2D Captchas from 3D Models

Mohammed E. Hoque, David J. Russomanno, Mohammed Yeasin
Electrical and Computer Engineering Department
The University of Memphis
Memphis, TN 38152 USA
{mhoque, d-russomanno, myeasin}@memphis.edu

Abstract

Existing image-based Captchas (Completely Automated Public Tests to Tell Computers and Humans Apart) utilize huge, public image databases to limit access to websites via image matching attacks. In this paper, a different image based Captcha was developed and prototyped to address the mislabeling and other shortcomings of traditional schemes, which utilize huge public image databases. A database was populated with n 3D models. Then, random rotations, distortions, translations, lighting effects and warping were applied resulting in 2D images. Those 2D images were presented to users for identification to gain access to a website. Users are prompted to identify the object with the label provided through a menu bar. This approach creates an infinite number of 2D images from a 3D model. Therefore, it appears to be impractical for an intruder to algorithmically identify the resulting 2D image without resorting to random guesses.

1. Introduction

Networked computers are susceptible to cyber attacks by software robots and spiders posing as humans to gain access to and possibly abuse systems intended for human use only. Prevention of such attacks has motivated the study of a new security protocol called Captchas: Completely automated public tests to tell computers and humans apart [1, 2].

Captchas are currently used to prevent automated acquisition of email accounts, inserting web pages into search engines, skewing online polling, and browsing websites by software robots. Among the existing Captchas, the text-based ones are currently most widely used. In a text-based Captcha, users are presented with a random string of distorted letters and asked to identify them. With the advancement of machine-vision technology, only a few text-based Captchas [3, 4, 5] have provided strong resistance to attacks while maintaining a decent legibility rate. Along with the text-based Captchas, image-based Captchas have shown great promise since humans have extensive background knowledge and context to identify objects much better than machines. This ability gap between humans and machines is exploited by image-based Captchas.

The traditional image based Captchas employ huge public image databases to randomly select a couple of images of a random object and asks user to identify it. Unfortunately, huge public image databases, such as images.google.com, often contain mislabeled images yielding false images for any specific object. A different solution to the mislabeling problem in public databases is proposed in this paper by generating a unique 2D representation of any 3D models, from a precompiled correctly labeled database, by applying random distortions, rotations and lighting variations.

2. Background

In November 1999, http://www.slashdot.com released an online poll asking which university has the best computer science program. The IP addresses of the voters were traced to prevent one user from casting multiple votes. However, a few students from Carnegie Mellon University (CMU) developed a program which voted for CMU thousands of times. Similarly, the next day, students from MIT wrote their own program to cast vote for MIT. After the end of the poll, MIT and CMU finished with 21,156 and 21,032 votes, respectively, whereas other universities ended with less than 1,000 votes. This incident highlighted the problems involved with online polling [1].

Most of the companies who provide free email accounts, such as Yahoo and Microsoft, suffer from malicious programs that automatically sign up for email accounts and use those accounts to spam. Prevention of such abuses requires a protocol to differentiate between human users and bots. Captchas are one approach to this problem.

Web pages that are unindexed are intended to be isolated so that others may not locate these pages using search engines. There is an html tag available that requests crawlers not to read the web pages. This html tag works against bots of the large companies, but does not absolutely guarantee that a bot will not enter the site. Therefore, a security protocol like Captcha is necessary to limit the crawlers' access on the web [1].

It has been recommended in [6] that using Captchas may also prevent dictionary attacks to retrieve passwords in which bots iteratively try different combinations of alphanumerical variables to guess passwords.

3. Related Work

In 1997, Altavista first developed Captchas to prevent offensive submission of URLs to their search engines by software robots [7]. Altavista presented an image-based text for the users to read for verification. These images were difficult for the state-of-the-art machine-vision systems to identify at that time. Their effort reduced spam additions to the search engines by over 95%. Since then, many others Captchas have been developed including CMU's EZ-Gimpy [1, 8], PARC's PessimalPrint [9], BaffleText [3], Paypal's Captcha (www.paypal.com), and Microsoft's Captcha [10].

Building an effective Captcha, which is easy for an average user to identify, but is difficult for machine-vision systems, is an elusive goal and still remains an active area of research. Among the several types of Captchas that already exist, image-, text- and speech-recognition-based Captchas have gained the most popularity. However, most of the text-based Captchas in use are either identifiable by machines or insufficiently studied [3, 11, 12, 13]. Therefore, a few studies [1, 11] have been conducted in image-based Captchas due to the legibility and vulnerability issues involved with text-based Captchas.

CMU's Captcha website [1] has two different prototypes of image-recognition Captchas. However, both of them use a limited number of images and thus are susceptible to a simple brute force image-matching algorithm. Chew et al. [11] introduced the image-based Captchas where the Google database is queried with random words from a precompiled dictionary. A couple of images are retrieved from the Google database and presented to the user asking the open question of "What do you see?" as shown in Figure 1.

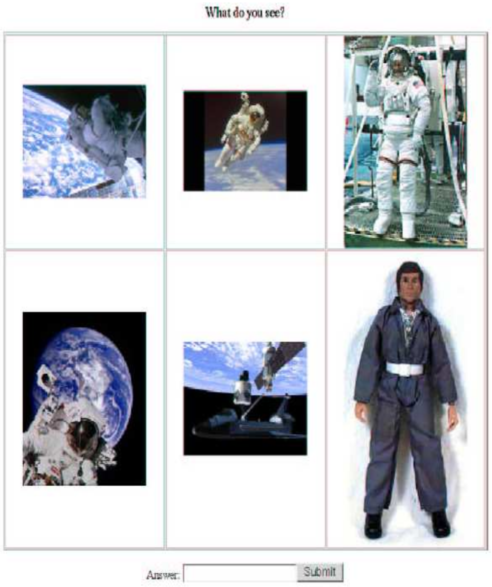

What do you see?

Figure 1. The naming Captcha [11].

However, if the image database is publicly available, the database can be algorithmically searched resulting in the automated classification of the image. Also, huge image databases, such as images.google.com, often generate incorrect labels for any given image. These mislabeled images have the adverse affect of causing humans to be less likely to pass the Captcha, since they will correctly identify the images, thus failing the Captcha. However, computers that perform database queries are unaffected by mislabeled images, that is, if the computer is capable of finding the matching image in the database, it will also have the matching, although incorrect, label.

4. Design

In this paper, a different approach to the image-based Captchas is implemented which addresses the mislabeling problem without maintaining a huge image database. The logical steps are as follows:

- Create a database of 3D models along with their corresponding labels
- Randomly select a model from the database
- Generate an intermediate 2D image of this model using random translation, rotation, scaling and lighting effects
- This final, 2D distorted image is presented to the user, who is required to label the object using a menu bar

The goal of this design was not to be able to reject improper access 100% of the time. If the Captcha raises the cost of using a software robot above the cost of using a human, the approach can be regarded as effective [11].

Thirty 3D models were carefully selected so that humans can easily recognize them regardless of their size, orientation, lighting and distortion within reasonable values. For each test, one of the models is randomly chosen to apply rotation, translation, scaling, warping in random order with random parameters. The warped image was merged with the original image to further complicate the transformation. Variation of lighting effects is also applied to the 2D images generated from the 3D models to provide resistance against color and texture based image retrieval attacks. A high-level diagram of the random sequence of operations available on the 3D models is shown in Figure 2. The Visualization Toolkit (VTK) [17] was used to implement our prototype. Figure 3 shows the diagram describing the VTK pipeline used to achieve the desired goal of distorting images randomly to make it more difficult for automated identification while preserving the ability for human recognition.

Due to the nature of the Captcha, the objects are not going to resemble the real world objects that they are modeled after. If a human is given a distorted and oddly lit object, and asked "what do you see?" the human will likely be overwhelmed by the possibilities. This also opens the door to a huge variety of free format answers. Therefore, a list of possible objects is provided to eliminate any confusion.

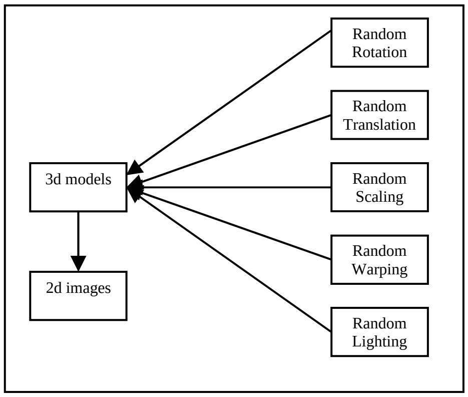


Figure 2. Operations from 3D to 2D.

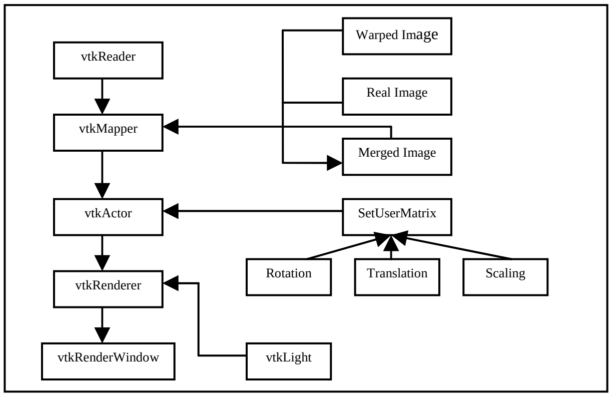


Figure 3. VTK Pipeline to generate 2D images.

Implementation of minor random distortions were similar to those applied to traditional 2D image based Captchas. However, to further complicate feature extraction, there may exist distortions unique to 3D based Captchas that will have great impact on feature extraction while only having minor impact on human performance.

The mapping from a 3D to a 2D image is not one-to-one; therefore, a database query based on the 2D image for the original 3D object is non-trivial. Converting a 2D image containing random distortion to its native 3D model is a difficult problem in machine vision, which makes it hard to search through the 3D model database for the correct match. As a result, the proposed 3D object based image classification Captcha scheme does not have the deficiencies as those used in [1, 11] for image classification.

A method of attack could involve the generation of a feature database based on the original 3D models. However the features extracted from the image will be distorted to varying degrees based on the random rotations, translations, scaling and morphing of the models used to generate the image.

Since the features will not be accurately extracted, an attacker will most likely be unable to perform range queries on the extracted features without allowing for a large number of false positives. Current appearance based 3D recognition methods [14, 15] tend to be unreliable and computationally expensive for objects rotated in random and with the presence of occlusions, etc. The common weaknesses of these approaches, as well as other known 3D appearance based recognition methods, can be exploited when performing the rotations, transformations, lighting and distortions, so that these 3D recognition methods perform at near worst case in terms of accuracy.

An attacker could employ the color based image retrieval as explained in [16] to match the 2D image with the corresponding 3D models. However, all the models in our work were tested for having somewhat uniform distribution of luminosity, so that the 2D images can not be indexed by dominant color regions [16]. See the examples in Fig. 4.

5. Evaluation and Future Work

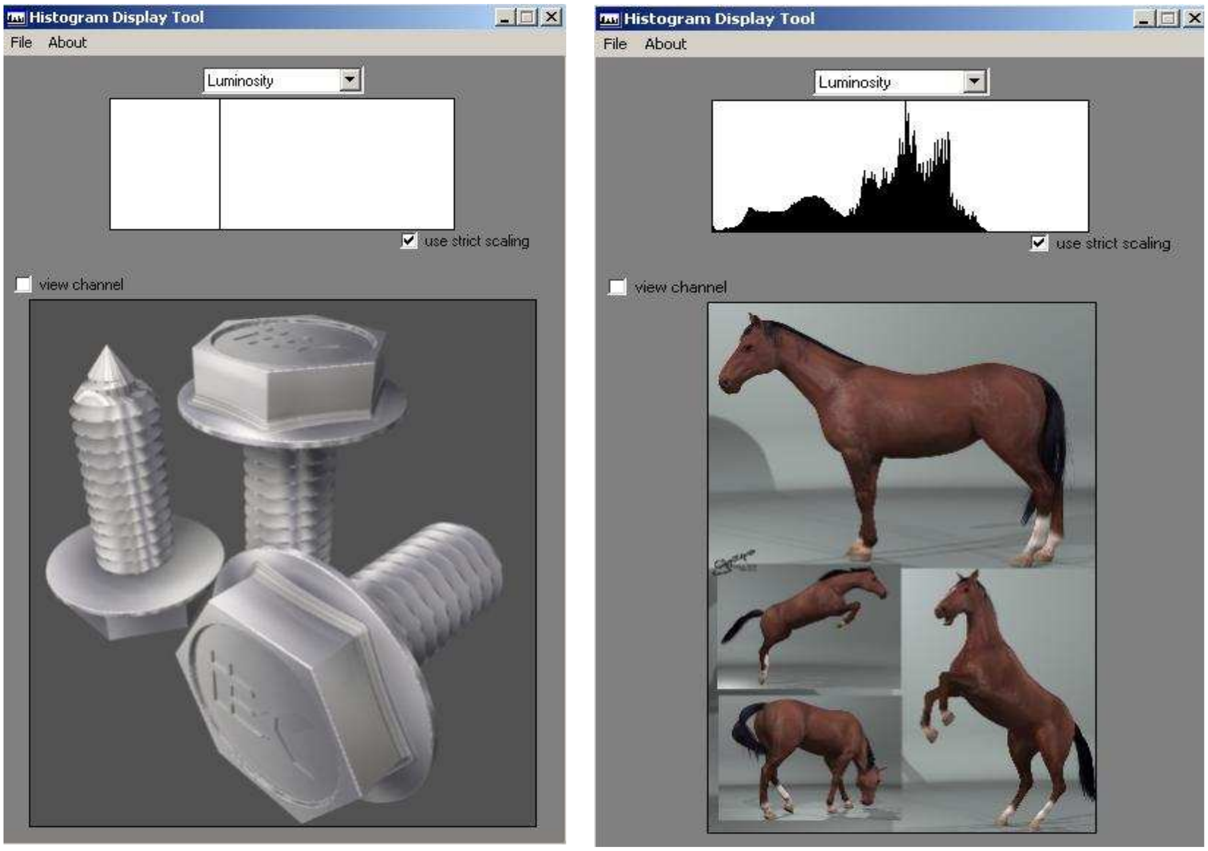
Evaluation of Captchas is the most important yardstick in determination of their effectiveness. Captchas must satisfy the following three basic properties [11]:

1. Easy for humans to identify
2. Easy for a tester machine to generate and grade
3. Hard for a software robot to pass. The only automaton that should be able to pass a Captcha is the one generating the Captcha.

Our image based Captcha is very easy to generate and grade yet very difficult for a software robot to pass within a limited amount of time. Apart from the complexity involved in converting back to 3D models from 2D images with random distortions, the number of images that a program has to traverse is simply enormous. All of the distortion parameters are limited such that the resultant images are still distinguishable by a human. Considering the huge number of generated 2D images, an attacker may not be able to reduce the number of possible choices of objects by a significant amount; therefore, resorting to random guessing will be the most likely approach. The probability of passing the Captcha with random guesses, ignoring the distortions, is as follows:

$$N = (N_m \times N_w \times O \times \frac{m!}{(r! \times t! \times s! \times l!)})^{-1}$$

N_m = No. of models
N_w = No. of variations in warping
O = No. of orders of performing distortions
r = No. of possible rotations along x, y and z axes
t = No. of possible translations along x, y and z axes
s = No. of possible scaling
l = No. of possible lighting variations
m = $r + s + t + l$



Luminosity

☑ use strict scaling

□ view channel

Luminosity

☑ use strict scaling

□ view channel

(a) (b)

Figure 4. (a) Bad example: gray values are localized which is an indication of dominant color; (b) Good example: gray values are distributed over the whole dynamic range. Such an image is a difficult candidate for modeling only dominant color.

For example, for 30 models, and 50 different warping variations, 24 different orders for four operations, 50 (it is actually way more than 50) different possibilities separately for rotations, translations and scaling (limiting the ranges of x, y, and z to 0 to 50), 50 different lighting variations, the least possible images that an attacker might have to traverse is 2.33e^{+314}. This provides a probability of 1.83e^{-139} to pass the Captcha with random guesses.

Answering the question of whether this Captcha is easy for humans to identify requires conducting perception experiments with humans of various backgrounds and ages. Numerous experiments need to be conducted including limiting the range of parameters of the distortions to the threshold where it becomes difficult for humans to identify objects.

Future work may include incorporating a common theme with the 3D models to increase human's accuracy to identify the objects and decrease their response time. For example, a common theme may be "an office environment," where objects presented will be commonly found in an office. Because the object database will be public, the attacker will already have their object possibilities limited, thus using a theme will have no favorable affect on an attacker's speed or accuracy. Textures that are overly unique should be avoided, since they could be used as clues by an attacker to eliminate possible models that were used in forming the image.

The probability of finding the correct image in the database can be reduced by requiring a response to the Captcha in a period of time that is sufficient for a human, but insufficient for adequate database traversal. Objects that can be identified with multiple names should have each name present on a list so that user is not frustrated if their initial name for the object is not on the list. The images that were used in developing our prototype were taken from www.**3d**shop.com for demonstration purpose only.

Our proposed image-based Captcha has demonstrated that it is possible to generate a huge database which not only solves the mislabeling of images problems, but also provides a scheme for improved security against cyber attacks.

Figure 5 shows a few examples of Captchas after transformations were applied to 3D models in random order. Figure 6 shows an example of the prototype's user interface where a user is asked to identify the correct object through the menu bar.

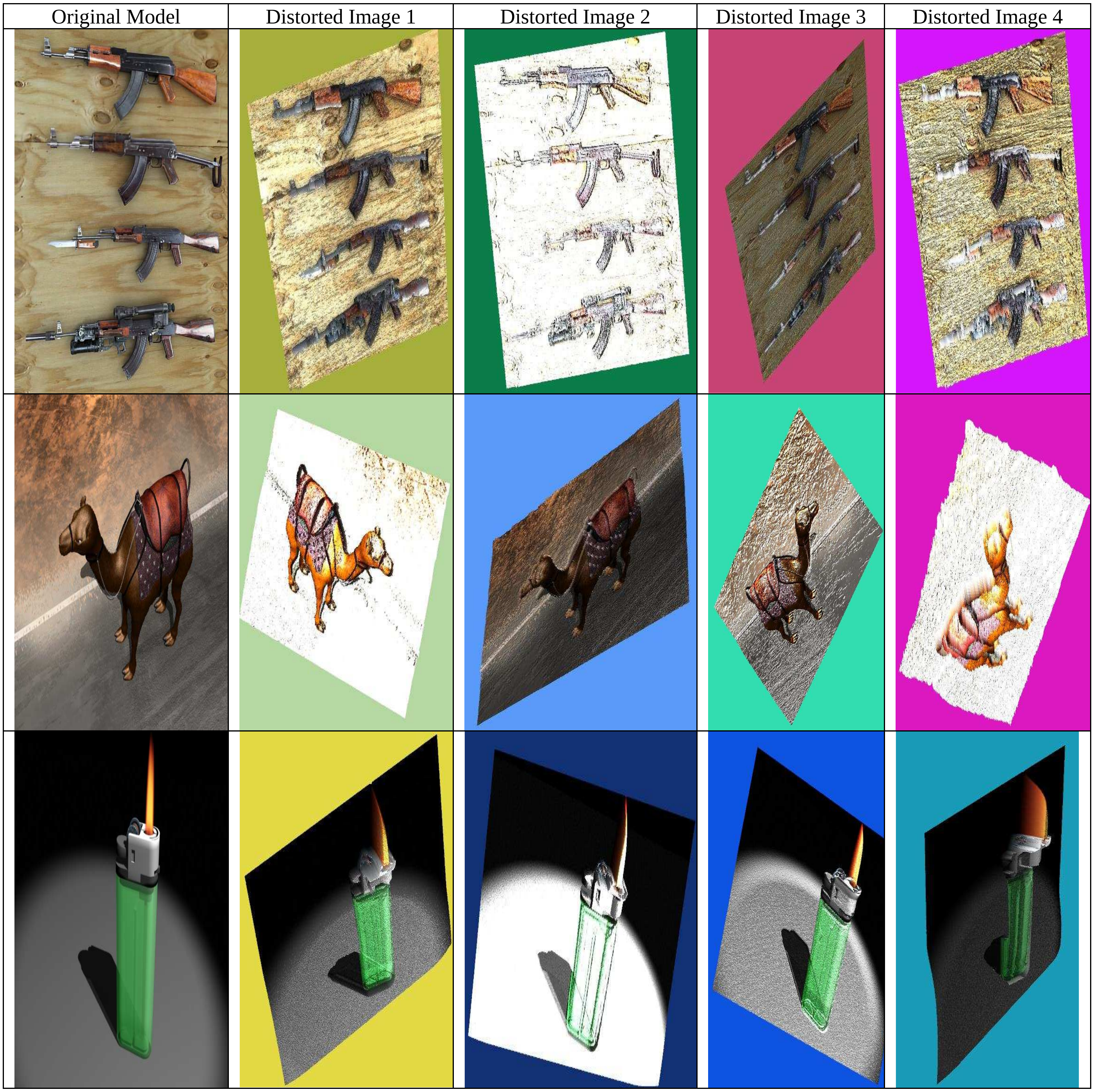
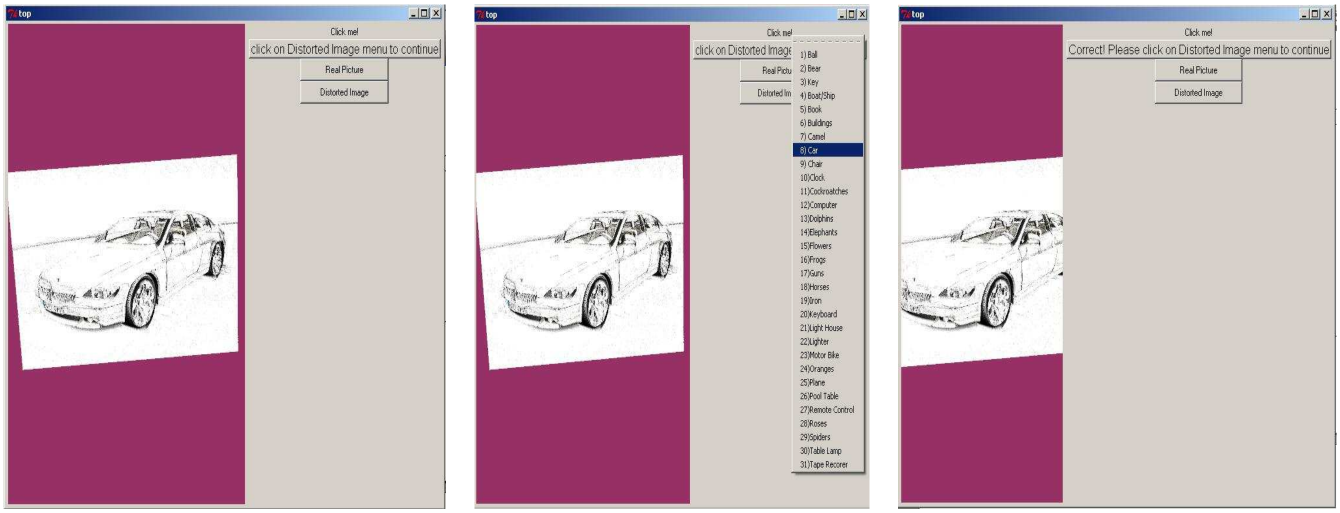


Figure 5. A few examples after applying various transformations in random order.



click on Distorted Image menu to continue

Correct! Please click on Distorted Image menu to continue

(a) (b) (c)

Figure 6. (a) 2D image is presented to user; (b) Menu bar is used to identify the correct object; (c) Object identified.

6. References

[1] M. Blum, L. Ahn, J. Langford, and N. Hopper, The CAPTCHA Project, http://www.captcha.net, November 2000.

[2] L. von Ahn, M. Blum, N. Hopper, and J. Langford, “Captcha: Using hard AI problems for security,” In *Advances in Cryptology Eurocrypt '03* (Volume 2656 of Lecture Notes in Computer Science), pp. 294-311, 2003.

[3] M. Chew and H. S. Baird, “BaffleText: a Human Interactive Proof," *Proc. 10th SPIE/IS&T Document Recognition and Retrieval Conf. (DRR2003)*, Santa Clara, CA, January 23-24, 2003.

[4] H. S. Baird and T. Riopka, "ScatterType: a Reading CAPTCHA Resistant to Segmentation Attack," *Proc. SPIE/IS&T Conf. on Document Recognition and Retrieval XII* (*DR&R2005*), San Jose, CA, January, 2005.

[5] H. S. Baird, Michael A. Moll, and Sui-Yu Wang, "ScatterType: a Legible but Hard-to-Segment CAPTCHA," *Proc. IAPR 8th Int'l Conf. on Document Analysis and Recognition*, Seoul, Korea, August 29 - September 1, 2005.

[6] B. Pinkas and T. Sander, “Securing Passwords Against Dictionary Attacks” *Proc. of the ACM Computer and Communications Security Conference*, pp. 161-170, Washington, DC, USA, November, 2002.

[7] AltaVista's \Add-URL site: altavista.com/sites/addurl/newurl, protected by the earliest known CAPTCHA.

[8] N. J. Hopper and M. Blum, “Secure Human Identification Protocols," *In: C. Boyd (Ed.) Advances in Crypotology, Proc. of Asiacrypt 2001, LNCS 2248*, pp. 52 -66, Springer-Verlag, Berlin, 2001.

[9] A. L. Coates, H. S. Baird, and R. Fateman, “Pessimal Print: a Reverse Turing Test," *Proc. IAPR 6th Intl. Conf. on Document Analysis and Recognition*, pp. 1154-1158, Seattle, WA, September 10-13, 2001.

[10] P. Y. Simard, R. Szeliski, J. Benaloh, J. Couvreur, I. Calinov, “Using Character Recognition and Segmentation to Tell Computer from Humans," *Proc. IAPR Int'l Conf. on Document Analysis and Recognition*, Edinburgh, Scotland, August 4-6, 2003.

[11] M. Chew and J. D. Tygar, “Image Recognition CAPTCHAs”, *Proc. of the 7th Annual Information Security Conference (ISC’04)*, pp. 268–279, Palo Alto, CA, September 2004.

[12] G. Mori and J. Malik, “Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’03)*, Vol. 1, pp. 134-141, June 18-20, 2003.

[13] G. Moy, N. Jones, C. Harkless, and R. Potter, "Distortion Estimation Techniques in Solving Visual CAPTCHAs," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, Vol. 2, pp. 23-28, 2004.

[14] P. Mittrapiyanuruk, G. N. DeSouza and A. C. Kak, “Calculating the 3D-Pose of Rigid Objects Using Active Appearance Models," *Proc. International Conference in Robotics and Automation*, May 2004, New Orleans, LA.

[15] P. Mittrapiyanuruk, G. N. DeSouza and A. C. Kak, "Accurate Rigid Objects with Occlusion Using Active Appearance Models," *The IEEE Workshop on Motion and Video Computing*, January 2005, Breckenridge, CO, USA.

[16] B. G. Prasad, K. K. Biswas and S.K. Gupta, “Region based image retrieval using integrated color, shape & location index,” *Computer Vision & Image Understanding*, Vol. 94, pp. 193-233, January, 2004.

[17] W. Schroeder, K. Martin and W. Lorensen, *The Visualization Toolkit Third Edition*, Kitware Inc., Clifton Park, NY, 2004.