Avatar CAPTCHA: Telling Computers and Humans Apart via Face Classification

Darryl D'Souza
Computer Engineering and Computer Science
University of Louisville
Louisville, Kentucky 40292
Email: darryl.dsouza@louisville.edu

Phani C. Polina
Computer Engineering and Computer Science
University of Louisville
Louisville, Kentucky 40292
Email: p0poli01@louisville.edu

Roman V. Yampolskiy
Computer Engineering and Computer Science
University of Louisville
Louisville, Kentucky 40292
Email: roman.yampolskiy@louisville.edu

***Abstract*—This paper introduces Avatar CAPTCHA, an image based approach to distinguish human users from computer programs (bots). The proposed CAPTCHA asks users to identify avatar faces from a set of 12 grayscale images comprised of a mix of human and avatar faces. Experimental results indicate that it can be solved 62% of the time by human users with an average success time of 24 seconds and a positive user rating of 90%. It is designed to be secure against computer programs (bots). Using brute force attack the success rate for a bot to solve it is 1/4096.**

***Index Terms*—CAPTCHA, avatars, security, ASIRRA, bots.**

I. Introduction

Human Interactive Proofs (HIPs) are challenges meant to be easily solvable by humans while being infeasible for computers. HIPs lately have been increasingly used to protect web services against automated scripting attacks [1]. Examples involve online registrations, ticket reservations, online polling, etc. HIPs help discourage scripting attacks by raising the computational or developmental costs and are easy enough for the human user to solve [1]. HIPs are used to protect computational resources and disk storage space from computer-generated programs (bots). Work on distinguishing computers from humans traces back to the original Turing Test (TT) [2]. TT asks a human judge to distinguish between another human and a machine by interrogating both, via a text chat interface. Similarly, CAPTCHAs (Completely Automated Public Turing Tests to Tell Computer and Humans Apart) aim to distinguish between computer programs (bots) and humans [3], [4] as shown in Fig. 1.

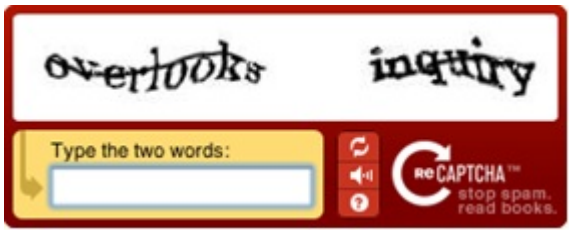

Fig. 1. A text CAPTCHA by CMU [3]

Companies such as Google, Yahoo, Microsoft, etc. are using them to protect their online services by requiring a user to solve the reverse Turing-test challenge before permitting them to sign up for an account. In the absence of such challenges, it would be possible for computer programs to misuse companies online services by submitting thousands of requests by each bot which could even cause denial of service to human users. Text-based CAPTCHAs increase the distortion of the letters making it more difficult for the bots to recognize them correctly. However, the underlying problem with this approach is that by increasing the distortion we make it difficult for the user to identify all the letters correctly. The primary question here is what characteristics should an effective CAPTCHA have? It must be easy for humans to understand and solve but extremely difficult for a computer program to break [5]. However, powerful, intelligent and advanced computer programs can break a sufficiently large number of CAPTCHAs. A seemingly viable solution is to use the concept of Image Recognition.

Its role will be to replace text identification CAPTCHAs with images. Humans can easily identify images whereas they are very difficult for computer programs to decipher. Generating a prototype is easy. But unless the images picked are from a large database and generated dynamically, one faces the risk of a brute-force attack on the system. Image CAPTCHAs are not completely immune to bot attacks, but raising the computational time and costs can be very effective. The idea of using image recognition in CAPTCHAs is not new. There are existing prototypes showcased in CMUs CAPTCHA website [3] as well as in Microsofts ASIRRA project [6]. Such challenges widen the gap between human and non-human success rates as image recognition is a much harder problem for bots to break than text recognition. Moreover, it is more convenient and less complex for a human user. In our proposed Avatar CAPTCHA the user is presented with 12 images consisting of a random number of avatar faces and the rest consisting of human faces. The user has to select all the avatar images present in the 12 images. If the user selects only and all the avatar images the user is considered human. If incorrect, it is viewed as an attempt by a bot (non-human user). A bot has a 1 in 2 probability to select an image as an avatar. With 12 images the success rate for a computer bot to pose as a human is 1 in 4096. If we have to make the brute

978-1-4673-0818-2/12/$31.00 ©2012 IEEE

force attack difficult we can increase the number of images to 18. The success rate for a bot would then be 1 in 262144. The increase in complexity for the bot will not increase the complexity for the human solver.

II. Related Work

Several variations of the CAPTCHA design have appeared since it was first introduced. Text CAPTCHAs [7], Image CAPTCHAs [6], [8], Motion CAPTCHA [9] and 3D animation CAPTCHAs [10], [11]. Most of them are text-based. The computer generates a sequence of letters or digits and distorts them with a certain amount of noise before rendering them on the screen (Fig. 2).

(a) Simple text CAPTCHA easily defeated by OCRs [12]

(b) Text CAPTCHAs distorted with noise

Fig. 2. Text CAPTCHA variants

Such CAPTCHAs are quite robust to random guess attacks. However, it has been proven that Optical Character Recognition (OCR) can achieve human-like accuracy in figuring out distorted letters as long as they can be reliably segmented into their constituent letters shown in Fig. 2a [12]. Distorted text in CAPTCHAs makes it hard for humans to read them. This has led to the introduction of images as CAPTCHAs. Chew and Tygar [5] were among the first to use labeled pictures to generate a CAPTCHA.

However, their database was small enough to permit a brute-force attack by manually labeling all images. CAPTCHA principles have been recently applied in numerous areas such as bot detection in online games such as poker [13], [14], graphical CAPTCHAs have been embedded in playing cards in online poker [15], as well as image distortion and random initial placement of chess pieces in an online chessboard game, Fischer Random Chess [16]. Distorted human faces are used to design a universal HIP system known as ARTiFACIAL [17]. There has been notable work in building face recognition CAPTCHAs [18].

Microsoft's ASIRRA [6] addresses the image generation and subsequently the database populating scheme in a novel way by working together with Petfinder.com, a popular pet adoption website for homeless pets. It generates challenges by displaying 12 images of cats and dogs from a hugely populated database of three million pictures manually classified as cats and dogs. Nearly 10,000 more are added every day through the United States and Canada. The size and accuracy of this database is the key to ASIRRAs security. Users have to select all the cat images from the 12 displayed images correctly in order to be classified as human. In exchange for access to Petfinders database, ASIRRA provides an Adopt Me link beneath each picture to help promote Petfinders primary mission to expose the pets to the public in the hopes of having them adopted. To maximize the adoption probability ASIRRA will employ IP geolocation to determine the users approximate region and preferentially display pets that are nearby. Thus, ASIRRA has several positive features. It is quick and solved by humans with a high accuracy. Computers cannot solve it easily and it requires no prior or specialized knowledge to solve it. This makes it less frustrating for humans. However, ASIRRA has several disadvantages too. Its security is lost if its database is compromised, it requires more screen space than a regular text CAPTCHA and it is inaccessible to the visually impaired. Any click of the forms Submit button causes ASIRRA to score the challenge, even though the user had a different intent in mind [6].

Another interesting CAPTCHA worth mentioning is Googles CAPTCHA based on image orientation [8] which uses images as an alternative to text. Here users are presented a set of images that have to be rotated to align them in an upright position. Fig. 3 illustrates this.

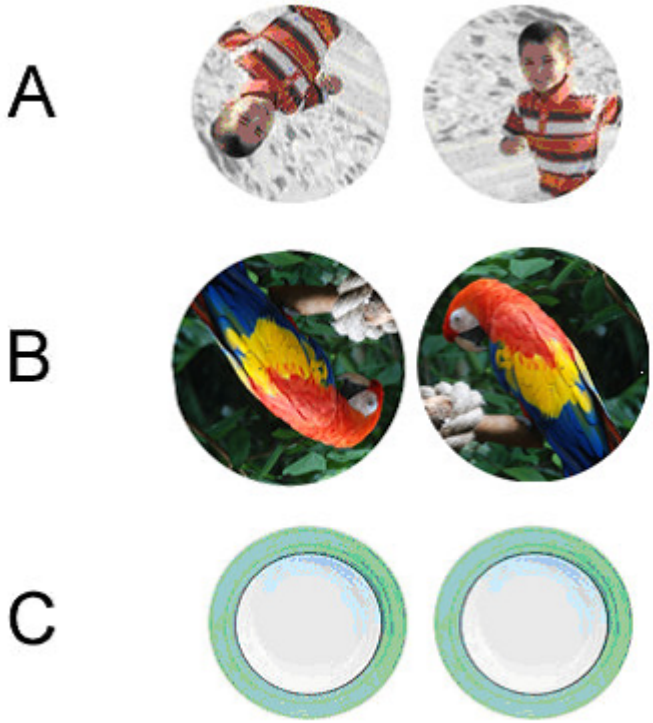

Fig. 3. A snapshot of Whats up CAPTCHA? [8]

Setting an upright orientation is easy for people whereas it is difficult for bots. They discard images that are easily identifiable by bots as well as difficult for humans to orient.

III. Architecture

We are motivated by Microsofts ASIRRA CAPTCHA [6] and Luis von Ahns art of harnessing human capabilities to address problems that computers cannot solve [19]. The idea is to build an image (graphic) CAPTCHA using biological (human) and non-biological (virtual world avatar) faces. From

our survey feedback we believe that faces are easily identifiable and distinguished by the human eye yielding better and accurate results.

Our CAPTCHA comprises of 2 rows with 6 images each. These images are randomly picked from datasets comprised of human and avatar faces. Each image has a checkbox associated with it for the user to make his choice. These images are converted to grayscale before being rendered onto the screen. This is to prevent computer programs from breaking the CAPTCHA by taking advantage of the varying color spectrum difference between human and avatar images. The goal of the users here is to select all the avatar faces. Their choices are validated for accuracy thus preventing unauthorized access to malicious computer programs.

This architecture is based on the popular client server architecture as shown in Fig. 4.

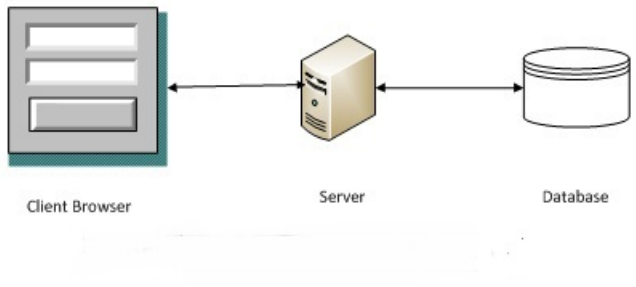

Fig. 4. Client-Server Architecture

The client machine (browser) requests the server for an authentication service. The server randomly picks 12 images of humans and avatars. Of these, 5 or 6 are avatar images. Later it transfers them to the client. The users are asked to select all the avatar faces. Their choices are subsequently validated by the server. The users are notified of their decision and a decision is made classifying them as a genuine human user or a malicious computer program (bot). In the future we aim to build this as a web service to help clients directly integrate it into their respective websites.

IV. Experiment

A. Datasets

We considered images with upright frontal faces, complex backgrounds and varying illuminations. Converting them to grayscale helped avoid color-based image recognition algorithms to detect the unusually bright and uncommon colored avatar faces and consequently breaking the CAPTCHA.
In our experiments we used the following datasets:

1) Humans: :
The Nottingham_scans dataset [20] was used. It contains grayscale facial images of 50 males and 50 females. These are mainly frontal and some profile views with some differences in lighting and expression variations. Resolution variance was from 358 x 463 to 468 x 536. For efficiency, thumbnail-sized images with resolutions of 100 x 123 were used.

2) Avatars: :
100 samples of grayscale, frontal face avatar images [21] from the popular online virtual world Entropia Universe [22] were used. Thumbnail-sized images with a resolution of 100 x 135 were used for efficiency.

B. System Description

The 12 images of humans and avatars are displayed on the webpage with corresponding checkboxes for each. A snapshot is shown in Fig. 5. A survey requesting users to attempt and solve this CAPTCHA was posted at the following URL http://darryl.cecs.cecsresearch.org/avatarcaptcha/.

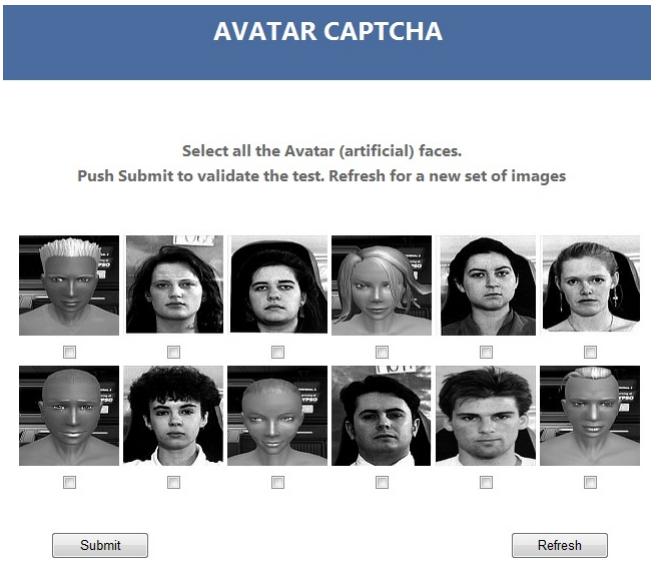

Fig. 5. Snapshot of the Avatar CAPTCHA

This screen represents the home screen of the Avatar CAPTCHA. Here the user is requested to select all the avatar faces from a mixed set of human and avatar faces. Their scores are validated and they are classified as humans as shown in Fig. 6, or bots as shown in Fig. 7, and led to a survey feedback page. Here their feedback responses are obtained, which are analyzed and discussed in the Results section, to help improve our CAPTCHA. The outcomes are stored in two tables named Test_Results and Survey_Feedback within a database.
The feedback survey form gets the following responses from the users:

Gender, age, education background, their experiences in solving text and image CAPTCHAs before, rating the fun factor in solving this Avatar CAPTCHA, justifying the choice of faces here, how challenging is it, their preferences in solving text or image CAPTCHAs, usage of this CAPTCHA on their websites, rate it and finally their comments or feedback on it. Once the user hits the Submit key the outcomes are stored in the Test_Results table. We capture the user's IP address, success or failure outcome, number of avatars not selected, number of humans selected and time taken to give the test. Now if the users fill out the feedback form they are stored in the Survey_Feedback table.

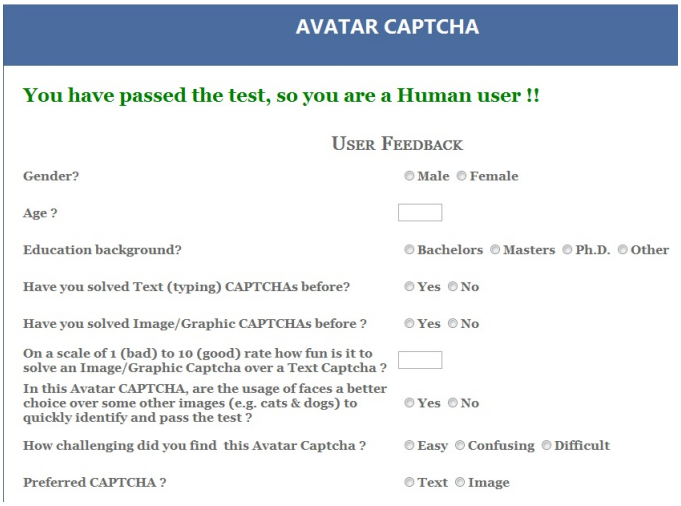

Fig. 6. Outcome of a human user solving the test

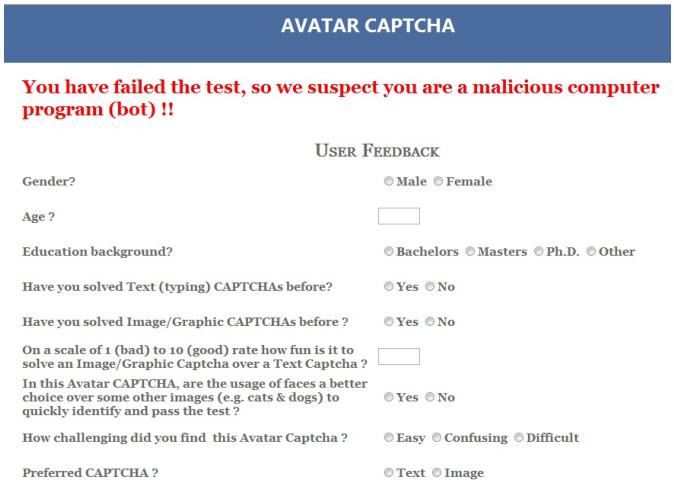

Fig. 7. Outcome of a bot attempting to solve the test

C. Security Risks

If a bot tries to break the CAPTCHA using a brute-force approach it will have a success probability of 0.5. So guessing 12 images will yield a success probability of 1 in 4096, which is considerably low. Users are unable to access the datasets. In the future we aim to generate datasets dynamically. These datasets, obtained in real-time, will comprise of human and avatar images from popular online websites such as Flickr and ActiveWorlds. These dynamic datasets will help us combat manual brute-force attacks on the database and lend a real-time dimension to it. Moreover, so far to our knowledge no work has been done to differentiate human and avatar faces.

V. RESULTS

The results evaluated so far are records from 163 user test evaluations stored in the Test_Results table and 50 user feedback responses stored in the Survey_Feedback table within the database.
Table 1 and Tables 2 & 3 depict an overview of their data.

TABLE I
OVERVIEW OF THE USER_RESULT DATA

Outcome			**Average Submit Time (seconds)**	**Average Success Time (seconds)**
Success	***Failure***			
101/163 = 61.96%	62/163 = 38.04%		3466/163 = 21.2638	2410/101 = 23.8614
	Avatar Missed (Avg)	***Humans Checked (Avg)***		
	213/62 ≈ 3	124/62 = 2		

We observe that 62% of the users solved the CAPTCHA successfully. The failure rate of 38% also includes those users testing the CAPTCHA with attempts to randomly select a few images and submit the challenge.This is chiefly done to validate the system. Amongst the failures on an average, 3 avatars missed out on being selected and 2 human faces were accidentally selected. The submit time is the time when the user hits the Submit button to validate the test. The success time is the time reported when the CAPTCHA is solved. Average submit and success times of 21 and 24 seconds were reported respectively.

Results from the Survey_Feedback table are split into two tables. Table 2 contains an overview of the results regarding information about the user, text and image CAPTCHAs in general. Table 3 depicts an overview of the Avatar CAPTCHA results.

TABLE II
OVERVIEW OF PART 1 OF USER_FEEDBACK DATA

Gender		**Age**		**Text CAPTCHA Knowledge**	
Male	***Female***	***Min***	***Max***	***Yes***	***No***
30/50 = 60%	20/50 = 40%	18	61	42/50 = 84%	8/50 = 16%
Education				**Image CAPTCHA Knowledge**	
Bachelor's	***Master's***	***Ph.D.***		***Yes***	***No***
14/50 = 28%	16/50 = 32%	20/50 = 40%		22/50 = 44%	28/50 = 56%

From Table 2 we observe that 60% of the test takers (users) were male and 40% female. Their ages ranged between 18 and 61 years. 40% of the users held a Ph.D. degree, 32% had a Masters degree and 28% had a Bachelors degree. 84% of the users had some knowledge and experience in facing and solving text CAPTCHAs before. 56% of the users had never seen or solved an image CAPTCHA before. This signifies the need for this approach to be put forward to the users to test and judge it.

From Table 3 we observe that 88% of the users felt that faces played a pivotal role in quick identification and easily solving the CAPTCHA. 94% of the users preferred solving an image CAPTCHA over its textual counterpart. 90% of the users voted to use this CAPTCHA on their personal websites. This CAPTCHA was rated excellent by 52% and good by

TABLE III
OVERVIEW OF PART 2 OF USER_FEEDBACK DATA

Role of Faces		**Preferred CAPTCHA**		**Website usage**		
Helping	***Unhelping***	***Image***	***Text***	***Yes***	***No***	
44/50	6/50	47/50	3/50	45/50	5/50	
88%	12%	94%	6%	90%	10%	
Rating				**Solvability**		
Excellent	***Good***	***Average***	***Poor***	***Easy***	***Confu -sing***	***Hard***
26/50	19/50	3/50	2/50	46/50	3/50	1/50
52%	38%	6%	4%	92%	6%	2%

38% of the users. Its solvability rate was judged as 92% excellent. Specific user comments are summarized and stated below Table 3.

"*Image CAPTCHAs are very easy.*"
"*A smaller, configurable interface would be nice as this takes a large amount of screen space.*"
"*I would prefer colored images over black and white (grayscale).*"
"*Not everyone understands what an avatar is.*"
"*It is like a virtual keyboard where one uses a mouse.*"
"*Easy to solve.*"

The users also rated the fun factor of solving an image/graphic CAPTCHA over the traditional text CAPTCHAs on a scale of 1 (bad) to 10 (best). The outcomes are shown in the labeled histogram below in Fig. 8.

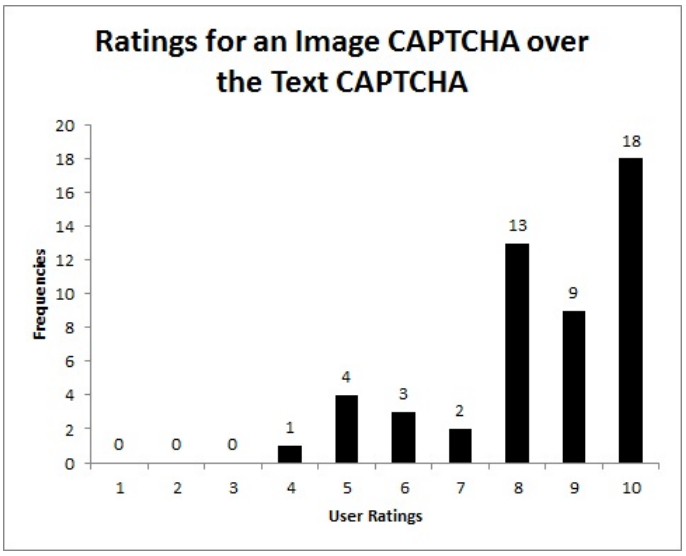

Fig. 8. Histogram for user ratings of an image CAPTCHA over the text CAPTCHA with labeled frequencies

VI. FUTURE WORK

We aim to build this CAPTCHA as a web service to be integrated within websites as a security measure over the traditional text CAPTCHAs. We appreciate the user comments on our CAPTCHA and work on user-friendliness. Building a smaller, configurable interface definitely seems a valuable opinion. Configuring this CAPTCHA with real-time images is a priority. We plan to use human and avatar images from popular online websites such as Flickr and ActiveWorlds dynamically and display them on our CAPTCHA.

VII. CONCLUSION

This proposed Avatar CAPTCHA is a novel approach based on human computation relying on identification of avatar faces. Of the 163 user tests recorded, 62% of the users solved the CAPTCHA. This excludes user attempts to validate the system. The average success time was 24 seconds. Of the 50 user feedback responses recorded 56% of the users had absolutely no knowledge about solving image CAPTCHAs. 94% of the users preferred image CAPTCHAs over text CAPTCHAs. 88% of the users stated that facial-image CAPTCHAs are easier to solve. 92% of the users rated it as easily solvable and 90% of the users had positive ratings for it. 90% of the users voted to use this on their websites. These statistics prove it to be a convenient tool to filter out unauthorized access by bots. Designing CAPTCHAs indeed proves to be a challenge in building foolproof systems. A good approach is to make it fun and convenient for users to solve them.

ACKNOWLEDGMENT

The authors would like to thank the users for their valuable feedback and time towards solving the CAPTCHA. They would also like to thank Taylor Smith for his help in setting up the necessary tools required to host this CAPTCHA for the user survey on the department web server.

REFERENCES

[1] K. Chellapilla, K. Larson, P. Simard, and M. Czerwinski, "Designing human friendly human interaction proof (HIPs)," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, Portland, Oregon, USA, 2005.
[2] A. M. Turing, *Parsing the Turing Test*. Springer Netherlands, 2009, ch. Computing Machinery and Intelligence, pp. 23–65.
[3] L. v. Ahn, M. Blum, N. Hopper, and J. Langford. The CAPTCHA project. [Online]. Available: http://www.captcha.net/
[4] L. V. Ahn, M. Blum, and J. Langford, "Telling humans and computers apart automatically," *Communications of the ACM*, vol. 47, no. 2, pp. 56–60, 2004.
[5] M.Chew and J.Tygar, "Image recognition CAPTCHAs," in *Information Security Conference*, Palo Alto, California, 2004.
[6] J. Elson, J. Douceur, J. Howell, and J. Saul, "Asirra: a CAPTCHA that exploits interest-aligned manual image categorization," in *14th ACM conference on Computer and communications security*, Alexandria, Virginia, USA, 2007.
[7] L. V. Ahn, M. Blum, N. Hopper, and J. Langford, "CAPTCHA: Using hard ai problems for security," in *International Conference on the Theory and Applications of Cryptographic Techniques*, Warsaw, Poland, 2003.
[8] R. Gossweiler, M. Kamvar, and S. Baluja, "What's up CAPTCHA?: a CAPTCHA based on image orientation," in *18th International Conference on World Wide Web*, Madrid, Spain, 2009.
[9] M. Shirali-Shahreza and S. Shirali-Shahreza, "Motion CAPTCHA," in *Human System Interaction*, Krakow, Poland, 2008.
[10] C. Jing-Song, M. Jing-Ting, Z. Wu-Zhou, W. Xia, and Z. Da, "A CAPTCHA implementation based on moving objects recognition problem," in *International Conference on E-Business and E-Government (ICEE)*, Shanghai, China, 2010.
[11] C. Jing-Song, M. Jing-Ting, W. Xia, Z. Da, and Z. Wu-Zhou, "A CAPTCHA implementation based on 3d animation," in *International Conference on Multimedia Information Networking and Security, MINES '09*, Hubei, China, 2009.

[12] P. Y. Simard, D. Steinkraus, and J. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Seventh International Conference on Document Analysis and Recognition*, Edinburgh, Scotland, 2003.
[13] R. V. Yampolskiy, "Embedded CAPTCHA for online poker," in *20th Annual CSE Graduate Conference (Grad-Conf2007)*, Buffalo, NY, 2007.
[14] R. Yampolskiy and V.Govindaraju, "Embedded noninteractive continuous bot detection," *Computers in Entertainment (CIE) - Theoretical and Practical Computer Applications in Entertainment*, vol. 5, no. 4, pp. 1–11, 1st October 2007.
[15] R. V. Yampolskiy, "Graphical CAPTCHA embedded in cards," in *Western New York Image Processing Workshop (WNYIPW) - IEEE Signal Processing Society*, Rochester, NY, 2007.
[16] R. C. McDaniel and R. V. Yampolskiy, "Embedded non-interactive CAPTCHA for fischer random chess," in *16th International Conference on Computer Games (CGAMES)*, Louisville, Kentucky, 2011.
[17] Y. Rui and Z. Liu, "ARTiFACIAL: Automated reverse turing test using FACIAL features," *Multimedia Systems*, vol. 9, no. 6, pp. 493–502, 2004.
[18] D. Misra and K. Gaj, "Face recognition CAPTCHAs," in *International Conference on Internet and Web Applications and Services (AICT-ICIW '06)*, Guadeloupe, French Caribbean, 2006.
[19] L. von Ahn, "Human computation," in *46th ACM/IEEE Design Automation Conference*, San Francisco, CA, 2009.
[20] P. I. C. at Stirling (PICS). Nottingham_scans. [Online]. Available: http://pics.psych.stir.ac.uk/2D_face_sets.htm
[21] J. N. Oursler, M. Price, and R. V. Yampolskiy, "Parameterized generation of avatar face dataset," in *14th International Conference on Computer Games: AI, Animation, Mobile, Interactive Multimedia, Educational & Serious Games*, Louisville, Kentucky, 2009.
[22] Entropia universe. [Online]. Available: www.entropiauniverse.com

Darryl DSouza received the B.Eng. degree in Computer Engineering from University of Mumbai, India, in 2006 and the M.S. degree in Computer Science from University of Louisville, Louisville, USA, in 2009, respectively. Currently,he is working towards the Ph.D. degree at the Department of Computer Engineering and Computer Science, University of Louisville, Louisville, USA. His current research interests include applying biometric principles in virtual worlds, face detection of avatars, digital forensics.

Phani Polina received the masters degree in computer science in 2006 from the Western Kentucky University (WKU) at Bowling Green, Kentucky, USA.He is currently doing his Ph.D at University of Louisville, USA. His current research interests include cloud, mobile computing and security.

Roman V. Yampolskiy is the director of the Cybersecurity Research Lab at the CECS department, University of Louisville. Dr. Yampolskiy holds a PhD degree from the Department of Computer Science and Engineering at the University at Buffalo. There he was a recipient of a four year NSF fellowship. After completing his PhD dissertation Dr. Yampolskiy held a position of an Affiliate Academic at the Center for Advanced Spatial Analysis, University of London, College of London. He had previously conducted research at the Laboratory for Applied Computing at the Rochester Institute of Technology and at the Center for Unified Biometrics and Sensors at the University at Buffalo. Dr. Yampolskiys main area of expertise is biometrics. He has developed new algorithms for action-based person authentication. Dr. Yampolskiy is an author of over 60 publications including multiple journal articles and books.