An Image-Based CAPTCHA Using Sophisticated Mental Rotation

Yuki Ikeya, Masahiro Fujita, Junya Kani, Yuta Yoneyama, and Masakatsu Nishigaki

Graduate School of Informatics, Shizuoka University, Japan
nisigaki@inf.shizuoka.ac.jp

Abstract. As one of the advanced Completely Automated Public Turing tests to tell Computers and Humans Apart (CAPTCHAs), the CAPTCHA using mental rotation has been proposed. Mental rotation is an advanced human-cognitive-processing ability to rotate mental representations of "one" single 2D/3D object. However, as have already been reported, the mental rotation CAPTCHA can be overcome by pattern matching and/or machine learning. Therefore, this paper proposes to enhance the mental rotation CAPTCHA by using "two" distinct 3D objects in the task of mental rotation, which we call "sophisticated mental rotation". We implemented a prototype of the sophisticated mental rotation CAPTCHA, and carried out basic experiments to confirm its usability. Also, we conducted a comparison between the proposed CAPTCHA and existing CAPTCHAs. The obtained results were satisfactory.

Keywords: CAPTCHA, Mental Rotation, 3DCG.

1 Introduction

With the expansion of Web services, denial of service (DoS) attacks by malicious automated programs (malwares) are becoming a serious problem. Thus, the Turing test is becoming a necessary technique to discriminate humans from malicious automated programs and the Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) [1] system developed by Carnegie Mellon University has been widely used. The simplest CAPTCHA presents distorted or noise added text (Fig. 1) to users who visit Web sites and want to use their services. We refer to this simple CAPTCHA as text recognition based-CAPTCHA. If they can read the given text, they are certified as human. If they cannot read the text, they are certified to be malwares.

However, many researchers have recently pointed out that automated programs with optical character reader (OCR) and/or machine learning can answer those conventional text recognition based-CAPTCHA [2]. Indeed, these sophisticated malwares have been spreading and they have cracked the text recognition based-CAPTCHA [3, 4]. It can be made more difficult for automated programs to pass tests (i.e. read texts) by increasing the distortion or noise. However, it also becomes more difficult for humans to read such texts. We therefore need to adopt even more advanced human cognitive processing capabilities to enhance CAPTCHA to overcome this problem.

T. Tryfonas and I. Askoxylakis (Eds.): HAS 2014, LNCS 8533, pp. 57–68, 2014.

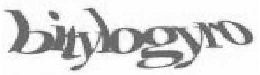
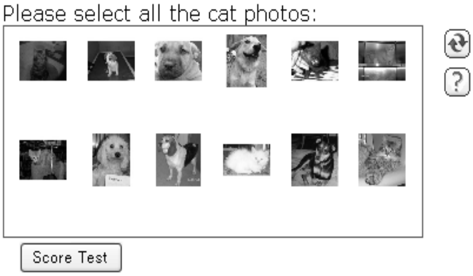


Fig. 1. CAPTCHA used by Google



Please select all the cat photos:

Fig. 2. Asirra

Image recognition based-CAPTCHA such as Asirra [6] (Fig. 2) is known as one of the effective solutions for enhancing CAPTCHA, because image recognition is a much more difficult problem for automated programs than character recognition [5]. Labeled images are used in the image recognition based-CAPTCHA to confirm that a user can recognize the meaning of the image. In Asirra, several photos of animals (e.g. images of cats and dogs with diverse backdrops, angles, poses, and lighting) are presented to a user, and the user is then asked to select a specific animal in a test. For example, suppose that the user is asked to select “cat”; if he/she can select all photos labeled as cat in the test, then he/she is certified to be human. If not, he/she is certified to be an automated program.

However, a technique that has effectively been used to breach the image recognition based-CAPTCHA has been reported and shocked researchers [7, 8]. Advancements to cracking capabilities (CAPTCHA cracking algorithms and CPU processing speeds) will continue indefinitely. No matter how advanced malicious automated programs are, a CAPTCHA that will not pass automated programs is required. Hence, we have to find another human cognitive processing capability to tackle this challenge.

As one of the interesting possibilities to deal with this challenge, mental rotation has been used in YUNiTi's CAPTCHA [11] (Fig. 3). Mental rotation [9, 10] is the advanced human-cognitive-processing ability to rotate mental representations of two-dimensional (2D) and/or three-dimensional (3D) objects. In YUNiTi's CAPTCHA, Web page visitors need to choose an appropriate object from a candidate image list matching the same 3D object as the question image. However, a report suggested that this 3D CAPTCHA could be vulnerable to template matching attack [12]. The CAPTCHA using only simple rotation of “one” single 3D object is not safe enough.

Therefore, this paper proposes to enhance the mental rotation CAPTCHA. The approach taken in this paper is make the task of mental rotation more complex by using “two” distinct 3D objects. We call this enhanced mental task “sophisticated mental rotation”. Our CAPTCHA is expected to improve the decrypting tolerance for automated programs without noticeable degradation in understandability for humans.

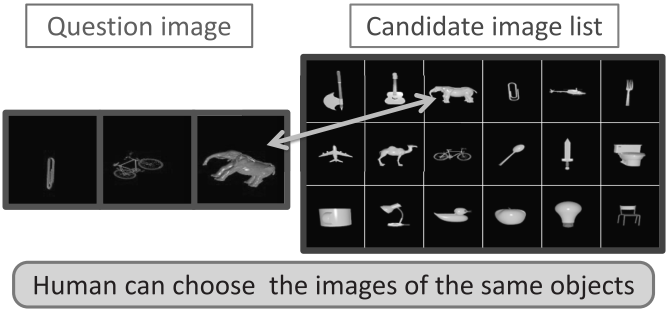


Fig. 3. YUNiTi's 3D CAPTCHA

2 Mental Rotation and YUNiTi's CAPTCHA

Humans are good at spatial reasoning capacity. For this reason, it is not difficult for humans to understand the three-dimensional (3D) shape of the object from the two-dimensional (2D) image. This kind of "ability to recognize 3D objects from 2D image" is considered to be an advanced human-cognitive-processing ability [14]. Also, it is possible for humans to rotate 2D/3D objects in an imagination and to recognize the shape figure, which has been photographed from a different point of view. This human ability is called “mental rotation” [9, 10]. Therefore, by looking at the 2D images of the two sheets of copies from different viewpoints of one single 3D object, humans infer the shape of the 3D object in the 2D images, and understand the change of the viewpoint.

Mental rotation is used in YUNiTi's CAPTCHA [11] (Fig. 3). In this CAPTCHA, Web page visitors need to choose appropriate objects from a candidate image list (containing 18 objects) by matching the same 3D objects as the question image (containing 3 objects). If they can choose all the correct 3D objects corresponding to each of three question images, they are certified as human. If they cannot, they are rejected as automated programs. The question images are automatically generated by randomly selecting a 3D object from the candidate image list and then photo shooting the object from different viewpoints.

However, it has been reported that this 3D CAPTCHA could be vulnerable to template matching attack [12]. In YUNiTi's CAPTCHA, all 3D objects which are used in the candidate image list are unchanged and immutable. Therefore, an attacker can collect a variety of question images shot from all angles for each object in a limited number of CAPTCHA trials, and exploit them for template matching attack and/or machine learning attack. Additionally, automated programs (malwares) can utilize the technology of three-dimensional object recognition. In the current technology, it is possible to figure out almost exactly the 3D shape of the 3D object from two images, which are taken of a 3D object from two viewpoints [13]. In YUNiTi's CAPTCHA, the response to the CAPTCHA test is to select an answer image among the candidate images. This means that malwares may be able to identify the correct object by restoring the 3D shape of the correct object from two images displayed in a CAPTCHA test (a question image and the corresponding image in the candidate image list). That is why the CAPTCHA using only simple rotation of one single 3D object is not safe enough.

3 Sophisticated Mental Rotation CAPTCHA

3.1 Concept

In this paper, we propose to use two distinct 3D objects in the task of mental rotation (which is referred to as “sophisticated mental rotation” in this paper). As long as these two distinct objects are the semantically same as each other, it is expected that this task is still not too difficult for humans, but enhances the safety by increasing complexity of analysis by automated programs (malwares).

Fig. 4 and Fig. 5 show the overview of the sophisticated mental rotation CAPTCHA. Sophisticated mental rotation CAPTCHA is presented with a pair of 2D images (the “question image” and its “response image”) of two distinct 3D objects shot from two different viewpoints. In the question image, a marker (like a red sphere) is added to any portion of the 3D object. There is no marker in the response image. The user is then asked to click the location on the response image, which is corresponding to the position where the marker in the question image is located. If the user is a human, he/she can use (sophisticated) mental rotation to identify the correct position of the response image.

On the other hand, malware can utilize the technologies of pattern matching, machine learning, and three-dimensional recognition. However, these technologies are all basically designed for finding the “visually identical” object to a target object. In contrast, in the sophisticated mental rotation proposed here, two different 3D objects (to be more precise, two objects are semantically same but “visually different” from each other) are being used. Therefore, it can be expected that it is still difficult even for these technologies to overcome the task of sophisticated mental rotation. Alternatively, if an appropriate level of transformation/deformation is applied against two images, it is also expected that three-dimensional shape identification for malwares become markedly difficult. Therefore, in this paper, we consider two types of systems; two distinct 3D objects are produced by deforming one 3D model (type-α) (Fig. 4), or by two different 3D models (type-β) (Fig. 5).

Generally speaking, for automated programs, producing 2D images from the 3D object is significantly easier than identification of 3D objects from a 2D image. In (sophisticated) mental rotation CAPTCHA, this “one-way property” contributes to the automatic generation of a pair of the question image and its response image. As shown in Fig. 6, automated programs (Web servers) can generate new images using 3D computer graphic technology every time and achieve the automatic generation of the pair of images. The pair of images can be generated innumerably by registering a large number of 3D models with a system and changing some parameters such as the object, size of the object, marker position, viewpoint, and so on.

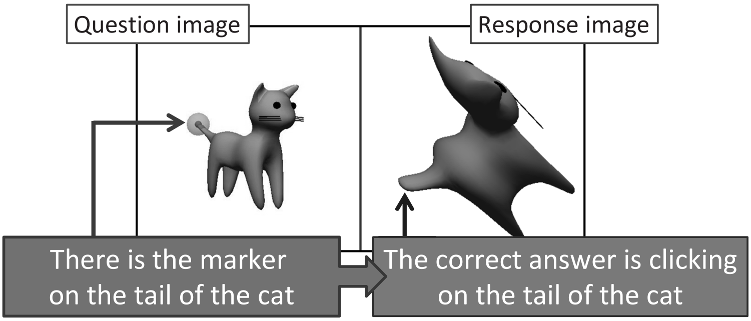

Question image
Response image
There is the marker on the tail of the cat
The correct answer is clicking on the tail of the cat

Fig. 4. Sophisticated mental rotation CAPTCHA (type-α)
(Left: question image; Right: response image)

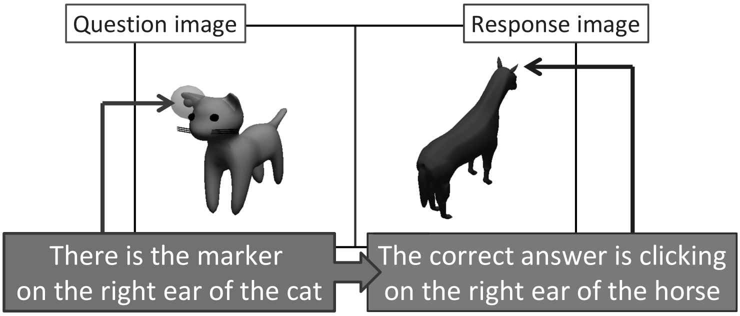

Question image
Response image
There is the marker on the right ear of the cat
The correct answer is clicking on the right ear of the horse

Fig. 5. Sophisticated mental rotation CAPTCHA (type-β)
(Left: question image; Right: response image)

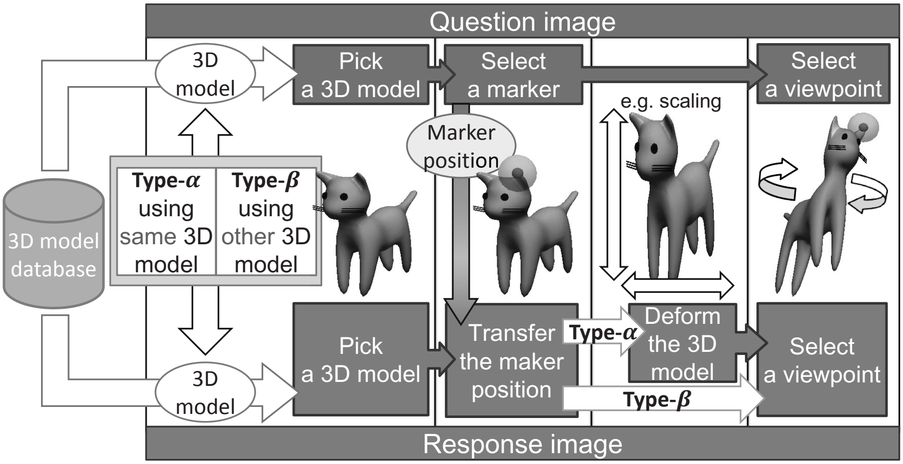


Fig. 6. Automatic generation procedure in sophisticated mental rotation CAPTCHA

3.2 Authentication Procedure

Authentication procedure of the sophisticated mental rotation CAPTCHA is as follows. It is here assumed that the sophisticated mental rotation CAPTCHA system has a 3D model database, in which enough number of 3D models are archived.

Step1. The system picks up a 3D model for the question image (defined to as "question object") at random.

Step2. The system randomly selects a marker position on the question object.

Step3. The system randomly selects a viewpoint of the question image.

Step4. The system generates the question image, by photo shooting the question object (selected in step 1) with the marker (selected in step 2) from the viewpoint (selected in step 3).

Step5. The system picks up the question object (selected in step 1) again as a 3D object for the response image (defined to as "response object") (type-α), or the system picks up another 3D model for the response object at random (type-β).

Step6. The marker position (selected in step 2) in the question object is transferred to the response object (selected in step 5). That is, the marker positions are quite identical (type-α), or semantically identical (type-β) between the question object and the response object.

Step7. In the case of type-α, the system deforms the response object randomly.

Step8. The system randomly selects a viewpoint of the response image.

Step9. The system generates the response image, by photo shooting the response object (selected in step 5 (and deformed in step 7)) without the marker from the viewpoint (selected in step 8). Note that the response object has the marker (set in step 6), but no marker is visually shown.

Step10. The system shows a user (Web page visitor) a pair of the question image and the response image.

Step11. The user clicks a position of the invisible marker in the response image.

Step12. If the clicked position of the response image is correct, the user is identified as a human, and if the position is incorrect, the user is identified as a malware.

Fig. 4 shows an image example of type-α. A response object (right of Fig. 4) is generated from the question object (left of Fig. 4) by scaling at any magnification independently in each of the x/y/z direction. In this paper we use the affine transformation as a deformation processing in step 7, but some other deforming may also be able be applicable.

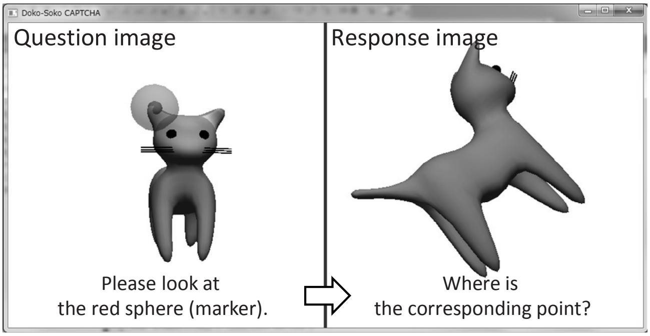
Fig. 5 shows an image example of type-β. In Fig. 5, the question object is a cat (left of Fig. 5), and the response object is replaced with a horse (right of Fig. 5). To achieve the transfer of the marker position from the question object to the response object in step 6, the system needs a database in which the relationship between the parts of all objects is described.

In the sophisticated mental rotation CAPTCHA, it is difficult for malwares to identify the invisible marker position in the response image by using only the question image (with a marker) and response image (without a marker). On the other hand, our system knows the marker position in the response image in step 6. Because this knowledge forms a trapdoor, our system (Web server) can automatically generate the challenges that malwares cannot answer, and then the system can determine whether the user (Web page visitor) clicked the correct position. By randomly choosing the object, the position of the marker, and the position of the viewpoint every time of the authentication, the system with a large amount of 3D models can automatically generate a myriad of challenges. Therefore, the sophisticated mental rotation CAPTCHA is expected to be also resistant to template matching attack and/or machine learning attack.

3.3 Implementation

We implemented a prototype sophisticated mental rotation CAPTCHA (type-α). Fig. 7 shows an authentication screen example of our CAPTCHA: the question image is in the left of Fig. 7; the response image is the right of Fig. 7. The red sphere that is drawn on the question image is the marker. The Web page visitor needs to identify and click the location on the response image, which corresponds to the position of the marker on the question image. If the distance between the clicked position and the correct position is the threshold value or less, he/she is certified as human. In the example of Fig. 7, as the marker is pointing to the right ear of the cat in the question image, it is correct if the visitor clicks the right ear of the cat in the response image.

Note here that the coordinate of the (invisible) maker position in the response object is a 3D data, whereas a mouse click by the user is a 2D data (because it is obtained as the coordinate information on the display). Therefore, the system computes the 2D coordinates of the marker on the display from the 3D marker position. In this implementation, the correct answer range (threshold value) is a circle with a 30-pixel radius.



Please look at the red sphere (marker).

Where is the corresponding point?

Fig. 7. Prototype sophisticated mental rotation CAPTCHA (type-α) (Left: question image; Right: response image)

4 Verification Experiment

We conducted basic experiments to evaluate the authentication rate of the proposed method (type-α). In addition, after the experiment, we did a survey on subjects for usability. Due to time constraints, the prototype system implementation and the experiment have yet to be performed on the sophisticated mental rotation CAPTCHA (type-β).

4.1 Experiment Method

The subjects included twenty volunteers, subjects 01-20, who all are college students in the Faculty of Informatics and Faculty of Engineering at Shizuoka University. Each subject solved five challenges of the sophisticated mental rotation CAPTCHA in a row. In this experiment, the first and the second trials were treated as a tutorial. We only evaluated the remaining three trials.

Five 3D objects (A - E) were used in the experiment. In the tutorial, the object A and B were shown in this order. In the following three trials, the object C – E appeared in random order. The subjects were told to answer according to the center of the sphere (the small opaque sphere inside rather than the big translucent outer sphere). For each challenge, we recorded success or failure, response time, and the click position.

After completing all of the CAPTCHA challenges, we had the subjects respond to the following questionnaires. Question 1, 3, 5 were answered on a 5-point scale.

Question 1. Is it easy solving the CAPTCHA? (Easy) : Yes (5) – No (1)

Question 2. If you chose 1 or 2 in Question 1, please write why you think that it is not easy.

Question 3. Is it user-friendly? (User-friendly) : Yes (5) – No (1)

Question 4. If you chose 1 or 2 in Question 3, please write why you think that it is not user-friendly.

Question 5. Is it pleasant? (Pleasant) : Yes (5) – No (1)

Question 6. If you chose 4 or 5 in Question 5, please write why you think that it is pleasant.

Question 7. How many challenges would you be able to consecutively solve? Also, please write why you think that.

Question 8. Which would you choose: text recognition based-CAPTCHA or sophisticated mental rotation CAPTCHA in at real Web service? Also, please write your reason.

4.2 Experiment Results

Correct Response Rate and Response Time. The experiment results are shown in Table 1, which summarizes the correct response rate and the average response time for each subject. Because the order of the 3D object was random in this experiment, we show the results summarized in experimental order in Table 2 and by object in Table 3.

Table 1. The experiment results for each subject

	Subject									
	01	02	03	04	05	06	07	08	09	10
Correct response rate	2/3	2/3	3/3	3/3	3/3	2/3	3/3	2/3	1/3	3/3
Average response time [sec]	6.7	5.3	3.9	8.7	3.9	8.9	6.5	4.6	4.3	5.4
	Subject									
	11	12	13	14	15	16	17	18	19	20
Correct response rate	2/3	2/3	2/3	3/3	2/3	3/3	2/3	1/3	1/3	2/3
Average response time [sec]	7.1	3.9	3.2	3.7	4.8	6.4	3.4	5.9	6.1	4.5
	Average									
Correct response rate	73.3% (44/60)									
Average response time [sec]	5.4									

Table 2. The experiment result of each order

	First time	Second time	Third time
Average correct response rate	60.0% (12/20)	80.0% (16/20)	80.0% (16/20)
Average response time [sec]	5.9	5.5	5.2

Table 3. The experiment result of each 3D object

	Object C	Object D	Object E
Average correct response rate	80.0% (16/20)	55.0% (11/20)	85.0% (17/20)
Average response time [sec]	6.1	4.6	5.4

From Table 1, the correct response rate of the sophisticated mental rotation CAPTCHA is 77.3% on average (a total of 60 times, 44 successes, 16 failures). The correct response rate is too low for practical use. Future improvements are necessary. As can be seen from Table 2, the correct answer rate for the first trial is the lowest. Thus the correct response rate is expected to increase, as a user gains familiarity. As Table 3 shows, the correct response rate depends strongly on which 3D object is used.

We analyzed why the subjects failed. There are three main reasons. The first is a mistake due to confusing left and right sides of the 3D object shown in the response image. There were five failures by this reason among the 16 total failures. The second reason is a slight deviation of the click position. There were four failures by this reason. If we increase the radius of the correct answer range (threshold value) to 35 pixels from 30 pixels, then the correct response rate rises to 80.0%. In the future, we will consider the appropriate correct answer range. The third reason is a difficulty in the recognition of the depth information of the image. In particular, when the 3D object is viewed from just in front, behind, above, beside, or under, the depth is difficult to be grasped and it is more likely to make a mistake. Since incorrect recognition of depth is associated with the other two reasons, further examination is required on the selection of viewpoint.

From Table 1, the average response time per challenge is 5.4 seconds; the shortest time is 3.2 seconds, and the maximum time is 8.9 seconds. The expected response time for the text recognition based-CAPTCHAs is around 10 seconds at the most. Therefore, it can be said that the proposed CAPTCHA can be solved in a shorter time compared with the text recognition based-CAPTCHA. As Table 2 shows, the response time is not dependent on the execution order. As can be seen from Table 3, there is a difference in response time by 3D object. We will study the trends of 3D objects that can be solved in a short time.

Usability. The results of the survey are shown in Table 4.

In Questions 1, most subjects answered 4 (5 if easy), and the average value was 3.3. The subjects who answered difficult (1 or 2) were asked to write the reason in Question 2. Several reasons are as follows: viewing in three dimensions is difficult; understanding of left and right of object is difficult; it is difficult as the place to click (in the response image) is not visible.

In Questions 3, most subjects answered 5 (5 if user-friendly), and the average value was 4.3. The subjects who answered user-hostile (1 or 2) wrote the following reasons in Question 4: considering the structure of the solid object is troublesome; it is troublesome if it has to be solved every time.

In Questions 5, most subjects answered 4 (5 if pleasant), and the average value was 4.3. The subjects who answered pleasant (4 or 5) wrote the following reasons in Question 5: using the image is fun; it is fun like a game; it is interesting because it tests spatial reasoning capacity.

In Question 7, most subjects, 13 people, answered "three challenges". The second most common answer, chosen by five subjects, was "two challenges". There was one subject who answered one and four challenges respectively. This result indicates that many subjects do not feel like solving the CAPTCHA four or more times consecutively. The primary reasons include the following: "I would fail if there are too many", "It is troublesome and takes too long". In addition, there were also opinions saying, "An appropriate number of challenges will vary with the importance of Web services", and "I think even one challenge is painful if I grow older".

Table 4. Result of survey

	Subject																				Average
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	
Q1	2	2	4	4	4	5	4	4	2	2	4	3	3	4	2	4	4	3	2	3	3.3
Q3	4	4	4	2	5	5	4	5	5	1	5	5	5	3	5	5	4	5	4	5	4.3
Q5	4	4	4	4	4	3	5	3	5	4	5	4	4	4	5	5	5	5	4	5	4.3
Q7	3	3	2	1	3	2	4	2	2	2	3	3	3	3	3	3	3	3	3	3	2.7
Q8	T	T	M	M	M	M	M	M	M	M	M	M	T	T	M	M	M	M	T	M	-

Q1. Easy: Yes (5) – No (1)
Q3. User-friendly: Yes (5) – No (1)
Q5. Pleasant: Yes (5) – No (1)
Q7. How many questions?
Q8. Which would you choose?
(T: Text recognition based-CAPTCHA,
M: proposed CAPTCHA)

In Question 8, five subjects chose the text recognition based-CAPTCHA, while 15 subjects chose the sophisticated mental rotation CAPTCHA. We believe subjects who felt inconvenience of the text recognition based-CAPTCHA chose the sophisticated mental rotation CAPTCHA. The main reasons for choosing the text recognition based-CAPTCHA are as follows: text recognition based-CAPTCHA is easier to understand; it can be done with only the keyboard; the sophisticated mental rotation CAPTCHA requires time in order to answer correctly. The main reasons for choosing the sophisticated mental rotation CAPTCHA are as follows: text recognition based-CAPTCHA is more difficult; it can be done with only the mouse; it is pleasant to use a picture.

5 Conclusion and Future Work

In this paper, we propose the sophisticated mental rotation CAPTCHA, which is an image recognition based-CAPTCHA focusing on the advanced human-cognitive-processing ability of mental rotation. We implemented a prototype of our CAPTCHA, and the system was evaluated in a verification experiment. Twenty human subjects solved the challenges of our CAPTCHA in the experiment. The results show that the correct response rate is 77.3% and the average response time per one challenge is 5.4 seconds. Although the response time required per question is short, the correct response rates have to be improved. Our survey of the results of usability is satisfactory.

At present, there is still room for improvement in terms of both security and usability, so we plan to make improvements to the proposed method based on the knowledge obtained through the experimental results in this paper. For example, we are planning to consider 3D objects, which are more suitable for (sophisticated) mental rotation CAPTCHA, and improve selection of the viewpoint and threshold value of the correct answer range. In addition, we plan to use our results in the implementation and evaluation of type-β.

It is expected that sophisticated mental rotation is one of difficult tasks for automated programs (malwares). However, the attack techniques of malware vary, and the sophisticated mental rotation CAPTCHA's resistance to decipherment is not proven theoretically. We will conduct studies to determine whether our CAPTCHA is truly resistant to malware attacks.

References

1. The Official CAPTCHA Site, `http://www.captcha.net`
2. PWNtcha-Captcha Decoder, `http://caca.zoy.org/wiki/PWNtcha`
3. Yan, J., Ahmad, A.S.E.: Breaking Visual CAPTCHAs with Naïve Pattern Recognition Algorithms. In: 2007 Computer Security Applications Conference, pp. 279–291 (2007)
4. Elson, J., Douceur, J., Howela, J., Saul, J.: Asirra: a CAPTCHA that exploit interest-aligned manual image categorization. In: 2007 ACM CSS, pp. 366–374 (2007)

5. Chellapilla, K., Larson, K., Simard, P., Czerwinski, M.: Computers beat humans at single character recognition in reading-based Human Interaction Proofs (HIPs). In: 2nd Conference on Email and Anti-Spam (CEAS) (2005)
6. MSR Asirra Project, `http://research.microsoft.com/asirra/`
7. Golle, P.: Machine Learning Attacks Against the ASIRRA CAPTCHA. In: 2008 ACM CSS, pp. 535–542 (2008)
8. Vaughan-Nichols, S.J.: How CAPTCHA got trashed, Computerworld (July 15, 2008), `http://www.computerworld.com.au/article/253015/how_captcha_got_trashed/`
9. Shepard, R., Cooper, L.: Mental images and their transformations. MIT Press, Cambridge (1982)
10. Shepard, R., Metzler, J.: Mental rotation of three dimensional objects. Science, New Series 171(3972), 701–703 (1971)
11. YUNiTi.com, `http://www.yuniti.com/`
12. TechnoBabble Pro: How they'll break the 3D CAPTCHA, `http://technobabblepro.blogspot.jp/2009/04/how-theyll-break-3d-captcha.html`
13. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2000)
14. Stafford, T., Webb, M.: Mind Hacks. Oreilly & Associates Inc. (2004)