Image-based face detection CAPTCHA for improved security

Brian M. Powell and Adam C. Day

Department of Computer Science and Electrical Engineering,
West Virginia University,
Morgantown, WV 26506, USA
E-mail: brian.powell@mail.wvu.edu
E-mail: aday2@mix.wvu.edu

Richa Singh and Mayank Vatsa

Indraprastha Institute of Information Technology (IIIT) Delhi,
New Delhi, 110078, India
E-mail: rsingh@iiitd.ac.in
E-mail: mayank@iiitd.ac.in

Afzel Noore*

Department of Computer Science and Electrical Engineering,
West Virginia University,
Morgantown, WV 26506, USA
E-mail: afzel.noore@mail.wvu.edu
*Corresponding author

Abstract: In this paper, we propose a novel image-based completely automated public turing test to tell computers and humans apart (CAPTCHA) that relies on detecting human faces to provide an additional layer of security in web-based services. Face images were selected from the CMU face database and subjected to different types of distortions at different intensity levels to make the automatic face detection very challenging. An extensive experimental study involving 1,100 individuals was undertaken to determine the efficacy of the proposed approach and evaluate the performance of humans compared to computers. We also used two image quality metrics to objectively study the characteristics of the composite CAPTCHA images. Unlike a text-based CAPTCHA, a major benefit of the proposed image-based face detection CAPTCHA is that it does not have any language barriers. In addition, the proposed CAPTCHA can easily be implemented on handheld devices to provide an additional level of security.

Keywords: completely automated public turing test to tell computers and humans apart; CAPTCHA; web security; face detection, human-interactive proof.

Reference to this paper should be made as follows: Powell, B.M., Day, A.C., Singh, R., Vatsa, M. and Noore, A. (2010) 'Image-based face detection CAPTCHA for improved security', Int. J. Multimedia Intelligence and Security, Vol. 1, No. 3, pp.269–284.

Biographical notes: Brian Powell received his MS in Computer Science from West Virginia University, USA in 2006. He is currently a Doctoral candidate in the Lane Department of Computer Science and Electrical Engineering at West Virginia University. His areas of interest are human interactive proofs, human computation, user interface design and computer science education. He is a member of the IEEE, Computer Society and the Association for Computing Machinery. He is also a member of the Upsilon Pi Epsilon and Sigma Zeta honour societies. He was the recipient of the West Virginia University Foundation Distinguished Doctoral Fellowship.

Adam Day received his MS in Electrical Engineering from West Virginia University, USA in 2010. He is currently a System Administrator/Software Engineer with Lockheed Martin. His areas of interest are signal and image processing with an emphasis on digital image watermarking and computer recognition. He is a member of the Eta Kappa Nu and Tau Beta Pi honour societies.

Richa Singh received her MS and PhD in Computer Science from West Virginia University, USA in 2005 and 2008, respectively. She is currently an Assistant Professor at the Indraprastha Institute of Information Technology (IIIT) Delhi, India. Her areas of interest are biometrics, pattern recognition and machine learning. She has more than 90 publications from refereed journals, book chapters and conferences. She is a member of the Golden Key International, Phi Kappa Phi, Tau Beta Pi, Upsilon Pi Epsilon and Eta Kappa Nu honor societies. She is the recipient of seven best paper and best poster awards in international conferences.

Mayank Vatsa received his MS and PhD in Computer Science from West Virginia University, USA in 2005 and 2008, respectively. He is currently an Assistant Professor at the Indraprastha Institute of Information Technology (IIIT) Delhi, India. He has more than 90 publications from refereed journals, book chapters and conferences. His areas of interest are biometrics, image processing, watermarking and information fusion. He is a member of the Golden Key International, Phi Kappa Phi, Tau Beta Pi, Upsilon Pi Epsilon and Eta Kappa Nu honor societies. He is the recipient of seven best paper and best poster awards in international conferences.

Afzel Noore received his PhD in Electrical Engineering from West Virginia University, USA in 1987. From 1996 to 2003, he was the Associate Dean for Academic Affairs and Special Assistant to the Dean in the College of Engineering and Mineral Resources, West Virginia University. He is currently a Professor in the Lane Department of Computer Science and Electrical Engineering. His research interests include applications of machine learning and computational intelligence techniques. He has over 100 publications in refereed journals, book chapters and conferences and is the recipient of six best paper and best poster awards in international conferences.

1 Introduction

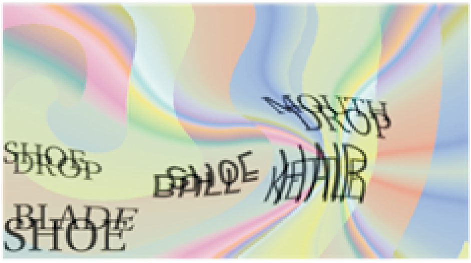
The pervasive growth of web-based services such as free e-mail, search engines and surveys, have transformed the way people communicate and share information. However, these services are exploited by spammers using computers to automatically create multiple online accounts, skew survey results, send unsolicited spam messages to e-mails, weblogs, and message boards, or guess usernames and passwords. Computers, unlike humans, have the advantage of performing such tasks quickly and in large volumes. To prevent these automated tasks, a program called completely automated public turing test to tell computers and humans apart (CAPTCHA) is used to provide an additional layer of security. There are different types of CAPTCHAs in use today that are based on text, images, audio, and video. An overview of existing CAPTCHA implementations is described.

1.1 Text-based CAPTCHAs

The most commonly used CAPTCHAs are text-based where distorted text is displayed. To solve the CAPTCHA, users must recognise the distorted characters and correctly enter them in a designated space. Text-based CAPTCHAs are easy to generate but are vulnerable to optical character recognition (OCR) attacks (Chellapilla et al., 2005; Kluever, 2008).

An early research-based CAPTCHA called GIMPY was developed by Carnegie Mellon University in 2000 (Baird and Popat, 2002; Kluever, 2008). In this CAPTCHA, seven English words are randomly selected and then have their character outlines warped. The text is presented in overlaid pairs on a colourful noisy background. Users are asked to type a certain number of these words, although variants exist that just use one word or string of characters (von Ahn et al., 2004; Baird and Popat, 2002; Kluever, 2008; Mori and Malik, 2003; Moy et al., 2004). Figure 1 shows an example of a GIMPY CAPTCHA used as an extra layer of security.

Figure 1 Example of a GIMPY CAPTCHA

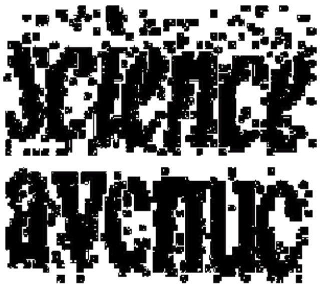


Source: von Ahn (2005)

GIMPY was one of the first CAPTCHAs to be systematically attacked from a research perspective. Mori and Malik proposed two different methods of using histograms to differentiate characters from the surrounding background. They then compared character shapes and outlines to guess the text. Using this method, they achieved a 92% success rate in identifying one word (as in EZ-GIMPY) and 33% rate in identifying three words as used by the main GIMPY version (Mori and Malik, 2003). By leveraging knowledge of the dictionary used by EZ-GIMPY, Moy was able to successfully attack with an accuracy of 99% (Moy et al., 2004).

Another form of text-based CAPTCHA called Pessimal Print was developed in 2001 and is shown in Figure 2. It randomly selects English words from a predefined list, modifies the character outlines by thickening or thinning lines, and then applies distortions such as salt-and-pepper background noise or blurring in an attempt to make OCR attacks difficult (Chew and Baird, 2003; Coates et al., 2001; Rice et al., 1999).

Figure 2 Pessimal Print CAPTCHA with thickening and high background noise

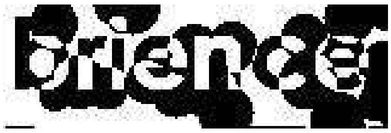


Source: Coates et al. (2001)

In developing the CAPTCHA, Coates and Baird evaluated a test set of 685 generated images on human volunteers. All were deemed to be human-legible. The fixed list of English words and fonts made this CAPTCHA vulnerable to attack. In computer tests, an attack rate of approximately 40% to 50% was achieved (Chew and Baird, 2003; Coates et al., 2001).

BaffleText CAPTCHA relies on image masking to help achieve its security. A black-and-white image mask consisting of circles, squares, and ellipses is created. A pronounceable non-dictionary word is generated and then placed on the background. Difference masking is applied for cases where black pixels from the background and text overlap, yielding an image as shown in Figure 3. In human testing, BaffleText achieved an 89% success rate with the average attempt taking 8.7 seconds, and computer attack rates were approximately 25% (Chew and Baird, 2003).

Figure 3 Example of a BaffleText CAPTCHA

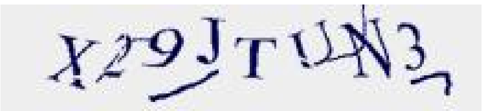


Source: Chew and Baird (2003)

Microsoft Research developed a CAPTCHA that combines traditional CAPTCHA techniques such as character warping and rotation with adding background arcs that connect multiple characters as shown in Figure 4 (Kluever, 2008; Simard et al., 2003).

However, Yan and Salah developed a process that could attack it with 90% accuracy (Simard et al., 2003; Yan and Salah, 2008).

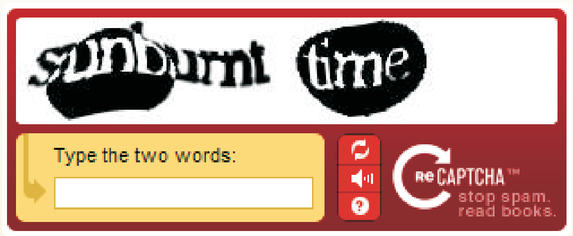
Figure 4 Example of the Microsoft Research CAPTCHA (see online version for colors)



Since text-based CAPTCHAs are attacked by using optical character recognition techniques, OCR-resistant CAPTCHAs have been developed. These CAPTCHAs use the text that has previously failed attempts at performing optical character recognition. It includes handwritten CAPTCHAs based on text from sources such as the US mail (Rusu and Govindaraju, 2004).

One of the most popular CAPTCHAs today is reCAPTCHA (2010), which uses words that were first scanned for book digitisation projects. reCAPTCHA presents users with a pair of words and asks the user to identify both (von Ahn et al., 2008). Originally, these words were modified by applying warping techniques although this has recently been paired with BaffleText-like image masking as shown in Figure 5 (von Ahn et al., 2008; reCAPTCHA, 2010).

Figure 5 The current version of reCAPTCHA incorporates image masking (see online version for colours)



Source: reCAPTCHA (2010)

1.2 Image-based CAPTCHAs

Existing image-based CAPTCHAs generally rely on image classification where users are presented with a series of images and asked to identify the relationship between them. One such CAPTCHA is ESP-PIX, which displays four images and asks users to select a common description from a drop-down list (Carnegie Mellon University, 2004). Since the image categories are selected from a fixed list, there is a high likelihood of random guessing yielding a correct answer.

The Asirra image-based CAPTCHA uses a closed database of animals from Petfinder.com (Elson et al., 2007). Users are asked to select all images of cats from a mixed set of 12 cats and dogs drawn from a large source database of over three million images. While random selection only has a 0.02% chance of correctly selecting the cats (Microsoft, 2010), Asirra is vulnerable to attack by a classifier trained to differentiate between cats and dogs with 82.7% accuracy (Golle, 2008).

1.3 Video-based CAPTCHAs

These CAPTCHAs use videos rather than static images or text. In one such CAPTCHA, users are shown YouTube videos and asked to tag them with descriptive keywords. In tests, humans achieved 90% accuracy while computer attack rates were approximately 13% (Kluever, 2008; Kluever and Zanibbi, 2009).

1.4 Audio-based CAPTCHA

For visually-impaired users, some websites have developed audio-based CAPTCHAs. These generally work by playing a recording of a set of words or characters with users being asked to type-in what they hear. Unfortunately, these CAPTCHAs are subject to attacks using speech recognition software (Bursztein and Bethard, 2009; Santamarta, 2008; Tam et al., 2008). The attack rate on audio CAPTCHAs used by Google and Digg was around 71% (Tam et al., 2008).

This paper presents a novel image-based CAPTCHA that relies on detecting human faces in a heterogeneous composite CAPTCHA image. Besides human faces, there are also faces of animals embedded in the CAPTCHA to impede face detection software in reliably distinguishing human faces. To make detection more challenging, the face images are distorted. An extensive experimental study is undertaken to evaluate the performance of humans compared to computers. This provides valuable insights to designing future CAPTCHAs for enhanced security. Section 2 discusses the proposed image-based face detection CAPTCHA design and different distortion techniques used. Section 3 describes two high-level human visual system (HVS) image quality metrics to objectively study the characteristics of images where humans performed better and where computers performed better. Section 4 analyses the data collected from the experiments. Finally, Section 5 discusses design considerations for implementing similar face detection CAPTCHAs.

2 Proposed CAPTCHA design

In recent years, improved optical character recognition techniques have successfully attacked text-based CAPTCHAs. As a result, the design of text-based CAPTCHAs has progressively become more complex to make it difficult for optical character recognition programs to successfully attack. At the same time, it has become challenging for humans to successfully solve CAPTCHAs on the first try. The proposed image-based clickable CAPTCHA presents the user with a composite image that includes several embedded human and non-human faces. The images are visually distorted and randomly placed on a noisy background. To successfully solve the CAPTCHA, the user must correctly click on all human faces. The proposed approach has several benefits over existing CAPTCHAs. Compared to text-based CAPTCHAs, our methodology avoids the continuing escalation in difficulty caused by improved OCR technology. It also avoids potential language barriers since there is no text used in the CAPTCHA, making the proposed image-based face detection CAPTCHA language-independent and can therefore be deployed to a large global audience. Since the proposed CAPTCHA does not require a keyboard, it can easily be used on handheld devices which lack a (convenient) keyboard. Compared to existing image-based CAPTCHAs, our proposed CAPTCHA does not rely on small classification sets like ESP-PIX or have a high likelihood of success by random guessing by computers using automated algorithms. Human face detection is a complicated task, especially when the faces are distorted.

Very little work has been done on image-based face detection CAPTCHAs. While there is an existing CAPTCHA that makes use of facial images, it is substantially different from our approach (Misra and Gaj, 2006). This existing CAPTCHA uses two copies of the same face image that appear different. The user is presented with a set of several face images and is asked to identify the two images that are based on the same original face. A major limitation of this approach is the small set of images presented to the user. A random guess has a high likelihood of being correct. Since the CAPTCHA makes repeated use of the same original images, it is possible for more advanced face detection technologies to perform matches. More importantly, there were no experiments conducted to quantitatively determine the efficacy of their approach.

Figure 6 Sample of the rendered background

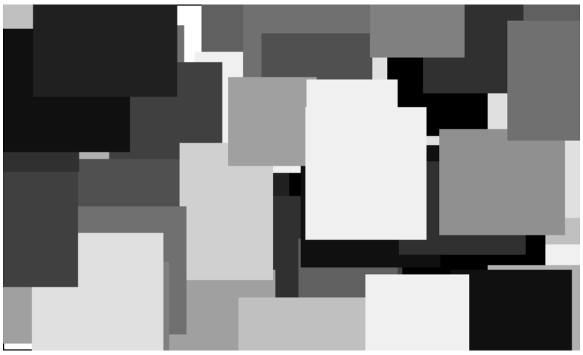


Table 1 Distortion types and associated parameter values for different distortion levels

Distortion type	*Parameters adjusted*	*Distortion level*		
		Low	*Medium*	*High*
Blurring	Standard deviation	4	8	20
Closing	Radius	3	5	-
Erosion	Radius	3	-	-
Laplacian filtering	a	-	10	5
Lightening	Histogram max-range	-	-	0.3
Periodic noise	% of image removed	67%	-	-
Piecewise scaling	Scale factor	2:1	3:1	-
Resolution modification	Scale factor	1:4	1:8	1:10
Rotation	Degrees rotated	-	90	180
Width scaling	Scale factor	4	5	-
Height scaling	Scale factor	2.5	3	4
Speckle noise	Variance	0.2	1	-

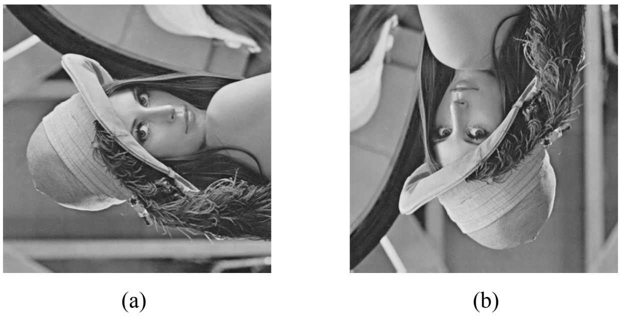
Creating the proposed image-based face detection CAPTCHA is a multi-step process. First, a subset of Carnegie Mellon University front face image database is used (Carnegie Mellon University, 2002). This is augmented by images of animal faces collected from Flickr (2010) and converted to grayscale images. We next create a 500 × 300 pixel background image consisting of a series of grayscale rectangles randomly superimposed over each other as shown in Figure 6. The noisy background is designed to thwart the effectiveness of using edge detection to identify the outline of embedded face images. We apply different distortion techniques and vary the distortion intensity by adjusting different parameters. Table 1 lists the distortion types and the parameters used in the experiment. The final CAPTCHA is designed by embedding at least one human face and at least one non-human face image in the background image.

As an example, the rotation distortion can be applied to an image using the equation,

$$R(x,y) = \begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} cos(a) & -sin(a) & 0 \\ sin(a) & cos(a) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

where, x and y represent the original coordinates and a is the angle of rotation converted to radians. In generating the 90° and 180° rotations, a is varied between 0.5π and π. When the rotation distortion is applied to the Lena image, the resulting images are shown in Figure 7.

Figure 7 Lena image with applied rotation distortions (a) $a = 0.5\pi$ (b) $a = \pi$


(a)

(b)

Another example of distortion is blurring that uses the 2D Gaussian equation,

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

where x represents the horizontal distance from the origin, y represents the vertical distance from the origin and σ is the standard deviation. The value of σ is set to 4, 8 and 20 to create low, medium and high distortion levels. When the Gaussian blur distortion is applied to the Lena image, the resulting distorted images are shown in Figure 8.

Different types of distortions at different intensity levels are applied to selected face images in the CMU database resulting in approximately 350 transformed images. Each image is randomly scaled and embedded on the background image. The distorted face images are randomly placed on the background such that no two images overlap each other. Figure 9 shows an example of the final CAPTCHA generated with the process described above.

The coordinates of human face bounding box are calculated and stored for each CAPTCHA. These reference coordinates are then compared against the coordinates of user clicks to determine if the CAPTCHA has been successfully solved. For the purposes of testing, the proposed CAPTCHA is deployed on the web for registration purposes. The web server has access to a database containing the CAPTCHA images and reference coordinates of the bounding boxes for each face image. When a user accesses the website containing the CAPTCHA, the web server will randomly select a CAPTCHA to display. After reading the instructions on how to complete the CAPTCHA, users use their mouse to click on all human faces. JavaScript code logs the coordinates of their clicks, which are compared with the reference coordinates of the bounding box for that CAPTCHA. If the user clicks on all human faces without clicking on any non-face object, the attempt will be treated as a success; otherwise it will be handled as a failure.

Figure 8 Lena image with applied Gaussian blur (a) $\sigma = 4$ (b) $\sigma = 8$ (c) $\sigma = 20$

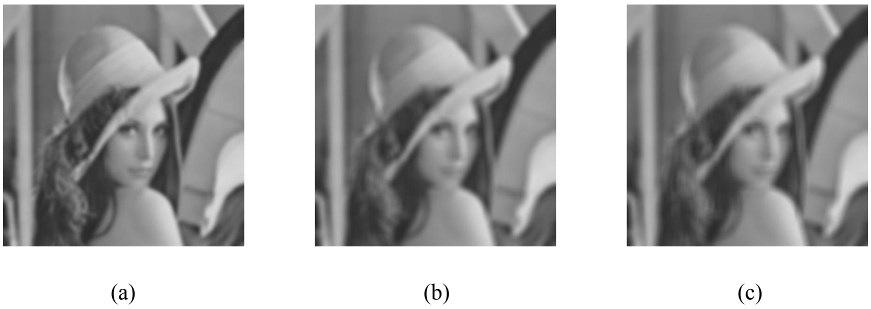
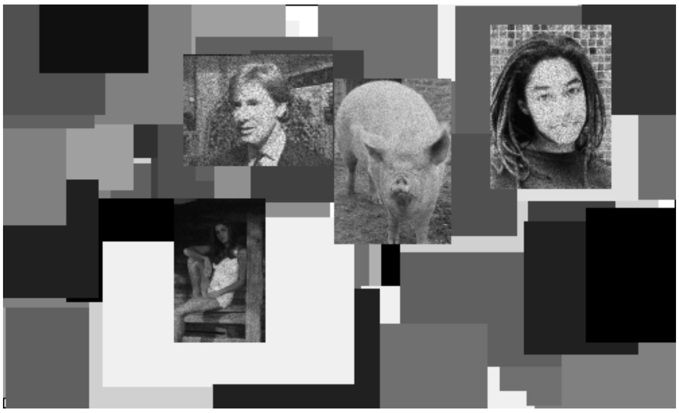


Figure 9 Example of final image-based face detection CAPTCHA



3 Image quality metrics

The final composite image-based face CAPTCHAs are analysed using image quality metrics to study the correlation between the objective measurements of image degradation and the success rates achieved by humans and computers. We selected two high-level HVS metrics: structural similarity (SSIM) (Wang and Bovik, 2002) and visual information fidelity (VIF) (Sheikh and Bovik, 2006). The advantage of the SSIM image quality metric is that it quantifies the HVS and analyses the high-level structural properties of an image on a local level. SSIM determines the differences in linear correlation, luminance, and contrast between two images (Wang and Bovik, 2002).

The SSIM is computed using,

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{3}$$

where $\mu_x\, mu_y$ are the mean of x,y; σ_x^2, σ_y^2 are the variance of x, y; and σ_{xy} is the cross-variance of x and y. $C_1 = (k_1L)^2, C_2 = (k_2L)^2$, serve to stabilise the denominator as it approaches 0, where k_1, k_2 are generic constants and L is the dynamic range of the pixel values.

An improved version of the Information Fidelity Criterion metric, the VIF metric, was used on our distorted images (Sheikh and Bovik, 2006). The VIF reference model, known as the natural scene model, is represented as:

$$C = S\,.\,U = \{S_i\,.\,\vec{U_i} : i \in I\} \tag{4}$$

where S is a random field of positive scalars, U is a Gaussian vector random field with zero-mean and covariance C_U, and I is the set of spatial indices for the random field (Sheikh and Bovik, 2006).

The VIF distortion model is expressed as:

$$D = GC + V = \{g_i\vec{C_i} + \vec{V_i} : i \in I\} \tag{5}$$

where C is a random field from the reference image, D is the corresponding random field in the distorted image, G is a deterministic attenuation field, I is the set of spatial indices, and V is a stationary zero-mean Gaussian noise random field with covariance $C_V = \sigma_V^2 I$.

The VIF is expressed as:

$$VIF = \frac{\sum_j \in \mathit{subbands}\; I(\vec{C}^{N,j}; \vec{F}^{N,j} | s^{N,j})}{\sum_j \in \mathit{subbands}\; I(\vec{C}^{N,j}; \vec{E}^{N,j} | s^{N,j})} \tag{6}$$

where

$$I(\vec{C}^N; \vec{E}^N | s^N) = \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{M} log_2\left(1 + \frac{s_i^2\lambda_k}{\sigma_n^2}\right) \tag{7}$$

$$I(\vec{C}^N; \vec{F}^N | s^N) = \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{M} log_2\left(1 + \frac{g^2 s_i^2\lambda_k}{\sigma_v^2 + \sigma_n^2}\right) \tag{8}$$

$I(\vec{C}^N; \vec{E}^N | s^N)$ and $I(\vec{C}^N; \vec{F}^N | s^N)$ represent information that could be extracted from a subband of the reference C and distorted D images. σ_n^2 and σ_v^2 are parameters used to model Gaussian noise.

4 Experimental results

To evaluate the accuracy rates of the image-based face detection CAPTCHA, we solicited the assistance of 1,100 individuals. Over a period of several weeks, we collected data from 8,995 login attempts across 24 distortion types and distortion intensities. To evaluate the performance of automatic face detection, the Viola-Jones face detector (Viola and Jones, 2002) was used. Based on the analysis of the experimental results, the performance of humans and computers using automatic face detection algorithm was compared. The results show that the success rates for both humans and computers decrease as the intensity of distortion increases. When no distortion was applied to the images, the success rate of humans was 82.48% and the success rate of computer based face detector was only 16.67%. However, as the images were subjected to different types of distortion at different intensity levels, the success rate of humans varied from 73.45% to 79.03%, while the success rate of computer based face detector varied from 5.82% to 10.95%. Furthermore, an analysis of the human success rate and computer success rate for each distortion type is summarised in Table 2. The difference between success rates is also shown.

Table 2 Human success rates by distortion type and all intensity levels

Distortion type	*Success rate by humans*	*Success rate by computers*	*Difference in success rates*
No Distortion	82.48%	16.67%	65.81%
Periodic Noise	81.40%	14.29%	67.11%
Resolution modification	80.34%	15.08%	65.26%
Erosion	79.46%	16.67%	62.79%
Rotation	79.37%	1.59%	77.78%
Blurring	79.04%	16.67%	62.37%
Speckle noise	79.03%	13.10%	65.93%
Piecewise scaling	77.38%	7.14%	70.24%
Closing	75.89%	11.90%	63.99%
Height scaling	75.37%	0.00%	75.37%
Lightening	68.70%	9.52%	59.18%
Laplacian filtering	66.26%	0.00%	66.26%
Width scaling	64.07%	0.00%	64.07%

Integrating the image quality metrics with the experimental results, we observe that there is a general relationship between the image quality metric values and the success rates. As shown in Tables 3 and 4, distortions with higher SSIM values tended to have better human accuracy, while distortions with higher VIF values generally had better computer accuracy.

While there were cases where the computer algorithms were unable to correctly detect the human faces in a composite CAPTCHA, we know from the basic design of the CAPTCHA that at least one human face must be present. If a computer is unable to detect any face, it may attempt to solve the CAPTCHA by guessing where the human faces are located. If the guess is done completely at random, the likelihood of defeating the CAPTCHA is extremely remote. Assuming a CAPTCHA with five embedded images (one to four of which are human faces) and an average human face bounding box size of 48 × 51 pixels as in our test CAPTCHAs, the chance of a correct guess is:

$$\prod_{i=1}^{4} \frac{(48)(51)i}{(500)(300)} = 0.00017\%$$

If the computer is able to use edge detection or some other means to identify where the embedded images are located, its chance of correctly finding human faces increases somewhat but is still remote. Each embedded image is approximately 100 × 100 pixels, with the face bounding box comprising 24.48% of the space. Since at least one of the identified images is not a human face, we find the likelihood of a correct guess randomly is extremely low and is given by:

$$\prod_{i=1}^{4} \frac{(48)(51)i}{(100)(100)(i+1)} = 0.0718\%$$

Table 3 Comparison of human accuracy rates and the corresponding SSIM values

Distortion type	*Human rank*	*SSIM rank*	*Human success*	*SSIM value*
Periodic noise	1	7	81.40%	0.8161
Resolution modification	2	2	80.34%	0.9608
Erosion	3	4	79.46%	0.9303
Rotation	4	11	79.37%	0.7468
Blurring	5	3	79.04%	0.9403
Speckle noise	6	5	79.03%	0.9146
Piecewise scaling	7	8	77.38%	0.7609
Closing	8	1	75.89%	0.9677
Height scaling	9	9	75.37%	0.7540
Lightening	10	6	68.70%	0.8556
Laplacian filtering	11	12	66.26%	0.7293
Width scaling	12	10	64.07%	0.7526

Table 4 Comparison of computer accuracy rates and the corresponding VIF values

Distortion type	*Computer rank*	*VIF rank*	*Computer success*	*VIF value*
Erosion	1	5	16.67%	0.5799
Blurring	2	4	16.67%	0.6239
Resolution modification	3	1	15.08%	0.7276
Periodic noise	4	6	14.29%	0.5574
Speckle noise	5	3	13.10%	0.6342
Closing	6	2	11.90%	0.7071
Lightening	7	7	9.52%	0.4983
Laplacian filtering	8	8	0.00%	0.3705
Piecewise scaling	9	9	7.14%	0.3579
Rotation	10	10	1.59%	0.3433
Height scaling	11	11	0.00%	0.3376
Width scaling	12	12	0.00%	0.3327

5 Implementation considerations

Through our experimental work, we have found several key factors that must be taken into consideration when implementing the CAPTCHA to ensure optimal balance between high human success rates and low computer success rates. One of the most critical choices is in selecting the human face images to be embedded. Attention should be given to the pose, size, illumination, and facial expression of the subject as these factors significantly impact a human's ability to recognise the face once it is distorted and embedded in the final composite CAPTCHA. As an example, consider the two images in Figure 10, which were paired together in 49 different composite CAPTCHAs.

Figure 10 The original face image can have a significant impact on how well humans are able to identify it once embedded

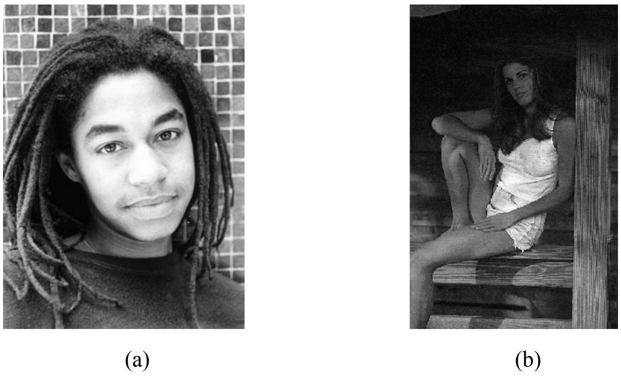


Figure 10(a) is a bust shot with good contrast. Humans correctly identified it as a human face 98.99% of the time. By comparison, Figure 10(b) is dark, has poor contrast, and the face comprises a relatively small portion of the full image. Humans only identified it in 61.82% of attempts. Effort should be given to removing items similar to Figure 10(b) that are unlikely to perform well from the source face image set, but there may be further value in using a standardised image set such as a driver's license photograph database. Standardised images with neutral pose, expression, and illumination will help to ensure the effect of applying distortions is predictable and also provides a consistent appearance for the human users attempting to recognise the human faces. The large variation in the CMU face database images we used is reflected in lowered human detection rates.

The two other key parameters to ensure optimal CAPTCHA effectiveness are the distortion type and distortion intensity applied to embedded images. We initially evaluated 24 different distortion types, considering their effects and differences from other distortions, before selecting 12 to be used in our CAPTCHA implementation. There were additional distortion types that may merit consideration, but were too similar to other types for our experimental purposes.

In our review and subsequent experiments, we made several observations regarding the performance of different CAPTCHA types and their value. We found geometric distortions such as scaling and rotating images functioned best, with noise-based distortions also of significant benefit. Mathematical morphologies such as closing and erosion have mixed value, performing well for human detection but also having above-average computer detection rates. Distortions impacting the contrast ratio are of limited use due to high computer success rates relative to human performance.

For each distortion type, one or more distortion intensities must be selected. In selecting the distortion intensities, it is beneficial to test several different intensities to identify which work best with a given image set. Depending on the underlying characteristics of the images, certain values may not be appropriate. For instance, heavily lightened face images that were already brightly illuminated may appear completely washed out. After testing several different parameter settings, it is frequently possible to divide the intensities into a few similar groups by their relative level of effect. In our testing, our selected distortion intensities coalesced towards high, medium, and low settings; with further evaluation, we removed some of the specific intensities because they provided little difference or resulted in unusable images.

While selecting standard distortion types and intensities to apply to all images is a viable means of generating CAPTCHAs, as demonstrated by our results here, other options exist. To achieve the best possible performance, the distortion type and intensity can be customised for each individual face image. The selection process may be automated by analysing image quality metrics. In our experiments, we used SSIM and VIF to model how well humans and computers, respectively, could identify each image. By creating a composite of these or other image quality metrics, one can easily compare a large number of distortion parameters to find those with optimal balance between human and computer detection accuracies. Similarly, different face images can be compared to remove images that are likely to perform poorly. In this way, when the composite CAPTCHA image is generated, we can know that it has the best possible likelihood of preventing computer attacks with minimal inconvenience to human users.

6 Conclusions

In this paper we demonstrate an implementation of a novel image-based face detection CAPTCHA to add an additional layer of security in web-based services. Existing CATCHAs are vulnerable to computer attacks. Text-based CAPTCHAs are vulnerable to advanced OCR technologies. Image-based CAPTCHAs use a small subset of images and are susceptible to random guessing. When the images or videos are selected from a large database, the users are presented with limited options making it susceptible to random guessing or machine learning techniques. Speech recognition software is used to exploit audio-based CAPTCHAs. Minimising the vulnerabilities to prevent computers from solving the CAPTCHAs also makes it challenging for humans, often requiring multiple attempts to successfully solve the CAPTCHA.

In this paper, we proposed an algorithm to generate an image-based CAPTCHA that uses the concept of face detection. The proposed algorithm embeds multiple human faces and non-human faces in a background image to create image CAPTCHAs. The background image contains randomly generated overlapping blocks of different shapes and contrast levels. The faces were selected from the CMU face database and were subjected to known distortions. By varying different parameters, the intensity of distortion is controlled to produce low, medium, and high levels of distortion. All these processing make it very challenging for face detection algorithm to accurately select all human faces embedded in the CAPTCHA image, while humans generally are able to identify the embedded human faces with relative ease. The design objective is to generate CAPTCHA images such that the computers attack rates are minimised while human accuracy to solve the same CAPTCHA is considerably increased. The use of image quality metrics to study the characteristics of images and design optimal images is briefly presented in the paper. An extensive experimental study demonstrates these important features of the image-based face detection CAPTCHA. In addition, key factors that need to be considered in designing image-based face CAPTCHAs are described in detail. The proliferation of new generation mobile devices increasingly uses Internet-based applications and it is imperative they be made secure and resilient to attacks. These devices generally do not have a convenient keyboard and therefore the proposed image-based face detection CAPTCHA is ideally suited for clicking to solve the CAPTCHA rather than typing. Since there is no text involved, this CAPTCHA is language-independent and can be widely used by a large audience.

References

Baird, H.S. and Popat, K. (2002) 'Human interactive proofs and document image analysis', *Document Analysis Systems V*, pp.531–537, Springer, Berlin.

Bursztein, E. and Bethard, S. (2009) 'Decaptcha: breaking 75% of eBay audio CAPTCHAs', *3rd USENIX Workshop on Offensive Technologies*, Montreal, August.

Carnegie Mellon University (2002) 'Carnegie mellon university image data base: frontal face images', available at http://bit.ly/cmuface (accessed on 11 April 2010).

Carnegie Mellon University (2004) 'ESP-PIX', available at http://server251.theory.cs.cmu.edu/cgi-bin/esp-pix/esp-pix (accessed on 9 September 2009).

Chellapilla, K., Larson, K., Simard, P.Y. and Czerwinski, M. (2005) 'Building segmentation based human-friendly human interaction proofs (HIPs)', *Human Interactive Proofs*, pp.1–26, Springer, Berlin.

Chew, M. and Baird, H.S. (2003) 'BaffleText: a human interactive proof', *Document Recognition & Retrieval X Conference*, Santa Clara, California, January.

Coates, A.L., Baird, H.S. and Fateman, R.J. (2001) 'Pessimal print: a reverse turing test', *6th International Conference on Document Analysis and Recognition*, Seattle, Washington, September.

Elson, J., Douceur, J., Howell, J. and Saul, J. (2007) 'Asirra: a CAPTCHA that exploits interest-aligned manual image categorization', *14th ACM Conference on Computer and Communications Security*, Alexandria, Virginia, October.

Flickr (2010) *Flickr*, available at http://www.flickr.com/ (accessed on 3 March 2010).

Golle, P. (2008) 'Machine learning attacks against the Asirra CAPTCHA', New York, NY.

Kluever, K.A. (2008) 'Evaluating the usability and security of a video CAPTCHA', Rochester Institute of Technology.

Kluever, K.A. and Zanibbi, R. (2009) 'Balancing usability and security in a video CAPTCHA', *5th Symposium on Usable Privacy and Security*, Mountain View, California, July.

Microsoft (2010) 'How secure is Asirra? – Microsoft research', available at http://research.microsoft.com/en-us/projects/asirra/security.aspx (accessed on 24 March 2010).

Misra, D. and Gaj, K. (2006) 'Face Recognition CAPTCHAs', *International Conference on Internet and Web Applications and Services/Advanced International Conference*.

Mori, G. and Malik, J. (2003) 'Recognizing objects in adversarial clutter: breaking a visual CAPTCHA', *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Madison, Wisconsin, June.

Moy, G., Jones, N., Harkless, C. and Potter, R. (2004) 'Distortion estimation techniques in solving visual CAPTCHAs', *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, June.

reCAPTCHA (2010) 'What is reCAPTCHA?', available at http://recaptcha.net/learnmore.html (accessed on 29 June 2010).

Rice, S.V., Nagy, G. and Nartker, T.A. (1999) *OCR: An Illustrated Guide to the Frontier*, Kluwer Academic Publishers, Dordrecht, Netherlands.

Rusu, A. and Govindaraju, V. (2004) 'Handwritten CAPTCHA: using the difference in the abilities of humans and machines in reading handwritten words', paper presented at the *Ninth International Workshop on Frontiers in Handwriting Recognition*.

Santamarta, R. (2008) 'Breaking Gmail's audio Captcha', available at http://blog.wintercore.com/?m=200803 (accessed on 31 August 2009).

Sheikh, H. and Bovik, A. (2006) 'Image information and visual quality', *IEEE Transactions on Image Processing*, Vol. 15, No. 2, pp.430–444.

Simard, P.Y., Szeliski, R., Benaloh, J., Couvreur, J. and Calinov, I. (2003) 'Using character recognition and segmentation to tell computer from humans', *7th International Conference on Document Analysis and Recognition*, Edinburgh, Scotland, August.

Tam, J., Hyde, S., Simsa, J. and von Ahn, L. (2008) 'Breaking audio CAPTCHAs', *22nd Annual Conference on Neural Information Processing Systems*, Vancouver, British Columbia, December.

Viola, P. and Jones, M. (2002) 'Robust real-time object detection', *International Journal of Computer Vision*, Vol. 57, No. 2, pp.137–154.

von Ahn, L. (2005) 'Human computation', PhD thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania.

von Ahn, L., Blum, M. and Langford, J. (2004) 'Telling humans and computers apart automatically', *Communications of the ACM*, Vol. 47, No. 2, pp.56–60.

von Ahn, L., Maurer, B., McMillen, C., Abraham, D. and Blum, M. (2008) 'reCAPTCHA: human-based character recognition via web security measures', *Science*, Vol. 321, No. 5895, pp.1465–1468.

Wang, Z. and Bovik, A. (2002) 'A universal image quality index', *IEEE Signal Processing Letters*, Vol. 9, No. 3, pp.81–84.

Yan, J. and Salah, A. (2008) 'A low-cost attack on a Microsoft CAPTCHA', *15th ACM Conference on Computer and Communications Security*, Alexandria, Virginia, October.