Directcha : A proposal of spatiometric mental rotation CAPTCHA

Directcha: A Proposal of Spatiometric Mental Rotation CAPTCHA

Ayane Sano[†], Masahiro Fujita[††], Masakatsu Nishigaki[††]
† Graduate School of Integrated Science and Technology, Shizuoka University, Hamamatsu, Japan
†† Graduate School of Science and Technology, Shizuoka University, Hamamatsu, Japan
nisigaki@inf.shizuoka.ac.jp

***Abstract*—A 3D CAPTCHA using mental rotation, called YUNiTi CAPTCHA, has been proposed as an advanced system to enable discrimination between computers and humans. YUNiTi CAPTCHA is performed in a "cognometric" mental rotation task. We point out that YUNiTi CAPTCHA has a vulnerability to pattern matching attacks. The pattern matching attack chooses the most similar images to a question image among a list of candidate images. We propose a new mental rotation CAPTCHA, called Directcha, to cope with the attacks. Directcha requires users to perform a "spatiometric" mental rotation task, in which users answer the direction of one 3D object in a question image. Directcha uses only one 3D object in the question image, so malware cannot break Directcha using pattern matching attacks. We implemented a prototype of Directcha and carried out basic experiments to test its usability. The results showed that, even though Directcha has higher attack tolerance than YUNiTi CAPTCHA, Directcha has nearly the same level of usability (correct response rate and response time) as that of YUNiTi CAPTCHA. We also describe threats to the security of Directcha.**

Keywords—CAPTCHA; Mental rotation; 3DCG; Space recognition; Spatiometric; Cognometric;

I. INTRODUCTION

Web services have been expanded greatly over the years. Unfortunately, this has given rise to malicious programs (malware) posting many spam comments on Web sites—submitting the same forms millions and millions of times. To cope with this issue, the Turing test plays an important role in discriminating humans from malware, 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.

Most Web sites utilize text-based CAPTCHA (Fig. 1) or image-based CAPTCHAs such as Asirra (Fig. 2) [2] to stop malware from interfering with their sites. However, an optical character reader (OCR) and machine learning could solve these CAPTCHAs [3, 4]. CAPTCHAs using higher human recognition abilities are needed to take measures against these malware [5]. One of these CAPTCHAs that has been proposed is an interesting three-dimensional (3D) CAPTCHA called YUNiTi CAPTCHA (Fig. 3) [6]. This CAPTCHA uses the ability of "mental rotation." Mental rotation involves the advanced cognitive-processing ability to rotate mental representations of two-dimensional (2D) and/or 3D objects.

YUNiTi CAPTCHA is performed in a "cognometric" mental rotation task. The cognometric task requires users to choose an appropriate image from a list of candidate images. In YUNiTi CAPTCHA, 18 candidate images (2D images of each 3D object) are displayed along with a question image (a 2D image of a 3D object). If Web page visitors can choose the same object corresponding to the question image, they are identified as humans. This approach is intuitive and easy-to-use, but YUNiTi CAPTCHA has a vulnerability to "pattern matching attacks." The pattern matching attack chooses the most similar image to its question image among its list of candidate images. Malware has no mental rotation ability, but it could break YUNiTi CAPTCHA by using the attacks. We will give a detailed description about this vulnerability in Section II.

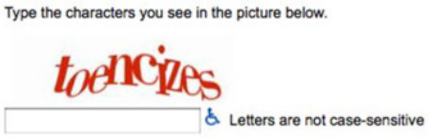
Type the characters you see in the picture below.

Fig. 1. CAPTCHA used by Google

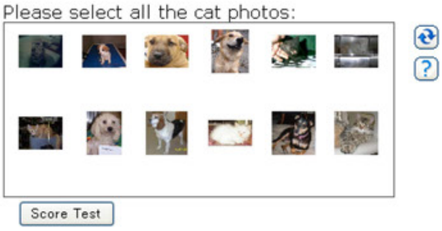
Please select all the cat photos:

Fig. 2. Asirra

This motivated us to design mental rotation CAPTCHAs that have tolerance against pattern matching attacks. In this paper, we propose one such mental rotation CAPTCHA, called Directcha. It requires users to perform a "spatiometric" mental rotation task. The spatiometric mental rotation task asks users to answer the direction of one 3D object in a question image. If users can correctly choose the direction of the 3D object, they are identified as humans. There are no candidate images and hence pattern matching attacks are useless. The spatiometric mental task uses an interesting feature of mental rotation: humans instantly distinguish the direction of an object between leftward and rightward. Therefore, even though Directcha has higher attack tolerance than YUNiTi CAPTCHA, it has nearly the same level of correct response rate and response time as that of YUNiTi CAPTCHA.

The organization of this paper is as follows. Section II describes YUNiTi CAPTCHA and a problem with it. Section III introduces Directcha, and Section IV shows basic experimental results of the CAPTCHA. Section V discusses automatic

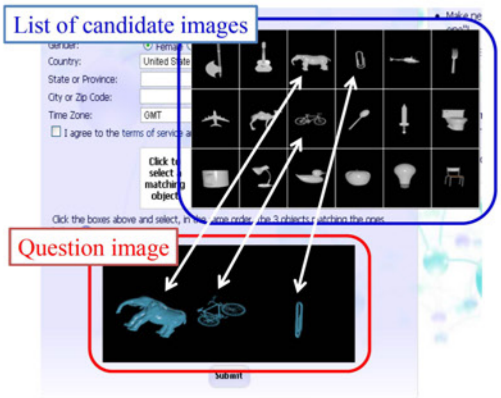
List of candidate images
Gender:
Female
Country:
United State
State or Province:
City or Zip Code:
Time Zone:
GMT
□ I agree to the terms of service a
Click to select matching objects
Click the boxes above and select, in the same order, the 3 objects matching the ones
Question image
Submit

Fig. 3. YUNiTi CAPTCHA

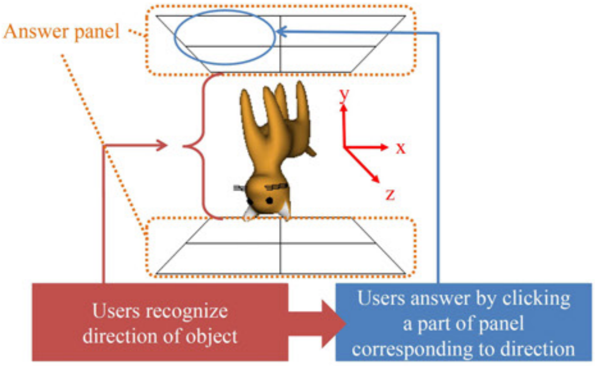
Answer panel
y
x
z
Users recognize direction of object
Users answer by clicking a part of panel corresponding to direction

Fig. 4. Directcha

generation of questions and Section VI discusses the effectiveness of the CAPTCHA. Finally, Section VII presents our conclusions and future work.

II. YUNiTi CAPTCHA

Humans have an ability to rotate 2D and/or 3D objects using their imagination and to recognize shape figures photographed from a different point of view. This human ability is called "mental rotation" [8, 9] and is one of the higher human recognition abilities.

YUNiTi CAPTCHA (Fig. 3) has been proposed as one of the 3D CAPTCHAs using mental rotation of 3D objects. YUNiTi CAPTCHA is performed in a "cognometric" mental rotation task. The cognometric task requires users to choose an appropriate image from a list of candidate images. YUNiTi CAPTCHA has a question image (a 2D image of a 3D object) and 18 candidate images (2D images of each 3D object). The question image is generated by randomly selecting a 3D object from the list and then photo shooting the object from different viewpoints. Web page visitors need to choose an appropriate image from the list, i.e., they need to choose the same 3D object as the object in the question image. Humans can choose the appropriate image by using mental rotation.

YUNiTi CAPTCHA is a very interesting approach, but it could be broken by using current image recognition techniques. Feature extraction techniques, such as SIFT [7] or SURF [10], have been developed rapidly. They enable malware to break YUNiTi CAPTCHA: malware can choose a candidate image that has the most similar features to the features extracted from the question image. We call the attacks "pattern matching attacks." Malware does not have mental rotation ability, but it could break YUNiTi CAPTCHA by using the attacks. Thus, we developed a new mental rotation CAPTCHA that has tolerance against the attacks.

III. Directcha

A. Concept

In this paper, we propose a new mental rotation CAPTCHA, called Directcha (Fig. 4). Directcha requires users to perform a "spatiometric" mental rotation task, in which users answer the direction of one 3D object in a question image. Directcha displays a 2D image of a 3D object and an answer panel. Users need to answer by clicking a part of the panel corresponding to the direction of the 3D object. Humans have the ability of mental rotation, so they can recognize how to rotate the 3D object from its front view into the question image [11]. This mean that we can answer the direction of a 3D object easily.

The spatiometric mental rotation task uses only one 3D object in a question image. Malware obviously cannot break Directcha using pattern matching attacks. In addition, using the task has an advantage. Mental rotation has an interesting feature. When an object angle increases, the time that people need to recognize the object increases; nevertheless, humans can instantly distinguish the direction of an object between leftward and rightward [9]. That is why we can easily distinguish the direction of the object between leftward and rightward (or between forward and backward). The spatiometric mental rotation task uses this feature, and Directcha uses this task. Thus, Directcha should respond relatively quickly even though it has higher attack tolerance than YUNiTi CAPTCHA.

B. Authentication Procedure

The authentication procedure of a Directcha system is as follows. The Directcha system is assumed to have many 3D models in a 3D model database. All the 3D models are "directed", i.e., they are clear in terms of top/bottom/left/right relationship.

Step 1. The system randomly picks up a 3D model.

Step 2. The system randomly rotates the 3D object picked up in Step 1 on the x-axis, y-axis, and z-axis.

Step 3. The system puts the 3D object rotated in Step 2 on the answer panel.

Step 4. The system projects the object and the answer panel onto a two-dimensional plane. This is used for the question image.

Step 5. The system shows a user (the Web page visitor) the question image.

Step 6. The user recognizes the direction of the 3D object in the question image and then clicks a part of the answer panel that corresponds to the direction of the 3D object.

Step 7. If the clicked position on the answer panel is correct, the user is identified as a human. If not, the user is identified as malware.

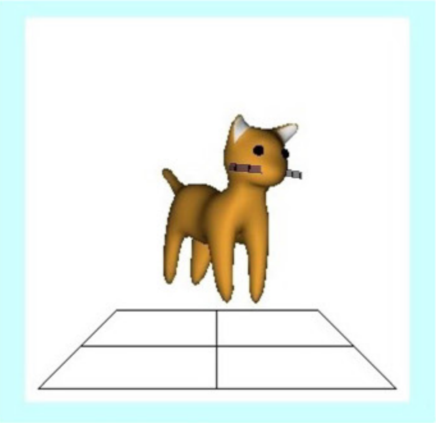

Fig. 5. Directcha-4

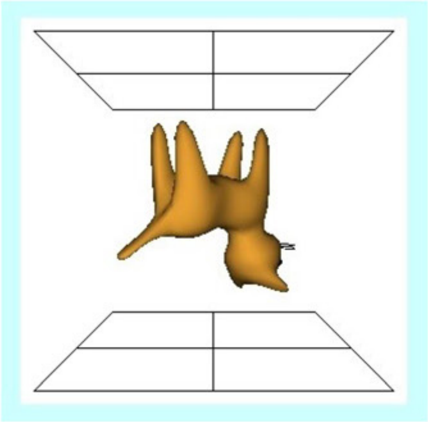

Fig. 6. Directcha-8

It is expected that malware cannot recognize the direction of the 3D object in a question image since it is not an easy task for computers to extract three-dimensional information from its 2D image. However, our system knows the rotation degree of the 3D object in Step 2. Because this knowledge forms a trapdoor, our CAPTCHA system (a computer) can automatically generate questions that malware (computers) cannot answer, and then the system can determine whether or not the positon clicked by the Web page visitor is correct.

C. Implementation

We conducted a basic experiment implementing two instances of a Directcha system: Directcha-4 and Directcha-8. Figure 5 and Figure 6 show authentication screen examples of Directcha-4 and Directcha-8, respectively. Directcha-4 requires users to click the direction in which an upright 3D object faces from the four directions (from the first quadrant to the fourth one on the y-axis). Thus, the number of possible answers is four. Directcha-8 requires users to click the direction that a 3D object faces from the eight directions (upright or upside down, in addition from the first quadrant to the fourth one on the y-axis). Specifically, if the 3D object in a question image is upright, users need to click a part of the lower panel. If the 3D object in a question image is upside down, users need to choose a part of the upper panel. Thus, the number of possible answers is eight. In Directcha-4 and Directcha-8, if the clicked panel corresponds to the direction of the 3D object, the users are authenticated. Figure 5 shows a cat that faces the right front in the upright state: if a user clicks the lower right front panel, the user is authenticated. Figure 6 shows a cat that faces right and backwards in the upside-down state: if a user clicks the upper right back panel, the user is authenticated.

When generating questions, the Directcha system needs to set some parameters such as the size of images, the point of view, and the rotation degree of 3D models. We set the parameters empirically through a preliminary experiment we conducted. We give a description about them in the following sections. Question image generation of our system is implemented by a C++ program including OpenGL [12]. The numerical values in the following descriptions are expressed in the unit used in OpenGL libraries.

1) Size of image

The size of a question image is 300 × 300 pixels. The coordinates [0, 0] are at the upper-left corner, the coordinates [299, 299] are at the lower-right corner.

2) Normalization of the object size and viewpoint

If a model is displayed too big in a question image, the model covers the answer panel. If a model is displayed too small in a question image, the user is unable to see it well. To prevent these, the system needs to normalize the size of a 3D object when picking up from the object data from the database. Although various normalization approaches could be used, our system scales models down to accommodate in a cube of 1.3 in size.

3) Point of view

The system created the smallest rectangular inscribed in the object and arranged the object so as to correspond to the center of the rectangular with the coordinates [0, 0, 0]. The point of view is [0, 0, 4.2], and the degrees are [0°, 0°, 0°].

4) Rotation degree

The system needs to choose degrees randomly on the x-axis, y-axis, and z-axis and then rotate a 3D object by the degrees. We applied some restrictions to this process.

(1) If the system chooses a degree near a border between two quadrants on the y-axis, users are not able to clearly determine the direction of the 3D objects in the image. To avoid this, the degree on the y-axis are chosen from 25° to 65°, 115° to 155°, 205° to 245°, and 295° to 335°.

(2) According to the preliminary experiment, the response time of users was significantly slow[1] when the rotation degrees on the x-axis or z-axis were large. To mitigate this, the degree on the x-axis was chosen from −10° to 10°. In Directcha-4, the degree on the z-axis was chosen from −5° to 5°. In Directcha-8, the degree on the z-axis was chosen from −5° to 5° and from 175° to 185°.

[1] The reason is unclear at this stage. Investigating the reason is future work, but we have considered that a possible reason is that we seldom see the objects rotated on the x-axis and/or the z-axis in the real world and therefore we are inexperienced at recognizing them.

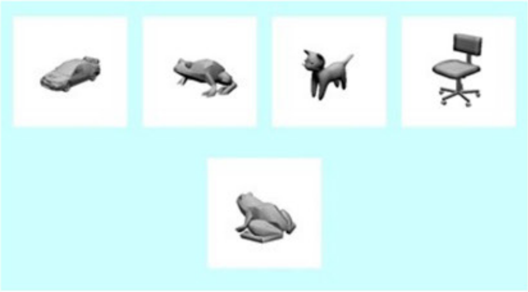

Fig. 7. YUNiTi-4

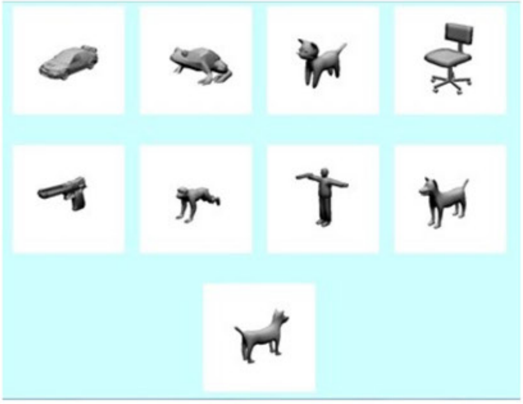

Fig. 8. YUNiTi-8

(3) The system rotates the 3D object in the order of the x-axis, z-axis, and y-axis.

IV. Basic Experiment

A. *Purpose*

We conducted basic experiments about the usability of the proposed Directcha CAPTCHA and YUNiTi CAPTCHA. We evaluated the usability of them in terms of the correct response rate and time.

B. *Experimental Method*

To compare the usability of Directcha, we developed two instances of a YUNiTi CAPTCHA system: YUNiTi-4 and YUNiTi-8. YUNiTi-4 has a question image and four candidate images. Thus, the number of possible answers is the same as that of Directcha-4. YUNiTi-8 has a question image and eight candidate images. Thus, the number of possible answers is the same as that of Directcha-8. The detailed description of the system is given in Section IV.B.2.

The subjects included twelve volunteers of college students in the faculty of computer science. We randomly divided the subjects into four groups: A, B, C, and D. Group A had three subjects who solved eight challenges of Directcha-4. Group B had three subjects who solved sixteen challenges of Directcha-8. Group C had three subjects who solved eight challenges of YUNiTi-4. Group D had three subjects who solved sixteen challenges of YUNiTi-8. Before the challenges, each subject could solve as many tutorial challenges as they wanted.

1) *Directcha System*

The system was the system implemented in Section III.C. The 3D objects used in this experiment were sixteen objects (Object A–P) provided free on Web sites. They were all "directed 3D objects" that were clear in terms of top/bottom/left/right relationship. Objects A–H were used in the tutorial challenges, and objects I–P were used in the challenges. As shown in Section IV.B.2, objects A–H did not have similar models of the same category. Objects I–P also did not have similar models of the same category.

In every tutorial challenge of group A, Directcha-4 picked up an object randomly from objects A–D and then generated a question image using it. In the challenges of group A, objects I–L were used twice each, for a total of eight question images that were generated by the system. The system showed the question images to the subject in random order. In every tutorial challenge of group B, Directcha-8 picked up an object randomly from objects A–H and then generated a question image by using it. In the challenges of group B, objects I–P were used twice each, for a total of sixteen question images that were generated by the system The system showed the images to the subject in random order.

2) *YUNiTi CAPTCHA System*

Figure 7 and Figure 8 show authentication screen examples of YUNiTi-4 and YUNiTi-8, respectively. This system was implemented under the same conditions in terms of brute-force attacks as Directcha-4's and Directcha-8's respective conditions. Group C needed to choose an image that had the same object as the object in the question image from the four candidate images (Fig. 7). Group D needed to choose an image having the same object as the object in the question image from the eight candidate images (Fig. 8).

The size of the question images and candidate images was 150 × 150 pixels. We used smaller images in size than the Directcha system's images (300 × 300 pixels) so that four candidate images are displayed in the transverse direction for easy observation. For specifications other than the size, YUNiTi-4 and YUNiTi-8 used the similar conditions corresponding to the specifications of the original YUNiTi CAPTCHA system[2]. The candidate images were always the same: the 3D objects in the candidate images were rotated by 315° on the y-axis, 20° on the x-axis, and 0° on the z-axis in this order. The question image was changed every time: the degree on the y-axis was randomly chosen from 0° to 359°, and then the 3D object in the question image was rotated by the chosen degrees on the y-axis, 20° on the x-axis, and 0° on the z-axis in this order. The 3D object in each image was made monochromatic ((R, G, B) = (204, 204, 204)) for it.

[2] The details of the original YUNiTi CAPTCHA system have not been published. So, we decided the conditions empirically.

TABLE I. Experiment results for each subject

Directcha-4

subject	Correct response rate	Average response time [s]
1	6/8	1.05
2	8/8	1.80
3	8/8	1.69
Average	91.7% (22/24)	1.52

YUNiTi-4

subject	Correct response rate	Average response time [s]
4	8/8	1.57
5	8/8	1.53
6	8/8	1.22
Average	100.0% (24/24)	1.44

Directcha-8

subject	Correct response rate	Average response time [s]
7	16/16	2.09
8	15/16	1.83
9	16/16	1.50
Average	97.9% (47/48)	1.80

YUNiTi-8

subject	Correct response rate	Average response time [s]
10	16/16	1.44
11	16/16	1.85
12	16/16	1.65
Average	100.0% (48/48)	1.64

In this experiment, the YUNiTi CAPTCHA system used the same sixteen objects (A–P) as the objects used in the Directcha system. YUNiTi CAPTCHA is a cognometric type of CAPTCHA. If two or more similar 3D objects were included in the candidate images, the correct response rate of users could decrease. To mitigate this issue, we used objects A–H that belonged to respective categories so that objects A–H were not similar to each other. Objects I–P also belonged to respective categories.

In every tutorial challenge of group C, YUNiTi-4 picked up an object randomly from objects A–D and then generated a question image by using it. The list of candidate images showed objects A–D. In the challenges of group C, objects I–L were used twice each, for a total of eight question images that were generated by the system. The system showed the images to the subject in random order. The list of candidate images showed objects I–L. In every tutorial challenge of group D, YUNiTi-8 picked up an object randomly from objects A–H and then generated a question image by using it. The list of candidate images showed objects A–H. In the challenges of group D, objects I–P were used twice each, for a total of sixteen question images that were generated by the system. The system showed the images to the subject in random order. The list of candidate images showed objects I–P.

C. *Experiment Results*

The experimental results are shown in Table I, which summarizes the correct response rate and the average response time for each subject.

1) *Correct Response Rate*

From Table I, the correct response rate of Directcha-4 was about 92% on average (a total of 24 times, 22 successes, 2 failures). The correct response rate of Directcha-8 was about 98% on average (a total of 48 times, 47 successes, 1 failure). Both the correct response rate of Directcha-4 and Directcha-8 was more than 90%. However, the correct response rate of both YUNiTi-4 and YUNiTi-8 was 100%. This result suggests that the correct response rate of Directcha was slightly lower than that of YUNiTi CAPTCHA. In the following, we analyzed why some subjects failed through subsequent interviews with the subjects and then explored their improvement in the correct response rate of Directcha.

The reason for the subject failing two challenges of Directcha-4 was due to misclicking in the answer panel. The subject recognized the correct answers, but he clicked out of the panel while moving the mouse. This mistake will be prevented by invalidating clicks outside of the panel. If the system was repaired to prevent users from misclicking out of the panel, it would increase the correct response rate of Directcha-4 to 100%.

The reason the subject failed a challenge of Directcha-8 was that he did not recognize what the object in the question image was. It suggests that, under specific conditions (e.g., when a specific object is rotated at a specific degree), correctly answering the directions of some objects is hard. We have two methods to mitigate this issue. First, through more comprehensive experiments, we can find the detailed conditions in which users tend to make a mistake. Then, we will add some restrictions, such as rotation degree or used objects, to Directcha. Second, we can use more detailed objects. The objects used in this experiment were free, so many of them were made simply. Using detailed objects will help users to recognize the direction of objects.

2) *Response Time*

From Table I, the average response time per challenge of Directcha-4 was 1.52 [s]; the shortest time was 1.05 [s], and the maximum time was 1.80 [s]. The average response time per challenge of YUNiTi-4 was 1.44 [s]; the shortest time was 1.22 [s], and the maximum time was 1.57 [s]. The average response time per challenge of Directcha-8 was 1.80 [s]; the shortest time was 1.50 [s], and the maximum time was 2.09 [s]. The average response time per challenge of YUNiTi-8 was 1.64 [s]; the shortest time was 1.44 [s], and the maximum time was 1.85 [s]. According to the results, Directcha had nearly the same level of response time as that of YUNiTi CAPTCHA. It shows the effectiveness of Directcha.

TABLE II. Expected value of correct response rate and response time

	Correct response rate	Average response time [s]
Directcha-4×6	100.0%	9.1
Directcha-8×4	91.9%	7.2

V. AUTOMATIC GENERATION

One of the requirements for CAPTCHAs is automatic generation of questions [17]. As shown in Section III.B, Directcha can generate questions automatically under the assumption that a database stores many “directed” 3D models. When constructing the database, we are able to use many 3D models on the Internet. However, they are not always directed (the top/bottom/left/right relationships are not clear), or they are directed but do not always face the front (the direction information are not available). This means that the Directcha system needs to extract directed 3D models from models on the Internet.

One good solution to this requirement is to use the idea of “covert filtering” [15]. Specifically, the Directcha system should show a question generated from a new object, picked up from the Internet, to people and test whether it is directed or not before adding it into the database. When solving a question, users see some images generated from these new objects to be evaluated (referred to here as “evaluation objects”) and some images generated from objects that are already known to be directed objects (referred to here as “vetted objects”). For example, a user may see three images, where two of the three are generated from vetted objects and where the other image is from an evaluation object. If a user correctly solves the captcha based on the two vetted images, then the system takes the given answer for the evaluation object as one user’s opinion. Once a certain number of user consistently correct a question generated from an evaluation object, the object is added to the database. However, if anyone provides a different answer, it is not added to the database.

VI. ATTACK TOLERANCE

A. Pattern Matching Attack

YUNiTi CAPTCHA is performed in a “cognometric” mental rotation task. We pointed out that YUNiTi CAPTCHA can be vulnerable to pattern matching attacks described in Section II. However, Directcha uses the “spatiometric” mental task, so it uses only one object in a question image. There are no candidate images to be matched and hence pattern matching attacks are useless. This is a major advantage; malware obviously does not break this CAPTCHA by using pattern matching attacks.

B. Brute-Force Attack

Elson et al. showed that by using token bucket scheme, CAPTCHAs with less than 1/4096 success probability of random guess can foil brute-force attacks [13]. We discuss our CAPTCHA's tolerance against brute-force attacks under the assumption.

To achieve a random guess probability of 1/4096, Directcha-4 needs to require users to solve six challenges (referred to here as “Directcha-4×6”), and Directcha-8 needs to require users to solve four challenges (referred to here as “Directcha-8×4”). On the basis of the results of the experiment shown in Section IV, we calculated the expected correct response rate and response time to Directcha-4×6 and Directcha-8×4 (Table II). The expected correct response rate of Directcha-4×6 and Directcha-8×4 was 100%[3] and about 92%, respectively. The expected response time of Directcha-4×6 and Directcha-8×4 was about 9.1 [s] and about 7.2 [s], respectively. The correct response rate of the text-based CAPTCHA was about 93%, and its response time was about 12.6 [s] [14]. These results show that Directcha with 1/4096 random guess probability has nearly the same level of correct response rate and response time as the text-based CAPTCHA.

C. Database Attack

Attackers can collect past questions and their answers for a CAPTCHA. If they collect large number of the pairs, they can solve the CAPTCHA by using the pairs. The attacks are called “database attacks” [15].

As shown in Section V, the Directcha system can automatically add directed 3D models to its database by using covert filtering. Ross et al. reported that the covert filtering technology contributes to not only “adding good models to the database” but also to “taking measures against database attacks” [15]. This means that, by using the covert filtering, our CAPTCHA will have the necessary tolerance against database attacks.

D. Machine Learning Attack

Directcha is a CAPTCHA requiring users to answer the direction of a 3D object. Malware may try to construct a classifier that recognizes the direction of an object in an image. The process is as follows.

Step 1. Attackers collect many question images having individual objects that are rotated by various degree. Attackers (humans) visually identify the rotation degrees of each image.

Step 2. Attackers extract features from each question image. Various features could show the directions of a 3D object such as the edge or intensity gradient. For example, attackers can use HOG features [16].

[3] The reason for the subjects failing some questions of Directcha-4 was due to misclicking out of the answer panel. We calculated this value under the assumption that the system will be repaired to prevent users from misclicking out of the panel. (Also, see Section IV.C.1)

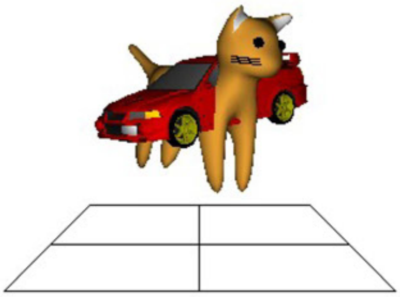

Fig. 9. Directcha using chimera object

Step 3. Attackers prepare training data, which are pairs consisting of features extracted in Step 2 and desired output values (rotation degrees obtained in Step 1).

Step 4. Using the training data, attackers construct a classifier that outputs the rotation degree by inputting a question image of a 3D object.

Step 5. Malware tries to break Directcha by using the classifier.

Many types of 3D models are developed daily. Intuitively, the more types of 3D objects that are developed, the more the features of 3D objects diversify. That is why we conclude that attackers will not be able to construct a classifier that calculates the rotation degree of 3D objects accurately.

In addition, the diversity of features of question images can be increased by processing the 3D objects. One of the processing methods is to use a "chimera object" [17] merged from two objects (object A and object B). It requires users to click both directions which two objects (object A and object B) face from four directions (Fig. 9). Figure 9 shows a cat that faces right front in the upright state and a car that faces left front in the upright state: if a user clicks the lower right front panel and the lower left front panel, the user is authenticated. In this case, a question image has the features of two 3D objects. This contributes to making it more difficult for attackers to construct the classifier.

The aforementioned discussion suggests that Directcha has the necessary tolerance against machine learning attacks. However, this has not been shown experimentally. We will conduct an experiment and show it in practice.

E. Remarks

In this section, we discussed the attack tolerance of Directcha. On the basis of the results, we stated that Directcha has higher attack tolerance against pattern matching attacks, which are the typical threats to YUNiTi CAPTCHA. Also, we showed that Directcha has higher attack tolerance against brute-force attacks, database attacks and machine learning attacks.

This paper is the first report of Directcha, so we focused on only these well-known threats against CAPTCHAs. We will investigate state-of-the-art work in the image recognition area, such as [18], and will re-think the tolerance of Directcha.

VII. Conclusion

In this paper, we proposed Directcha, which uses a spatiometric mental rotation task. The spatiometric mental task requires users to answer the direction of 3D objects. A major advantage of this CAPTCHA, compared to the conventional mental rotation CAPTCHA (YUNiTi CAPTCHA), is the higher attack tolerance against pattern matching attacks. We implemented a prototype of Directcha and carried out basic experiments to test its usability. The results showed that even though Directcha has higher attack tolerance than that of YUNiTi CAPTCHA, Directcha has nearly the same level of correct response rate and response time as that of YUNiTi CAPTCHA. We discussed the security of Directcha and the ability of automatic generation of questions.

We will conduct some experiments by changing various conditions: the type of 3D models, the rotation degree of 3D objects, the user interface, and so on. We will also conduct studies to determine whether or not our CAPTCHA is truly highly resistant to malware attacks, especially attacks using deep learning and neural network technology.

Acknowledgements

In this study, we used free 3D models published on Metaseko-Sozai (http://sakura.hippy.jp/meta/) and TurboSquid (http://www.turbosquid.com/). We would like to thank Prof. Isarida of Shizuoka Sangyo University for his valuable advice regarding "mental rotation." This work was supported by JSPS KAKENHI Grant Number JP25280046.

References

[1] *The Official CAPTCHA Site* [Online], Available: http://www.captcha.net/

[2] Microsoft Research. *ASIRRA* [Online], Available: http://research.microsoft.com/en-us/um/redmond/projects/asirra/

[3] J. Yan and A.S.E. Ahmad, "Breaking Visual CAPTCHAs with Naïve Pattern Recognition Algorithms," in *Proc. 2007 Comput. Security Applications Conf.*, Florida, 2007, pp. 279-291.

[4] P. Golle, "Machine Learning Attacks Against the ASIRRA CAPTCHA," in *Proc. 2008 ACM CSS*, Alexandria, 2008, pp. 535-542.

[5] T. Yamamoto et al., "CAPTCHA Using Strangeness in Machine Translation ," in *Proc. 24th IEEE Int. Conf. on Advanced Imform. Networking and Applications*, Takayama, 2010, pp. 430-437.

[6] D. Ngo. (2009,Mar. 25). *3D-based Captchas become reality - CNET* [Online], Available: http://www.cnet.com/news/3d-based-captchas-become-reality/

[7] D.G. Lowe, "Object Recognition from Local Scale-Invariant Features," in *Proc. Int. Conf. on Comput. Vision*, Toronto, 1999, pp. 1150-1157.

[8] R. Shepard and L. Cooper, "Mental Images and Their Transformations," Cambridge, MIT Press, 1986.

[9] R. Shepard and J. Metzler, "Mental Rotation of Three-Dimensional Objects," American Assoc. for the Advancement of Sci., New Series, vol. 171, no. 3972, 1971, pp. 701-703.

[10] H. Bay et al., "SURF: Speeded-Up Robust Features," in *Proc. European Conf. on Comput. Vision*, Graz, 2006, pp. 430-443.

[11] Y. Takano and M. Okubo, "Mental Rotation," in *Encyclopedia of Cognitive Science*, John Wiley & Sons, Tokyo, 2006.

[12] *Open GL* [Online], Availble: https://www.opengl.org/

[13] J.Elson et al., "Asirra: a CAPTCHA that exploit interest aligned manual image categorization," in *Proc. 2007 ACM CSS*, Alexandria, 2007, pp. 366–374.

[14] J. Kani et al., "Four-panel Cartoon CAPTCHA," in *IPSJ J.*, vol. 54, no. 9, Japan, pp. 2232-2243, 2013. (in Japanese).

[15] S. Ross et al. "Sketcha: A Captcha Based on Line Drawings of 3D Models," in *Proc. 19th Int. Conf. on World wide web*, New York, 2010, pp. 821-830.

[16] D. Navneet and T. Bill, "Histograms of Oriented Gradients for Human Detection," in *Proc. 2005 IEEE Computer Soc. Conf. on Comput. Vision and Pattern Recognition*, San Diego, 2005, pp. 886-893.

[17] M. Fujita et al., "Chimera CAPTCHA: A Proposal of CAPTCHA using Strangeness in Merged Objects," in *Proc. 3rd Int. Conf.* on *Human Aspects of Inform. Security, Privacy and Trust*, Los Angeles, 2015, pp. 48-58.

[18] S. Hao et al., "Estimating Image Depth Using Shape Collections," in *Proc. 41st Int. Conf. and Exhibition on Comput. Graph. and Interactive Techniques*, vol. 33, issue 4, no. 37, Vancouver, 2014.