Scene Tagging: Image-Based CAPTCHA Using Image Composition and Object Relationships

Peter Matthews
School of Electrical Engineering & Computer Science
University of Central Florida
Orlando, FL 32816
pmatt@cs.ucf.edu

Cliff C. Zou
School of Electrical Engineering & Computer Science
University of Central Florida
Orlando, FL 32816
czou@cs.ucf.edu

ABSTRACT
In this paper, we propose a new form of image-based CAPTCHA we term "scene tagging". It tests the ability to recognize a relationship between multiple objects in an image that is automatically generated via composition of a background image with multiple irregularly shaped object images, resulting in a large space of possible images and questions without requiring a large object database. This composition process is accompanied by a carefully designed sequence of systematic image distortions that makes it difficult for automated attacks to locate/identify objects present. Automated attacks must recognize all or most objects contained in the image in order to answer a question correctly, thus the proposed approach reduces attack success rates. An experimental study using several widely-used object recognition algorithms (PWD-based template matching, SIFT, SURF) shows that the system is resistant to these attacks with a 2% attack success rate, while a user study shows that the task required can be performed by average users with a 97% success rate.

Categories and Subject Descriptors
K.6.5 [**Management of Computing and Information Systems**]: Security and Protection

General Terms
Security

Keywords
CAPTCHA, HIP, access control; image/video recognition; multi-object composition; security

1. INTRODUCTION
A number of abuses of Internet services are only made possible by the use of automated programs, such as the mass posting of spam to comment sections and user forums, mass user account registration, brute force password attacks, and abuse of online polls. To prevent such abuses, services may require a user to pass a CAPTCHA [1] (Computer Automated Turing Test for telling Computers and Humans Apart), a challenge-response test that can be easily solved by human users but is extremely difficult for computer programs and, hence, determines whether a service request originated from a human or an automated program. Such test systems are now in wide use on the Internet, and play a critical role in ensuring the integrity of many of the most popular websites. In order to be effective in this role, it is important that CAPTCHAs are strongly resistant to automated attacks and at the same time do not cause problems for human users.

Text-based CAPTCHAs, in which users are required to transcribe text presented in a distorted image, make up the majority of CAPTCHA systems in real-world use. However, their vulnerability to attack has been repeatedly demonstrated by computer vision researchers [5, 15, 17]. For example, several commercial CAPTCHA implementations were attacked in 2004 by Microsoft researchers with 80%-95% success rates achieved [5]. To guard against these automated attacks, stronger and more elaborate distortions have been utilized in creating images used by text-based CAPTCHAs; this, however, greatly impairs human legibility, resulting in higher user error rates [18] and associated levels of user frustration. As the advance of character recognition algorithms continue to improve the attack capabilities of automated programs, human legibility problems will likely make text-based CAPTCHAs less effective in the future.

Because CAPTCHA systems are so important to the modern Internet, the need for investigation of alternative CAPTCHA formats is clear. Facing this challenge, in recent years researchers have begun to develop image-based CAPTCHA systems [2, 6, 7, 9, 11]. These typically require the user to identify the subject of an image, which is generally considered to be a significantly more difficult task to automate than that of distorted character recognition. In addition, user studies we have conducted have suggested that image-based CAPTCHAs are largely preferred by users over text-based CAPTCHAs for reasons such as ease of use and a preference for a mouse-based response format. However, it may be argued that the proposed systems have not been able to satisfactorily address the difficulties associated with creating a large, correctly tagged image database and that their security has not been demonstrated against sufficiently advanced automated attack methods.

We propose a novel form of image-based CAPTCHA that we term "scene tagging". Rather than asking a user to identify the subject of an image, a scene tagging problem consists of multiple object images composed with a background image into a single scene and tasks the user with understanding a relationship between several of the objects present. The advantages of this approach are:

- Answering a question correctly requires successful recognition of all or most of the present image objects in the relationship to which this question refers. This poses a more difficult task to attackers than that of the image-matching required in a number of proposed image-based CAPTCHAs.

ASIACCS'10, April 13–16, 2010, Beijing, China.

• A number of other systems are vulnerable to attacks based on image similarity metrics if the source image database is known and not sufficiently large. The combination and composition-based nature of the images generated by our system results in an extremely large space of possible images without requiring a large image database, even before considering image distortion. This minimizes the difficulty associated with building a large, correctly tagged image database by making more effective use of an image database of a relatively small size.
• The relationship-based format of the questions results in many possible questions for a given image. This allows multiple questions to be asked of a user without incurring an additional bandwidth cost.
• A carefully designed sequence of systematic image distortion and clutter is performed during the image composition process in order to make it extremely difficult for automated attacks to determine the quantity, identity, and location of the objects present amongst the large amounts of extraneous image information. Extensive user testing has been used in the design of this sequence in order to ensure that the impact upon human performance is relatively minor.
• Experimental and user study results show that the system presents a task that is strongly resistant to a number of state-of-the-art automated machine vision-based attacks, yet can be easily solved by the average user.

The remainder of the paper is organized as follows. Section 2 explores related work in developing alternative CAPTCHA techniques. Section 3 discusses scene tagging in greater detail. Section 4 discusses the forms that likely automated attacks would take, while Section 5 discusses the system's countermeasures in the form of the systematic distortion engine. Section 6 describes the experimentation and user study performed to determine the viability of scene tagging and analyzes the results. In Section 7 we conclude the paper and discuss further work.

2. RELATED WORK

A number of CAPTCHA systems based upon the understanding of semantic image content have been proposed. Many early image-based CAPTCHA systems simply ask the user to identify the subject of an image or a subject associated with a set of images. However, resistance to image-similarity based attacks requires the creation and maintenance of a large, correctly labeled image database – a task that is labor intensive and fraught with legal issues regarding usage. Approaches to dealing with this shortcoming include the sourcing of images from image search engines [6]; however, the manner in which image search engines tag images may result in mislabeled or potentially offensive results. Microsoft's Asirra [9] asks users to identify the images of cats out of a set of photographs of cats and dogs, sourced dynamically from a frequently updated pet adoption website. However, a successful machine learning attack [10] has been demonstrated against the system, a success that largely appears to be a result of the binary nature of the classification problem posed. The composition-based nature of our system largely avoids image database creation problems by generating an extremely large space of possible test images and associated questions from a relatively small database size.

Implicit CAPTCHAs [2] tasks the user with semantically understanding the objects present in an image. A problem with the system is that it requires significant manual labor in the annotation of images and creation of challenges, while our approach does not require manual annotation of images or creation of questions. A system named IMAGINATION [7] also utilizes image combination and distortion in generating the images presented to the user. Unlike this system, our approach utilizes overlay-based composition of irregularly shaped objects with a background image, utilizes a more complex sequence of image distortions such as non-linear image transformation, and tasks the user with understanding the relationships between multiple objects in an image rather than image center or image subject identification. Researchers in the same group [8] attempt to quantify the effect of distortions on human and machine recognition of image subjects, but have only tested with non-composed, single subject images. This is a different problem than that utilized by our proposed system, and thus testing a different set of automated attacks is appropriate.

Google Research's What's Up CAPTCHA [11] requires users to identify an image's proper upright orientation. One disadvantage of the system is that attackers may be able to build a pre-rotated image database given a known image database and then utilize image similarity metrics in order to determine correct image rotations. The task's user success rate when the images have undergone distortion, the proposed defense, has not been tested. It is possible that usability would become a problem in this case.

3. SYSTEM OPERATION

Scene tagging is a novel form of attack-resistant image-based CAPTCHA. It creates an image via composition of a background image and several irregularly shaped object images, applying a sequence of image distortions during this process, and asks a user to answer a series of questions based on the relationships between a number of objects in the image. Figure 1 displays a sample image used in our proposed system to which no distortion has been applied. A question for such an image might be "Name the object that is least closely related to the other objects in the image", and the correct answer would be "ball". Such a question requires identification of all or most objects in the image in order to answer correctly, posing a difficult task for automated attacks.

3.1 Composite Image Creation

In order to create a test CAPTCHA image, the system first randomly chooses a background image from the set of available backgrounds. It then performs the first round of image distortion, as described in a later section. Subsequently, a number of objects are selected from the object database. The associated object images are randomly placed over the distorted background image, with care taken to avoid overlapping of objects. Then, a second round of image distortion is applied to this composite image to generate the final image used for the test.

The object images are depictions of the associated objects that utilize transparent backgrounds to ensure that only the object is present, and not any extraneous object background information. The use of these irregularly shaped objects also makes it more difficult for automated attacks to determine object boundaries in the presence of image distortion and clutter.

3.2 Question Types

Question and answer pairs are generated from the relationships between objects placed in a scene tagging image. Our prototype generates the following three types of questions:

Figure 1. Scene Tagging CAPTCHA image with a frog, a second frog, a butterfly, and a soccer ball present.

***Relative Spatial Location*:** In this type of question, a user must identify and determine the relative spatial location of the objects present in the test image. An example that might accompany figure 1 is "Name the object that is directly to the upper-left of the butterfly", to which the answer would be "soccer ball". Care is taken by the system to only generate questions to which there exists a clearly correct answer, avoiding questions which may cause confusion due to the existence of multiple reasonable answers.

***Object Quantity*:** This type of question requires a user to identify an object based on the number of that particular object present in the test image. For example, the user might be asked "Name the object of which there are two present in the image" when presented with figure 1, to which the correct answer is "frog". The database may contain multiple image depictions of each object, such that the multiple instances of the object present may take different image forms.

***Logical Association*:** This type of question asks a user to determine the objects present in the scene, and also to determine the logical associations between these objects. A question of this type that might accompany figure 1 is "Name the object that is least like the others present" and the answer would be "soccer ball" given that the other three objects are all animals. These associations are based on being members of a shared logical object group and are set during object database creation via appropriate object-group tagging. Answering the question correctly typically requires recognition of all or most of the objects in the image, making it more difficult for automated attacks. It is possible that in the future a lexical database such as WordNet [14] could be used to automatically set the strength of concept associations [4].

3.3 Response Formats

For a given scene tagging problem, the question instructs the user to provide an answer via one of the following response formats.

Multiple Choice: The first response format is that of selecting the correct answer from a list of choices presented to the user. In our prototype system, the user is presented with 16 choices of which one is the correct answer to the associated question and the rest are objects selected randomly from the object database. The user selects their answer choice via mouse.

Image Point Selection: The second response format asks the user to click the center of the object in the image that corresponds to the correct answer. A circular tolerance region, the radius of which has been set in our prototype system based on a user study, is used to compensate for inaccuracies in user point selection.

It should be noted that we conducted portions of a preliminary user study with a text-entry response format, but later abandoned it based on user feedback and poorer task success rates. Many users prefer questions that may be answered quickly via use of a mouse over keyboard-based answer entry. Further benefits of avoiding text-entry are that it helps minimize the cultural issues of dealing with multiple names for an object and avoids problems involving user misspellings or typographical errors.

4. AUTOMATED ATTACK METHODS

4.1 Attack Model

The attack model we utilize first assumes that an attacker may know the algorithm utilized to create scene tagging instances. It is also assumed that the attacker has access to the database of tagged image objects that the system uses in composite image generation. This model satisfies the definition of a CAPTCHA [1] in that the system's security is based on the assumed difficulty of an AI problem, not on "the secrecy of a database or a piece of code".

4.2 Randomized Guess Attacks

The first method of attack that must be considered is that of a randomized guess attack. The user may be asked to answer more than one question to strengthen resistance to such attacks. A combination of two 16-answer multiple choice questions and one 50-pixel tolerance radius image point selection question would mean that 1 in every 10,000 guess-based attacks would succeed. The detailed analysis is omitted due to space limitations.

4.3 Composite Image Similarity Attacks

In a composite image similarity attack, an attacker collects a large set of scene tagging instance images and manually tags the objects involved. The attacker then measures the similarity between the presented image and those in the stored set, returning an answer based on the closest stored scene tagging instance. However, this type of attack is not practical against our system, as the object combination-based manner of image generation our system utilizes results in an extremely large space of possible images, even when utilizing a relatively small object database.

4.4 Object Recognition Attacks

An object recognition attack utilizes machine vision techniques in order to determine a) the number of objects that are present in the image, b) the identity of these objects, and c) the image location of each of these objects.

The first form of object recognition algorithms we consider are the simplest – those that measure the difference in pixel color values between object database images and possible object locations in the problem image. This measurement may be coarse-grain, as in comparing regional color histograms, or fine-grain, as in measuring corresponding pixel value differences between objects and potential image locations. A representative candidate from the group used in our testing is template matching via measuring normalized pixel-wise difference (PWD).

The second, higher-level form of object recognition algorithms we consider searches for correlation between object database images and the scene image with regards to various distinctive image locations, such as the intersections of detected edges or points of great discontinuity in scale-space representations. Strong feature matches may then be used in a voting process to determine object presence and location. One representative algorithm that has been shown to outperform many others [13] is Scale Invariant Feature Transform (SIFT) [12]. Another method for image feature generation is Speeded Up Robust Features (SURF) [3]. SURF is faster than SIFT and is claimed to be more robust than SIFT to various image transformations. We test our system using both of these object recognition algorithms.

5. SEQUENCE OF SYSTEMATIC IMAGE DISTORTIONS

5.1 Distortion Criteria

Our system of distortions consists of two sets, the first being applied only to the background image and the second being applied to the composite of the background and object images. The need for two different sets of distortions is due to the different purposes served by these sets. Distortion of the background image is performed to impede attacks that utilize knowledge of the database of background images. There is no necessity to preserve human recognition of background images, thus drastic distortions may be considered. Distortion of the composite image, conversely, primarily attempts to make machine recognition of objects more difficult. At the same time, human recognition of these objects is crucial. Thus, the sequence of distortions must be chosen as to make things difficult for automated attacks without distorting the objects in the image beyond the point of human recognition.

5.2 Distortions applied to background image

Randomized Clutter: Primary distortion of the background image is done via the use of randomly placed image clutter of randomly determined shape and color. While the primary benefit of this distortion is that it impairs attempts to identify the background image utilized, an additional benefit is that it may add elements of interest to areas of the background image that may otherwise be uninteresting. This not only makes object isolation more difficult, it also results in a larger number of extraneous image interest points.

Global Color Shifting: The color channel values of every image pixel are shifted by a randomly determined amount. This color shifting is performed to hamper attempts to utilize information about the background in reversing the more involved and extensive color shifting that follows during image composition.

5.3 Distortions applied to composite image

To perform distortions that are localized in a randomly determined fashion, a number of the system's image distortions utilize randomly generated orthogonal image partitions. These divide the image such that every image location is contained in exactly one of these contiguous, rectangular regions.

Object Scaling: Before their composition with the background, object images are scaled horizontally and vertically in a random fashion. The dimensions of scaling are independent, and thus the aspect ratio of the object image may change. This impairs object recognition techniques that rely heavily on the relative position of object image features in order to determine object presence and location.

Mesh Warping: Mesh warping [16] warps the images by means of two 2-d arrays of image coordinates S and D, such that the values of the control pixels specified in S are moved to the destination coordinates in D and all other pixel values are calculated via means of interpolation between the grids. The resulting nonlinear transformation between source and destination coordinates is valuable as a large number of object recognition algorithms have great difficulty in dealing with non-linear image transformations.

Localized Color Shifting: The image is first divided into a randomly chosen number of orthogonal partitions. Then, on a partition by partition basis, a random value is chosen for each color channel (red, green, and blue) and the color values of the pixels contained therein are shifted by these amounts. A new set of orthogonal partitions is then chosen and the color shifting process is repeated. This results in a large number of image segments that have undergone different color shifts. This impacts local color information strongly, breaks up contiguous object color segments, and introduces a large number of discontinuities to the image.

Semi-Regular Object Clutter: The system places lines of random color and thickness in a random, semi-regular fashion over the image. The use of clutter in this stage attempts to exploit the Gestalt perception abilities of humans, namely that human have a strong ability to recognize and understand images in the face of incomplete or fragmented visual information while machines have a difficult time doing the same thing. The partial object occlusions that result are especially effective against machine vision techniques that rely heavily upon object shape, edges, or segmentation.

Localized Texture Effects: The image is first divided into a randomly chosen number of orthogonal partitions. For each of these partitions, the corresponding sub-image will be either be dithered to a random number of colors using Floyd-Steinberg error diffusion, quantized to a random number of colors via the octree color quantization algorithm, or have colored Gaussian noise added. These effects have the most effect on machine vision techniques that rely on local texture information, including many interest point detectors.

The scene shown in figure 1 is shown again in figure 2, having had the full sequence of image distortions applied during its composition.

6. EXPERIMENTAL RESULTS

6.1 System Design

A prototype of the proposed image-based CAPTCHA was implemented for use in our preliminary experiments. The system uses an object database of size 125 and a set of background images of size 100. Images created by the system are of size 640x480 pixels and each contain 3 to 5 objects chosen randomly from the object database.

A preliminary prototype of our system may be seen at http://cns.eecs.ucf.edu/captcha. A reader who is interested in our system can go to this website to see more clearly the effects of the image distortions and to experience the use of the system first-hand.

Figure 2. Scene Tagging CAPTCHA image with distortion.

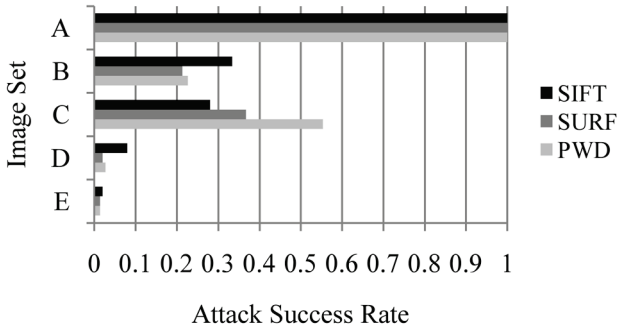

Figure 3. Success rates of automated attack techniques.

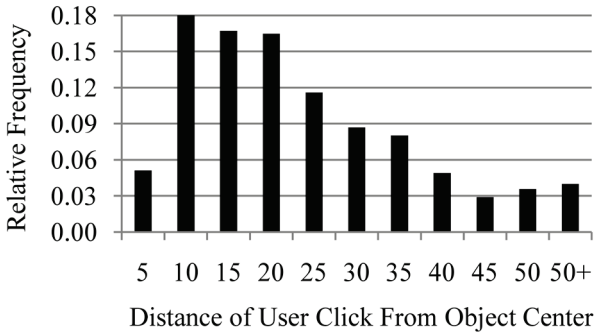

Figure 4. Relative frequency histogram of pixel distance of user answer from true object center

6.2 Automated Attack Study

In order to measure the likelihood of success via automated attack techniques, three machine vision techniques were utilized. These techniques take in the object database and a scene image, and return the likelihood of each object being present. We define an attack attempt success when the attack ranks the objects that are present in a scene as those most likely to be present.

The first automated attack technique tested is template matching via measuring the normalized pixel-wise difference (PWD) between object images and areas of corresponding size and shape in the scene tagging image. The second and third automated attack techniques utilize the feature point generators SURF and SIFT, respectively, in voting schemes based on the percentage of feature points in an object database image that have matches, filtered for uniqueness, found in the composite image.

The image sets described in Table 1 were utilized for automated attack testing. Each image set contains 100 randomly generated scene tagging instances with 3 or 4 objects present.

Table 1. Image Sets Tested

Set	Distortion Set Applied
A	No distortions, objects are placed on flat background
B	Object Scaling, mesh warping
C	Global color shifting, randomized clutter, localized color shifting, semi-regular object clutter
D	Combination of the distortions of set C and D
E	Combination of the distortions of set E and localized texture effects

The results of these automated attacks for the various image sets are presented in figure 3. The results show that with full series of distortions discussed above (data set E), automated attacks have a very low success rate. The most successful automated attack when dealing with the full set of distortions is SIFT, which is able to correctly identify the objects present in an image 2% of the time.

6.3 User Study

A user study with 20 individuals was conducted in order to measure human ability to successfully pass the test. The average user answered 63.9 questions in the time allotted. Approximately 1/2 of the problems used multiple choice and half of the questions used image point selection. Question formats were dynamically chosen based on the number of and associations between the objects present in the scene image. Users were asked to mark the images where content was so unclear that they were unable to answer the question. Table 2 shows user task success rates.

Table 2. User Study Task Success Rates

Response Format	Multiple Choice	Image Point Selection	Overall
User Success Rate	0.979	0.965	0.975

As table 2 shows, over 97% of the questions answered by users were answered correctly. This falls well within an acceptable range, as even widely implemented CAPTCHA systems such as reCaptcha find that users have problems answering a given question from 3% to 7% of the time [18]. An image was deemed to be unclear in 2.8% of cases, making a strong case for allowing a user a small number of image discards without penalty.

Figure 4 presents the distribution of image clicks with regards to distances from the center of the correct answer. These results indicate that accepting an answer that is less than or equal to a distance of 50 pixels from the center of the object will accept over 96% of the user responses within 100 pixels while only 2.56% of random guess clicks will be accepted.

During the study, the average number of seconds that elapsed between the point when a question was presented to the user and the point at which the user answered that question was 11.034 seconds. Over 99% of the user response times measured were less than 45 seconds, making it a good candidate for use as an answer

acceptance cut-off point. This may be used to make computationally expensive automated attacks infeasible. For example, our tests indicate that it takes approximately 4.436 seconds per database object image to perform a PWD based attack against a single problem on an Intel Core 2 Duo E6850 machine. Thus, if an object database of size 1000 were utilized then the task would take approximately 73.93 minutes to complete.

Table 3 displays the human user success rate and best automated attack success rate for a number of image-based CAPTCHA systems. For a fair comparison, the numbers provided are for completion of a single task and not a combination of multiple tasks. Note that the most successful known attack on all of these tasks except ASSIRA is that of randomized guessing. Where single task randomized guess attack success rates have not been presented explicitly they have been derived from the information given via straightforward geometry.

Table 3. Comparison of image-based CAPTCHA user and automated attack success rates

CAPTCHA Name	User Success Rate	Automated Attack Success Rate
Scene Tagging (Image Point Selection, 50 pixel region radius)	0.966	0.026
ASSIRA	0.99[13]	0.10[16]
What's Up CAPTCHA (single image rotation, 16 degree window)	0.94[14]	0.040
IMAGINATION (Image Center Selection, 25 pixel region radius)	0.70[17]	0.033

7. CONCLUSIONS AND FUTURE WORK

We have presented a novel form of image-based CAPTCHA that distinguishes between humans and machines based on their understanding of objects and object relationships in an automatically generated, composition-based image. Experimental results indicate that the system is secure to several automated attack techniques and a user study shows that the system has comparable usability with existing CAPTCHA systems. Based on these results, it is clear that scene tagging CAPTCHA is a viable alternative to text-based and other image-based CAPTCHA systems. Future work includes improving and expanding our automated attack testing array along with investigating the use of 3D object models in our image generation process.

8. ACKNOWLEGEMENT

We thank the anonymous reviewers for their valuable comments and suggestions. This research work was initially started from a course term project "Image and Scene Tagging: An Alternative Approach to Current CAPTCHA Techniques" conducted by graduate students Andrew Mantel and Peter Matthews in Spring 2009 in University of Central Florida.

9. REFERENCES

[1] L. von Ahn, M. Blum, and J. Langford. Telling Humans and Computers Apart (Automatically) or How Lazy Cryptographers do AI. *Comm. of the ACM, 47* (2), 57-60.

[2] H. S. Baird and J.L. Bentley. Implicit Captchas. In *Proceedings of the IST SPIE Document Recognition and Retrieval XII Conference*, (San Jose, CA, 2005), vol. 5676.

[3] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. 2008. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* 110, 3 (Jun. 2008), 346-359.

[4] Budanitsky, A., Hirst, G. Semantic distance in WordNet: An experimental, application--oriented evaluation of five measures. In *Proceedings of the North American Chapter of the Association for Computational Linguistics Workshop* (Pittsburgh, PA, USA, 2001), 29-34.

[5] Chellapilla, K., and Simard, P.Y. Using Machine Learning to Break Visual Human Interaction Proofs (HIPs). *Advances in Neural Information Processing Systems 17*, 265-272.

[6] M. Chew and J. D. Tygar. Image Recognition CAPTCHAs. In *Proceedings of the 7th Annual Information Security Conference* (Palo Alto, CA, USA, 2004), 268–279.

[7] Datta, R., Li, J., and Wang, J. Z. IMAGINATION: a robust image-based CAPTCHA generation system. In *Proceedings of the 13th Annual ACM international Conference on Multimedia* (Hilton, Singapore, 2005), 331-334.

[8] Datta, R., Li, J., and Wang, J. Z. Exploiting the Human-Machine Gap in Image Recognition for Designing CAPTCHAs. *IEEE Transactions on Information Forensics and Security, 4* (3), 504-518.

[9] Elson, J., Douceur, J.R., Howell, J., and Saul, J. Asirra: a CAPTCHA that exploits interest-aligned manual image categorization. In Proceedings of the *14th ACM Conference on Computer and Communications Security* (Alexandria, Virginia, USA, 2007), 366-374.

[10] Golle, P. 2008. Machine learning attacks against the Asirra CAPTCHA. In *Proceedings of the 15th ACM Conference on Computer and Communications Security* (Alexandria, Virginia, USA, 2008), 535-542.

[11] Gossweiler, R., Kamvar, M., and Baluja, S. 2009. What's up CAPTCHA?: a CAPTCHA based on image orientation. In *Proceedings of the 18th international Conference on World Wide Web* (Madrid, Spain, 2009), 841-850

[12] Lowe, D. G. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision* 60, 2 (Nov. 2004), 91-110.

[13] Mikolajczyk, K. and Schmid, C. 2005. A Performance Evaluation of Local Descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 10 (Oct. 2005), 1615-1630.

[14] G. A. Miller. 1990. Wordnet: a lexical database for English. *International Journal of Lexicography, 3* (4), 235-244.

[15] Moy, G., Jones, N., Harkless, C., and Potter, R. Distortion Estimation Techniques in Solving Visual CAPTCHAs. In *IEEE CVPR*, (Washington, D.C., USA, 2004), Vol. 2, 23-28.

[16] Wolberg, G. *Digital Image Warping*. IEEE Computer Society Press, Los Alamitos, CA, 1990.

[17] Yan, J. and El Ahmad, A. S. 2008. A low-cost attack on a Microsoft captcha. In *Proceedings of the 15th ACM Conference on Computer and Communications Security* (Alexandria, Virginia, USA, 2008), 543-554.

[18] Yan, J. and El Ahmad, A. S. 2008. Usability of CAPTCHAs or usability issues in CAPTCHA design. In *Proceedings of the 4th Symposium on Usable Privacy and Security* (Pittsburgh, Pennsylvania, 2008), Vol. 337, 44-52.