International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.12 (2015), pp.369-380
http://dx.doi.org/10.14257/ijmue.2015.10.12.35

OTRCaptcha: A Novel Object and Text Recognition Based Image CAPTCHA

Zhen Ye, Yufeng Wu, Wenyao Zhu and Mingjun Wang

Lishui University
yezhen@lsu.edu.cn

Abstract

CAPTCHA is an important technology to prevent auto-script attack. Currently most of the CAPTCHA systems are text based, which firstly distort, rotate different characters and then use some obfuscation, aiming to make the text difficult to be recognized by auto-script while still can be learnt easily by real users. However, such kind of CAPTCHA schema either too simple, which can be attacked easily by using optimal character recognition (OCR) or machine learning based technology, or it is too complex that even real users cannot tell it. By observing such contradiction between security and usability, we propose a novel object and text recognition based image CAPTCHA system called OTRCaptcha. In OTRCaptcha, some object images (each has a label) and their names are attached into a background image respectively. In order to pass this CAPTCHA, users have to identify all the object images, labels within those objects and their names. Besides, users also need to identify the semantic relationship between object and its name. Both the theoretical analysis and experiment result show that OTRCapthcha can provide both high security and strong usability.

Keywords: *Image CAPTCHA; Object and Text Recognition; CAPTCHA Design*

1. Introduction

CAPTCHA technology is used by lots of websites to block malicious request from auto-script [13]. Currently most of websites use text based CPATCHA. Figure 1 is one of such text based CAPTCHA called reCaptcha, which is used by Google [1]. The idea of such text based CAPTCHA is similar, it generates some characters randomly or picks a word from dictionary, then it distort, rotate these characters and adding some noises into it. By asking user to recognize such characters, the CAPTCHA can distinguish real users and auto-script.

Figure 1. Google's ReCaptcha

However, text based CAPTCHA system has the contradiction in terms of security and usability. If the CAPTCHA is too simple, it can be easily attacked by using Optical character recognition (OCR) or machine learning technology, means the system has serious security issue. In contrast, if characters are highly distorted or with very complex background, even the real user cannot tell the correct answer, thus the system has poor usability.

Some recent studies also propose different image based CAPTCHA systems [6, 2, 3], which usually ask users to identify some special elements in a CAPTCHA image or some

ISSN: 1975-0080 IJMUE

special relationships between different CAPTCHA images. In such studies, the main issue is: to prevent replay attack, the system needs to maintain a very large image database.

By observing the disadvantages exist in text based CAPTCHA and traditional image based CAPTCHA, we propose a new image CAPTCHA system called OTRCaptcha, which based on both object and text recognition. In OTRCaptcha, some random selected object images, each attached with a label, are composited into a background image. Their names also appear in that background image in some random chosen order. To pass the CAPTCHA, users need to input those labels in the same order with their corresponding object names.

Our main contributions include: 1.propose a novel image CAPTHCA based on both object and text recognition; 2. do a detail theoretical analysis on OTRCaptha's security; 3.conduct a comprehensive experiment to show OTRCaptha has high usability.

The remaining of the paper is organized as follows. Chapter 2 surveys the related work. Chapter 3 introduces OTRCaptcha system and the algorithm. We analyze the OTRCaptche's security in chapter 4. In Chapter 5, we conduct comprehensive experiment to show OTRCaptche's usability. Chapter 6 summarizes the paper.

2. Related Work

2.1. Text Based CAPTCHA System

Most of the CAPTCHA systems are text based, the reason is recognizing such CAPTCHA is intuitive and no need to pre-training or learning, and also the CAPTCHA generation algorithm is easy. Many famous websites have deployed text based CAPTHCA, *e.g*. Google, Microsoft [4], Alibaba. However, in this kind of CAPTCHA system, there is a contradiction between security and usability. Once it pursuits high security, usually the usability will be affected. Figure 2 presents a CAPTHCA that is hardly recognized by real user. Although this CAPTHCA can prevent almost all kind of auto-script attacks, it is meaningless if it cannot be recognized by real user.

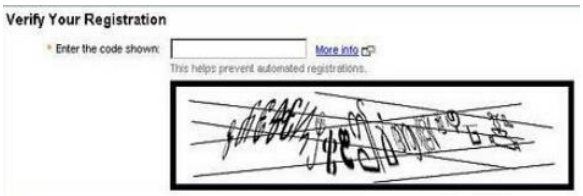


Figure 2. A Text CAPTCHA Which is Difficult to Recognize

For a system, the most basic requirement is that it can be used normally, so most of CAPTCHA systems choose to in the premise of keeping some level of usability, enhance the difficulty to auto-recognize the text CAPTCHA as much as possible.

However, as the development of machine learning and computer vision, OCR technology become stronger and stronger, thus such text CAPTCHA systems face more and more challenges [7].

[8] Presents an algorithm to attack the text based CAPTCHA system used in Microsoft MSN. Through a series of steps, the algorithm can locate and eliminate background noise around the CAPTCHA characters. After that, it locates each character and segments one by one. The experiment result shows that the success rate to segment text can be up to 90%, only cost tens of milliseconds. After segment characters, by applying state of the art machine learning algorithm to a single character, the recognition success rate is at least 95%. The average length of MSN CAPTCHA is 8, so the recognition success rate of the whole CAPTCHA can reach to 60% (90%*95%^8). In contrast, the design goal of this CAPTCHA is that the auto-recognition success rate should below 0.01%.

[15] Proposes a way to attack Google's text based CAPTCHA system. Although Google CAPTCHA uses different font size, style and even through strong distortion, the characters remain some invariant features. By detecting the unique feature model, the algorithm can identify the corresponding character and extract it from the whole CAPTCHA. The algorithm has three steps: preprocessing, model based character detection and character dividing. The experiment result shows that the segment success rate can be 62%. The average length of Google CAPTCHA is 5.5, so the whole recognition success rate is 46.75% (62%*95%^5.5).

[9] Introduces a semantic based text CAPTCHA system. In this system, each round of CAPTHCA is composed by three distorted words, two of which are synonyms. Users need to identify a word that has different meanings. Besides text recognition, this CAPTCHA adding a kind of semantic recognition thus can prevent auto attack. However, synonyms judgment has some requirements on user's knowledge background and is also easy to make mistake, thus the system has poor usability.

2.2. Image Based CAPTCHA System

Currently, some CAPTCHA researches are based on image, the idea is to ask user to recognize some specific objects within an image or some relationships between different images. [2] proposes a face recognition based image CAPTCHA system. In this system, each round of CAPTHCA includes several images, some of which contain real face while others not. Users need to identify all images that include real face. In HumanAuth [14] CAPTCHA, users need to recognize all images that describing the nature object, *e.g.* animals and plants. In [10], each round of CAPTCHA contains several images, each has an animal, and users need to pick up all the animals that meet some specific requirements. *e.g.* select all the cat images. The main issue of such image CAPTCHA systems is they need to build a very large image database to prevent replay attack.

[11] Introduces a image CAPTCHA that is generated by adding some object images into a background image. To use this CAPTCHA system, users need to answer questions about some specific relationships between different objects. In order to enhance the security, each object image can be distorted and conducts some color changes. At the same time, the background image can do further transformation and has some noises within it. The system has three kinds of relationship questions: 1.The relative position relationship; 2.Object number; 3.Correlation between different objects. However, the system has some disadvantages: 1.System's security and usability depend on the question being designed. If the questions are not well designed, even the real user cannot answer them easily. *e.g.* one of such bad questions is "select one object that is most unlike with other objects"; 2.if the number of object is small or the question is not well designed, it can be attacked by random guess. *e.g.* choose the object appear most in the images. If it only has 4 objects, the success rate of random guess can up to 25%. In contrast, if the object number is too large, it will affect CAPTCHA generation speed and user recognition success rate.

3. OTRCaptcha Image CAPTCHA System

In this paper, we propose a novel image CAPTCHA system, which is based on both object and text recognition, called OTRCaptcha. In OTRCapthca, an image CAPTCHA includes several objects, each has a label within it, and their names also appear in the image according to a random order. Figure 3 shows one such sample image CAPTCHA. When recognizing the CPATHCA, user should base on object names' order and input corresponding labels in the same order. In order to pass the CAPTCHA, user need to identify all objects, their names and the label within each object, thus has strong security. For real user, however, the corresponding relationship is obvious so that the recognition

process is easy and fast. There is no need to have strong background domain knowledge, so the system also has high usability.

Figure 3. Sample OTRCaptcha Image

The OTRCaptcha CAPTCHA system consists of two parts: 1. Image databases; 2. The algorithm to generate CAPTACHA image. The image databases include background image database, objects image database and the database contains the corresponding set of object names.

The role of background image database is to provide background image for the CAPTCHA system to make it more difficult to identify object and text by auto-script. To achieve this goal, the content and the color of background image should be rich. The number of background images can be relatively small.

The image in object image database should be identified easily by human user, thus should select those objects that are common in daily life and everyone knows as the candidates. Since the design of OTRCaptcha system can prevent replay attack effectively, there is no need to save very large number of object images. However, the number should larger than the number of background images. In order to further reduce the possibility of replay attack, in OTRCaptcha each object will have many names. *e.g.* in Figure 3, pigeon object can be named as: “pigeon”, “dove” and “bird”. Such object-names set are stored in object names database.

3.1. The Algorithm To Generate OTRCaptcha Image

The way to generate OTRCaptcha image is quite straightforward. In this chapter, we first define related concepts, then describe the process of generate OTRCaptcha image, at last we present how the users use OTRCaptcha.

We define the number of object in each OTRCaptcha image as n. The number of object name that appear in the OTRCaptcha image, which is also the number of labels user should input, is defined as m. O is the set of all images in object image database. B is the set of all images in background image database. D is the set of all object names contained in object name database. In each round of CAPTCHA image generation, S is defined as the set of selected object images. L is defined as the set of labels that will be added into each object image. According to their definition, we have $S \subset O$ & $|S| = n$ & $|L| = n$. N is

the set of object names that will appear in the background image. We have $|N| = m$. Figure 4 presents the details algorithm to generate OTRCaptcha image.

```
1.  S ← ϕ , N ← ϕ ,  L ← ϕ
2.  S ← randomSelectObject(O, n)
3.  m ← chooseNumberOfName();
4.  allIDs ← getIDs(S);
5.  ids ← randomSelectMIDs(allIDs, m)
6.  for each id ∈ ids
7.              name ← randomSelectName(id, D);
8.              N ← N ∪ name;
9.  end for
10. L ← selectLables(n);
11. b ← randomSelectBG(B);
12. captchaImage ← combine(S, L, N, b);
```

Figure 4. OTRCaptcha Generate Algorithm

Firstly, the algorithm randomly selects *n* images from object image database. Then it chooses *m* objects from those *n* images chosen before, and for each object, it selects one corresponding name from object name database. Then the algorithm generates *n* labels. The next step is randomly selects a background image from background image database. At last, it attaches labels into object images, one label each image, and then attaches object images and their names into background image to generate an OTRCaptcha image.

For users, it is quite easy to use OTRCaptcha. First, the user identifies the order of objects' names; then, finds all objects corresponding to those object names; after that, recognizes the labels whining those objects; and last, inputs those labels one by one in the same order with the object names. *e.g*. the correct input sequence of OTRCaptcha image shows in Figure 3 is: EDAC.

4. Security Analysis

For image based CPATCHA system, auto-script can use three ways to attack: 1.random guess; 2.replay attack; 3.use machine learning and computer vision based algorithm to do online recognition.

4.1. Random Guess Recognition

In OTRCaptcha system, user can pass the CAPTCHA when he input the labels in the same order with their corresponding object names, thus theoretically the auto-script can guess this order. The kind of labels within the objects is variable, for example in one round it uses English letters, and the next round it will use numbers. So the premise to do the random guess attack is the auto-script can recognize all labels within the objects. However, it is not an easy task with exist of complicated background image. Even it can recognize all labels, only some of the labels are required in the final input sequence, the number of labels, that is m, can vary from 1 to *n*. Besides, these m labels should be input following some specific order. Thus the success rate of random guess is $1/\sum_{m=1}^{n} P_n^m$. When the number of *n* is 4, the success rate is 1/40 (or 2.5%). While *n* is 5, the success rate is 1/185. If the auto-script can also recognize all object names, it can know the number of labels that should be input in the answer, so in this situation the success rate of random guess is $1/P_n^m$, when *n*=4, *m*=4, success rate is 1/24. When *n*=5, *m*=4, success rate is

1/120. From above analysis, we can see the success rate of random guess is very low, which means OTRCaptcha can prevent random guess attack very well.

4.2. Replay Attack Recognition

To do normal replay attack on image based CAPTCHA system, users need to download lots of CAPTCHA images by visiting the system continuously in advance, then manually tagging each image with a CAPTCHA answer. When encountering a new CAPTCHA image, it compares this image with saved images and thus retrieves the corresponding CAPTHCA answer this image represents. Since the CAPTCHA image in OTRCaptcha is generated by adding many objects images into background image, every time the CAPTCHA image is different, such replay attack is invalid in OTRCaptcha system.

A more advanced replay attack is on object level rather than the whole image level. The process is after downloaded lots of CAPTCHA images, using offline recognition algorithm to extract objects within the whole image and naming it. When meet a new CAPTCHA image, it locates and identifies object and then compares it with existed objects to get its name. In OTRCaptcha, firstly it's very difficult to extract and identify object image from background image. Secondly, even for the same object image, since every time it will be added into different backgrounds images and into different positions, do the similarity comparison between them is not so easy. Lastly, in OTRCaptcha each object has many names, so even auto-script can identify one object, knowing how to link it with their names is also a difficult task.

4.3. Machine Learning and Computer Vision Based Recognition

By using machine learning and computer vision technology, auto-script can do online recognition on image based CAPTCHA. In OTRCaptcha, the elements need to be recognized include all the objects, all the labels within the objects and all the objects names. Object and text recognition within complicated background is difficult and the success rate is relatively low [12, 5]. In order to pass OTRCaptcha captcha, it needs to recognize all above elements thus are very difficult.

More importantly, even the script can successfully recognize all the elements, how to relate the object name and the object image itself is almost impossible for the auto-script. However, this corresponding relationship is very nature and obvious in human eyes. *e.g.* even a child can link "snowman" with the object labeled with "*E*" in Figure 3.

5. Usability Experiment

We conduct a series of experiment to show OTRCaptcha's usability. The way we use is asking some users to recognize OTRCaptcha image and other CAPTCHA systems, and thus we can record their recognition speed and success rate respectively and do the comparison.

In this experiment, object image database includes 50 object images, which are downloaded from internet. Each image only contains a typical object and image's background is transparent. Background image database contains 20 background images, each of which has rich color and complicate content. Each object image is represented by a unique ID, which corresponding to 5 names stored in object name database.

The survey respondents we choose in the experiment are in different ages and with different educational level. The overall number of people is 50. Figure 5 shows the distribution of age. Figure 6 shows the educational level.

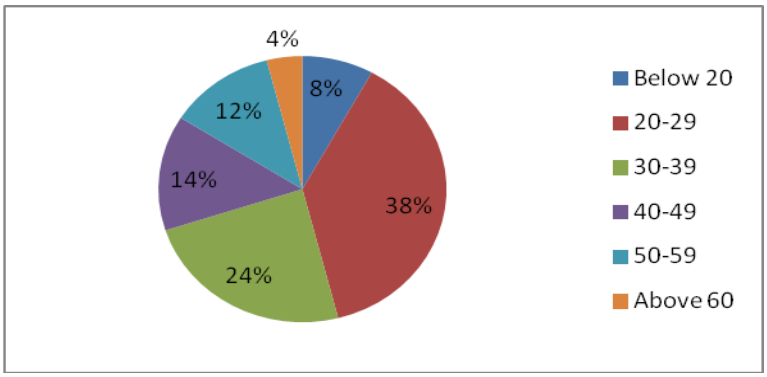


Figure 5. Distribution of Age

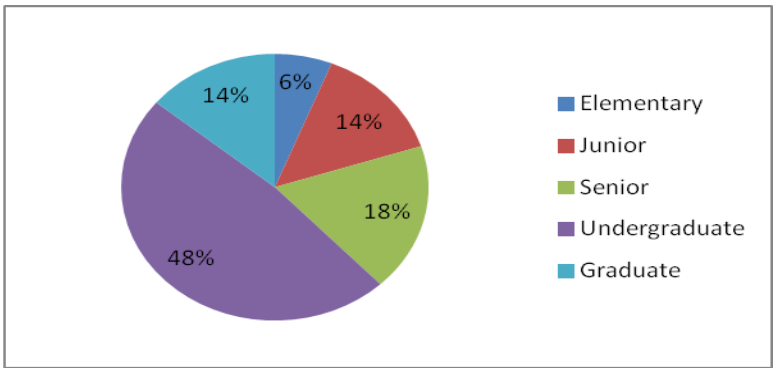


Figure 6. Distribution of Educational Level

5.1. Impact on Age and Educational Level

In this first set of experiment, the main purpose is to see the impact of age and educational level on CAPTCHA recognition speed and success rate. In this experiment, we choose Google reCaptcha [1] as our comparison algorithm. In this set of experiments, each user needs to identify 40 OTRCaptcha images and 40 reCaptcha images. Each OTRCaptcha image has 5 objects and 4 object names, which means users should input 4 labels as the answer.

Table 1. Impact of Age on Recognition Speed (Seconds)

	Below 20	20-29	30-39	40-49	50-59	Above 60
OTRCaptcha	8.1	8.8	9.1	14.5	21.8	35.8
reCaptcha	11.6	13.4	14.0	18.8	25.7	48.2

Table 1 presents the impact of age on recognition speed. From this Table we can see as user's age grows, recognition time also grows slowly. For users more than 60, as their slow reaction speed, little vison and unfamiliar with computer input mechanism, their recognition speed is very slow. Most of the users can recognize the OTRCapthc image within 10 seconds. In any age group, recognition speeds of OTRCaptcha are all faster than those in reCaptcha.

Table 2. Impact of Age on Recognition Success Rate

	Below 20	20-29	30-39	40-49	50-59	Above 60
OTRCaptcha	96.25%	98%	97.44%	91.93%	86.83%	79.12%
reCaptcha	88.22%	91.83%	92.06%	83.36%	79.42%	53.75%

Table 2 shows the impact of age on recognition success rate. In Table 2, the success rate of users below 40 is almost above 95%. For users older than 40, as the age continue grow, the recognition success rate decreases slowly. For the users above 60, their recognition success rate is very low. In each age group, the recognition success rate is much higher than those of reCaptcha. In our analysis, there are two reasons to cause the incorrect recognition in OTRCaptcha: 1.object names highly overlapped, causing them unrecognizable; 2. the colors of object image and their label are similar, which confuse the users.

The result in Table 1 and Table 2 shows when using OTRCaptcha, users below 50 can recognize the CAPTCHA image quickly and with very high success rate. Such users accounted for a great proportion of all internet users. Even for users above 50, they also can recognize OTRCaptcha image well.

Table 3. Impact of Educational Level on Recognition Speed

	Elementary	Junior	Senior	Undergraduate	Graduate
OTRCaptcha	27.5	15.2	11.1	9.2	9.3
reCaptcha	38.2	23.8	16.4	14.1	12.9

Table 3 presents the impact of educational level on CAPTCHA recognition speed. The result shows for users whose educational level above junior middle school, most of the recognition speeds are between 10 seconds and 15 seconds. The recognition speeds of other users are relatively slow. For the same group of users, their recognition speeds in OTRCaptcha are all much faster than reCaptcha.

Table 4. Impact of Educational Level on Recognition Success Rate

	Elementary	Junior	Senior	Undergraduate	Graduate
OTRCaptcha	83.33%	90.14%	96.11%	97.4%	97.86%
reCaptcha	59.17%	82.86%	88.45%	91.77%	92.86%

Table 4 shows the impact of educational level on recognition success rate. The Table tells us for users whose educational level above junior middle school, their recognition success rate almost above 90%. For users have just a primary school education, their recognition success rate is much lower. Again, for each group, the recognition success rate in OTRCaptcha is much higher than reCaptcha.

From Table 3 and Table 4 we can see, for users who use OTRCaptcha, when their educational level is junior middle school or above, they can recognition the CAPTCHA image quickly and correctly. For users whose educational level is elementary school, the recognition speed is relatively slow and also the recognition success rate is not very high.

5.2. Impact on Number of Object Images

This set of experiment shows the impact on number of object images. In this set and the next set of experiments, the respondents are aged between 25 and 30, and total number of person is 10.

In this set of experiment, the number of objects within one CAPTCHA image is increased one by one from 2 to 7. The number of object name appear in the background image is the same with the number of object image. For each number of objects, each user needs to recognize 40 OTRCaptcha images.

Figure 7 presents the result. From this Figure we can see, as the number of object increase, users' ability to recognize the CAPTCHA image become progressively difficult. After the number is larger than 5, recognition time increases significantly while the success rates decrease a lot. One of the reasons is as the number increase, user need to identify more objects. Another reason is as the number increase, the CAPTCHA image

becomes more crowded and thus makes it more difficult to be recognized. On the other side, as the number of object increase, OTRCaptcha's security also improved.

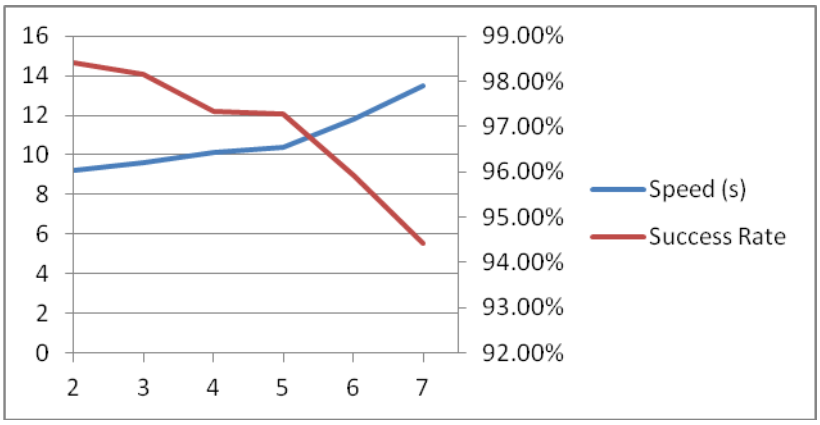


Figure 7. Impact on Object Number

5.3. Impact on Number of Object Names

In this set of experiments, we fix the number of object image to be 5 and increase object names, which appear in the background image, from 1 to 5 one by one, which means the number of labels user needs to input are increased from 1 to 5. For each number of object name, users need to identify 40 OTRCaptcha images.

Figure 8 presents the result. From the Figure we can see, as the number of objects names increase, the time to recognize is also increase. At the same time, the accuracy decreases slowly. However, from Chapter 4's analysis, we know as this number increases, OTRCaptcha system can prevent more attacks.

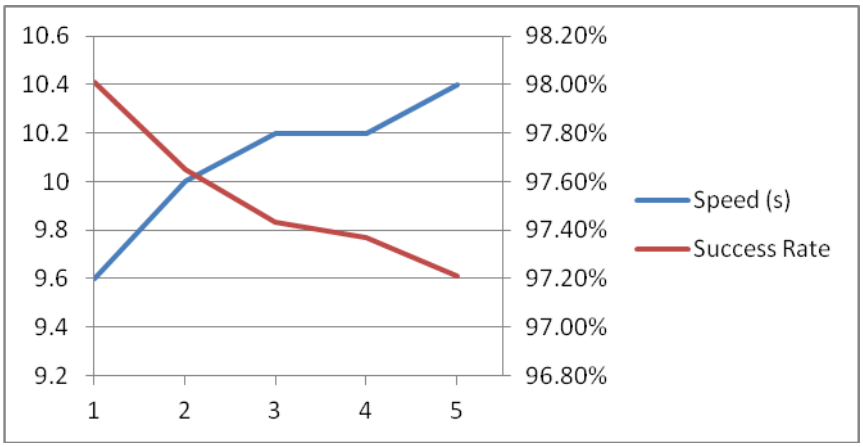


Figure 8. Impact on Number of Object Names Appear on the Background Image

6. Summary

In this paper, we propose a novel object and text recognition based image CAPTCHA system called OTRCaptcha. CAPTCHA images in OTRCaptcha are composed of object images, object names, labels within object images and the background image. In order to pass OTRCaptcha, user needs to recognize all above elements and also understand sematic relationship between object images and object names, thus the system has strong security. At the same time, since this recognition process and the sematic corresponding relationship is intuitive for real user, so it also has high usability.

Currently we only analysis OTRCaptcha's security in theory, in the future we will choose some state of the art machine learning or computing vision based CAPTCHA-attacking algorithms to attack our OTRCaptcha and other CAPTCHA systems, verifying if OTRCaptcha really has high security as we state in the paper.

We also plan to deploy OTRCaptcha in some website within our university's intranet, trying to see if it is usable in the real environment. In OTRCaptcha, it consumes lot of resource to generate one CPATCHA picture. We will observe such CAPTCHA generation process is acceptable from performance perspective in such large scale real environment and optimize it if necessary.

Acknowledgment

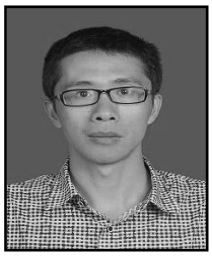
This work was partially supported by Scientific Research Fund of Zhejiang Provincial Education Department (No. FX2014120 and No. Y201432599)

References

[1] L. V. Ahn, B. Maurer, C. McMillen, D. Abraham and M. Blum, "reCAPTCHA: Human-Based Character Recognition via Web Security Measures", Science Magazine, vol. 321, **(2008)**, pp. 5895.
[2] G. Goswamia, B. M. Powellb, M. Vatsaa, R. Singha and A. Nooreb, "Face detection based color image CAPTCHA", Future Generation Computer Systems, vol. 31, **(2014)**.
[3] C. J. Hernandez-Castro and A. Ribagorda, "Pitfalls in CAPTCHA design and implementation: The Math CAPTCHA, a case study", Computers & Security, vol. 1, no. 29, **(2010)**.
[4] A. James, G. George and A. Yeldose, "A Survey on Spelling Based CAPTCHA", International Journal of Research in Computer and Communication Technology, vol. 3, no. 3, **(2014)**.
[5] V. N. Pawar and S. N. Talbar, "An Investigation of Significant Object Recognition Techniques", International Journal of Computer Science and Network Security, vol. 9, no. 5 **(2009)**.
[6] S. Vikram, Y. Fan and G. Gu, "SEMAGE: A New Image-based Two-Factor CAPTCHA", Proceedings of the 27th Annual Computer Security Applications Conference, Texas, USA, December 6-10, **(2011)**.
[7] B. Zhu, J. Yan, Q. Li, C. Yang, J. Liu, N. Xu, M. Yi and K. Cai, "Attacks and Design of Image Recognition CAPTCHAs", Proceedings of the 17th ACM conference on Computer and communications security, Chicago, USA, October 4-8, **(2010)**.
[8] J. Yan, A. Salah and E. Ahmad, "Low-cost Attack on a Microsoft CAPTCHA", Proceedings of the 15th ACM conference on Computer and communications security, Alexandria, USA, October 27-31, **(2008)**.
[9] P. Lupkowski and M. Urbanski, "SemCAPTCHA–the user-friendly alternative for OCR-based CAPTCHA systems", International Multi-conference on Computer Science and Information Technology, Wisla, Poland, October 20-22, **(2008)**.
[10] D. D'Souza, J. Matchuny and R. Yampolskiy, "Zoo CAPTCHA: Telling Computers and Humans Apart via Animal Image Classification", The Third ASE International Conference on Cyber Security, Standford, USA, May 27-31, **(2014)**.
[11] P. Matthews, A. Mantel and C. Zou, "Scene tagging: image-based CAPTCHA using image composition and object relationships", ACM Conference on Computer and Communications Security, Beijing, China, April 13-16, **(2010)**.
[12] C. Yi, X. Yang and Y. Tian, "Feature Representations for Scene Text Character Recognition: A Comparative Study", In International Conference on Document Analysis and Recognition, Washington DC, USA, August 25-28, **(2013)**.
[13] "CAPTCHA", http://en.wikipedia.org/wiki/Captcha.
[14] "HumanAuth", http://sourceforge.net/projects/humanauth/.
[15] A. S. Ahmad, J. Yan and M. Tayara, "The Robustness of Google CAPTCHAs", Computing Science Technical Report CS-TR-1278, **(2011)**.

Authors

Zhen Ye, receives his PhD degree in Computer Science and Technology in 2013 from Zhejiang University, China. Currently he is a lecturer in Lishui University, China. His areas of interest are Distributed System, Computer Vision and Machine Learning.



Yufeng Wu, receives his Master degree in Computer Science and Technology in 2005 from Zhejiang University, China. Currently he is a lecturer in Lishui University, China. His research area includes Software Engineering, Pattern Recognition and Data Mining.

Wenyao Zhu, was born in April, 1976. He receives his Master degree in Software Engineering from East China Normal University. Currently he is a lecturer in Lishui University, China. His research area includes Software Engineering and Pattern Recognition.

Mingjun Wang, received his Bachelor in Computer Science and Technology from Beijing Jiaotong University (2004), M.Sc. in Computer Application (2006) from China University of Geosciences (Beijing) and PhD in Pattern Recognition and Intelligent System (2010) from Donghua University. Now he is a full-time lecturer of informatics at Engineering College, Lishui University. His current research interests include different aspects Data Mining.