. Pixels with matching RGB values are shown in black, and pixels that are different in any one RGB component are shown in white. As in the disparity maps, the degree of noise is increased in order, which means that extracting the character region using image subtraction becomes more difficult as the values of parameters increase.



(a) Parameter set 5

(b) Parameter set 6

(c) Parameter set 7

Figure 12: Extraction of the character region from subtraction images using erosion and dilation operations. (Left) The order and number of erosion and dilation processes were adjusted so that many pixels in the actual character region were maintained as the foreground. (Right) These parameters were adjusted so that as many mispredicted white pixels as possible in the actual background region was removed. As in the disparity maps, with parameter set 5, it seems possible for machines to extract the character region by adjusting the erosion and dilation parameters. This seems difficult with parameter sets 6 and 7, for which the digits in the results are highly difficult to recognize, even for humans.

as support vector machine and convolutional neural networks should be investigated.

REFERENCES

Bhat, A., Bhagwat, G., and Chavan, J. (2015). A survey on virtual reality platform and its applications. *International Journal of Advanced Research in Computer Engineering & Technology*, 4(10):3775–3778.

Bursztein, E., Aigrain, J., Moscicki, A., and Mitchell, J. C. (2014). The end is nigh: Generic solving of text-based captchas. In *Proceedings of 8th USENIX Workshop on Offensive Technologies (WOOT 14)*.

Chellapilla, K., Larson, K., Simard, P. Y., and Czerwinski, M. (2005). Designing human friendly human interaction proofs (hips). In *Proceedings of the 2005 Conference on Human Factors in Computing Systems*, pages 711–720.

Chen, J., Luo, X., Guo, Y., Zhang, Y., and Gong, D. (2017). A survey on breaking technique of text-based captcha. *Security and Communication Networks*, page 6898617.

Chen, J., Luo, X., Liu, Y., Wang, J., and Ma, Y. (2019). Selective learning confusion class for text-based captcha recognition. *IEEE Access*, 7:22246–22259.

Gao, H., Yan, J., Cao, F., Zhang, Z., Lei, L., Tang, M., Zhang, P., Zhou, X., Wang, X., and Li, J. (2016). A simple generic attack on text captchas. In *Proceedings of Network and Distributed System Security Symposium (NDSS)*.

George, D., Lehrach, W., Kansky, K., Lázaro-Gredilla, M., Laan, C., Marthi, B., Lou, X., Meng, Z., Liu, Y., Wang, H., Lavin, A., and Phoenix, D. S. (2017). A generative vision model that trains with high data efficiency and breaks text-based captchas. *Science*, 358(6368):eaag2612.

Hasan, W. (2016). A survey of current research on captcha. *International Journal of Computer Science Engineering Survey*, 7:1–21.

Imsamai, M. and Phimoltares, S. (2010). 3d captcha: A next generation of the captcha. In *Proceedings of 2010 International Conference on Information Science and Applications*, pages 1–8.

Kluever, K. A. and Zanibbi, R. (2009). Balancing usability and security in a video captcha. In *Proceedings of the 5th Symposium On Usable Privacy and Security*.

Macias, C. R. and Izquierdo, E. (2009). Visual word-based captcha using 3d characters. In *Proceedings of the 3rd International Conference on Crime Detection and Prevention*, pages 1–5.

Roshanbin, N. and Miller, J. (2013). A survey and analysis of current captcha approaches. *Journal of Web Engineering*, 12:1–40.

Singh, N. and Singh, S. (2017). Virtual reality: A brief survey. In *Proceedings of 2017 International Conference on Information Communication and Embedded Systems (ICICES)*, pages 1–6.

Stark, F., Hazrba, C., Triebel, R., and Cremers, D. (2015). Captcha recognition with active deep learning. In *German Conference on Pattern Recognition Workshop*.

Starostenko, O., Cruz-Perez, C., Uceda-Ponga, F., and Alarcon-Aquino, V. (2015). Breaking text-based captchas with variable word and character orientation. *Pattern Recogn.*, 48(4):1101–1112.

Yang, T.-I., Koong, C.-S., and Tseng, C.-C. (2015). Game-based image semantic captcha on handset devices. *Multimedia Tools and Applications*, 74(14):5141–5156.

Ye, G., Tang, Z., Fang, D., Zhu, Z., Feng, Y., Xu, P., Chen, X., and Wang, Z. (2018). Yet another text captcha solver: A generative adversarial network based approach. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 332–348.

Ye, Q., Chen, Y., and Zhu, B. (2014). The robustness of a new 3d captcha. In *Proceedings of 11th IAPR International Workshop on Document Analysis Systems*, pages 319–323.