City Research Online

City, University of London Institutional Repository

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

A CAPTCHA Model Based on Visual Psychophysics: Using the Brain to Distinguish Between Human Users and Automated Computer Bots

Seyed Mohammad Reza Saadat Beheshti[1], Panos Liatsis[2] and Muttukrishnan Rajarajan[1]

[1]*Department of Electrical and Electronic Engineering, City University London, EC1V 0HB, UK*

[2]*Department of Electrical Engineering, The Petroleum Institute, PO Box 2533, Abu Dhabi, UAE*

A B S T R A C T

Demand for the use of online services such as free emails, social networks, and online polling is increasing at an exponential rate. Due to this, online service providers and retailers feel pressurised to satisfy the multitude of end-user expectations. Meanwhile, automated computer robots (known as 'bots') are targeting online retailers and service providers by acting as human users and providing false information in order to abuse their service provisioning. CAPTCHA is a set of challenge/response protocol, which was introduced to protect online retailers and service providers from misuse and automated computer attacks. Text-based CAPTCHAs are the most popular form, and are used by most online service providers to differentiate between the human users and bots. However, the vast majority of text-based CAPTCHAs have been broken using the Optical Character Recognition (OCR) techniques and thus, reinforces the need for developing a secure and robust CAPTCHA model. Security and usability are the two fundamental issues that pose a trade-off in the design of a CAPTCHA; a hard CAPTCHA model could also be difficult for human users to resolve, which affects its usability, and vice versa. The model developed in this study uses the unsurpassed abilities of the Human Visual System (HVS) to superimpose and integrate complex information presented in individual frames, using the mechanism of *trans-saccadic memory*. In this context, the model integrates in its design the concept of *persistence of vision*, which enables humans to see the world in a continuous fashion. Preliminary results from the proposed model based on this technique are encouraging. As a result of this research, we have achieved 65% improvement in terms of the character recognition success rate for human users compared to the current computer recognition programs for multi-frame scenarios. Its ability to remain unbroken by current OCR programs for single-frame scenarios is over 98%.

Keywords – CAPTCHA; Persistence of vision; Trans-Saccadic memory; Visual integration; Authentication; Security

1. Introduction

Nowadays, most of us are familiar with online retailer webpages asking their users to retype a selection of distorted characters from an image before allowing them access to their online services. At times, this process may be frustrating and can require a considerable amount of time and effort to visually inspect and type out these blurry characters and numbers. These distorted images are an example of CAPTCHAs (Completely Automated Public Turing test to tell Computers and Humans Apart). CAPTCHAs (Luis von Ahn, 2004) were developed to protect websites and online service providers from possible cyber-attacks originating from automated computer programs (bots), and are the principal shield in protecting websites from being abused by bots and spams. The term 'CAPTCHA' was first introduced by Luis von Ahn in the year 2000 at Carnegie Mellon University (Luis von Ahn, 2010). This process has also been referred to as the 'Reverse Turing Test' (Chellapilla, 2005). An example of one of the famous current CAPTCHA models, known as ReCAPTCHA is shown in **Fig. 1**.

As demand for online services rapidly grows, it is essential to have a mechanism that cannot only recognise the users as genuine, but at the same time can also ensure that users are real humans rather than automated computer programs (or 'bots'). There is a direct correlation between growing numbers of online users, and a rapid rise in the number of cyber-attacks in recent years. Automated computer attacks abuse online service providers by supplying false information to systems and acting as human users (Anon., 2014). Human Interactive Proofs (HIPs) or CAPTCHAs are a set of challenge/response protocols designed in the form of a challenge or a test that can be presented to the users in order to distinguish human users from computer automated programs (Chew, 2003). HIPs are designed to be easy for human users to solve, and should be very challenging (and ideally impossible) for automated computer programs to break.
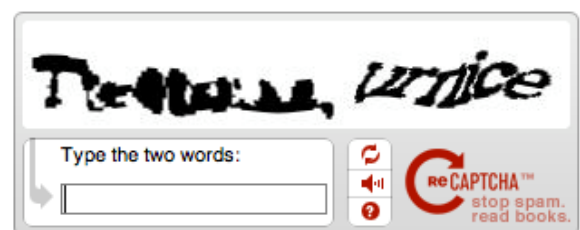

Fig. 1 - Example of one of the most popular current CAPTCHAs, known as ReCAPTCHA (Marc, 2011)

CAPTCHAs are used for a variety of online applications such as free email accounts, e-commerce, online polling, chatrooms, and many other interactive online services (Carnegie Mellon University, 2000-2010). According to von Ahn (Ahn, et al., 2003) the idea behind a CAPTCHA is to use the sophisticated abilities of the human perceptual system in order to resolve a problem, which cannot be addressed by computer programs. Over the past decade, various CAPTCHA models were introduced and used widely by major online service providers such as Yahoo!, Google and Microsoft as well as social networks such as Facebook, in order to provide better security against automated computer attacks.

However, there has been extensive research on the security of CAPTCHAs as most CAPTCHA models have been broken using sophisticated recognition techniques (Jeff Yan, 2009). In this paper, a novel CAPTCHA model is introduced called Visual Integration CAPTCHA (VICAP), which makes use of the abilities of the human visual perception system to superimpose and integrate fleeting image frames of noisy partial information in order to create a final object, which can then be recognised by the brain. The proposed approach involves generating a sequence of image frames, which are obtained by sampling the binarized version of the original image, consisting of a string of characters, and adding in background noise to increase the security of the model. Next, the sequence of these images is played back at an appropriate frame rate so that the brain perceives it as a continuous animation, and thus, recognising the original object.

When consecutive frames are displayed at very high speeds, visual information can be integrated into the Visual Short Term Memory (VSTM), which allows humans to perceive and complete the image of the original object. Each individual frame contains only a part of the original object pixels. Therefore, by analysing and processing a single frame, no useful information regarding the original object can be retrieved, thus rendering it unsuitable for OCR algorithms. The robustness of the proposed CAPTCHA model was tested using the state of the art CAPTCHA Breaker program. The final results demonstrate that the proposed CAPTCHA model is robust against various types of cyber security attacks.

Our main contributions to this research can be listed as follows:

- We have used the concept of visual psychophysics in order to design a novel CAPTCHA model, which would only be understandable for human users and not current computer programs.

- Persistence of vision has been applied on the new CAPTCHA model in order to superimpose and integrate all the CAPTCHA images to form an object in the brain.

- The proposed CAPTCHA model has been tested and evaluated on both human users and computer recognition programs in order to formulate the best possible CAPTCHA design with the optimal recognition success rate in terms of both computer security and usability.

- The *Recognition Improvement Level* (RIL) has been achieved for human users with having a 65% increase in recognition success rates compared to current computer recognition programs for multi-frame scenarios.

- The new CAPTCHA's ability to be unrecognisable to current OCR programs for single frame scenario has increased by 98%.

This paper is structured as follows: Section 1 will give a brief introduction regarding computer security and the importance of CAPTCHA in our daily online activities. Key background information on CAPTCHA, together with discussion of types of CAPTCHA categories, will then be discussed in Section 2. Sections 3 and 4 will present our proposed CAPTCHA model and the relevant methodologies behind it. In Section 5, some of the evaluation and experimental results will be provided and discussed. Finally, Section 6 will draw together the conclusions from our study.

2. Background

2.1. What is CAPTCHA?

As shown in **Fig. 2**, a CAPTCHA is generated by an online server and a client would respond in a challenge/response environment. When a client requests a service from an online entity, firstly, he or she sends a request message to the web server. Depending on the nature of the online service type, the web server will then decide whether to allow the user access to the requested resource, or else to authenticate the user before allowing them the online access. Where any requested resources are protected, the web server will invoke the CAPTCHA generator application to create a new CAPTCHA challenge, which is then sent to the user via the same communication channel. The user is then required to solve the challenge and provide the answer to the request in order to prove that they are a human user. A brief description of the operation of CAPTCHA is shown in **Fig. 2**.
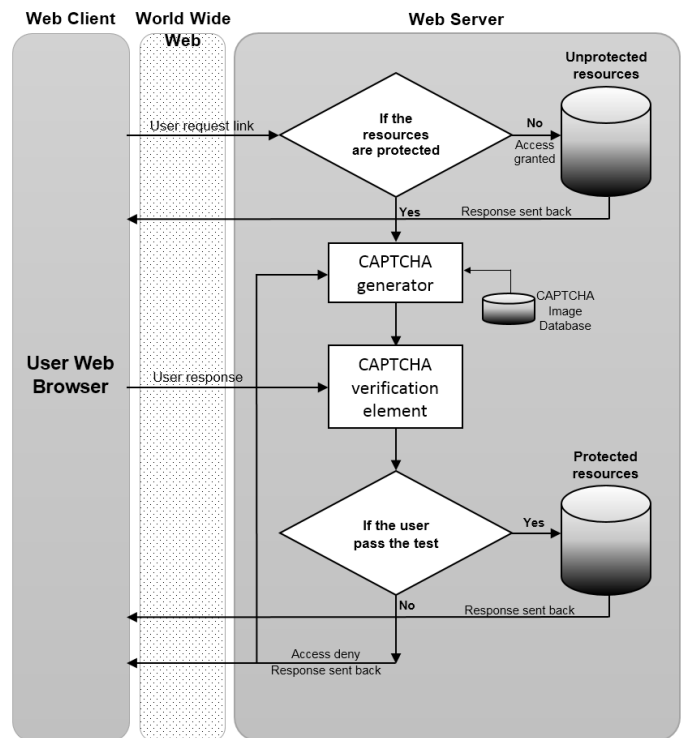
Fig. 2 - General framework for the CAPTCHA authentication process

The CAPTCHA generator application is made up of two separate elements. One is the CAPTCHA image database, which consists of all the information that a CAPTCHA application is using in order to measure the level of accuracy (in other words, whether or not the answer is correct). The second element is called the CAPTCHA verification element, which works as a judge. If the input data from the user matches with the CAPTCHA image stored in the database, then the user will be authenticated and will have access to the requested online services. However, if the user response from the client side does not match the database information on the server, then the CAPTCHA generator application will deny access, and the user will have to try this process again until he or she provides the correct response.

2.2. Types of CAPTCHAs

CAPTCHAs are available in many different shapes and formats depending on their specifications and security functionality. We can classify CAPTCHAs into three main categories: OCR-based, non OCR-based and non-visual-based (Moradi, M. & Keyvanpour, M., 2015). **Fig. 3** is a representation of the different CAPTCHA types based on their classification, where each class has its own conditions and specifications. Based on previous research, OCR-based or text-based CAPTCHAs are the most popular schemes implemented to date. These types of CAPTCHAs are easier in terms of web integration and implementation. In addition, they are more efficient in terms of design and evaluation. The reason that these types of CAPTCHAs are labelled 'OCR-based' is because the challenge is made up of different distorted characters and text. Also, character recognition software (OCR) needs to be used in order to break the CAPTCHA.

Non OCR-based CAPTCHAs utilise multimedia features such as video, images, or a picture of a natural scene. Examples of these types of CAPTCHAs are *Collage CAPTCHA* (Shirali-Shahreza, M. & Shirali-Shahreza, S., 2007), *PIX* (Luis von Ahn, et al., 2004), *Bongo* (Luis von Ahn, et al., 2004), *Asirra* CAPTCHA (J. H. J. S. Jeremy Elson, 2007), and *GeoCAPTCHA* (Te-En Wei, et al., 2012), all of which are based on the recognition of an image or a group of images using a specific criteria. There is another type of CAPTCHA not based on the visual recognition, known as an *audio-based CAPTCHA*. The purpose behind audio-based CAPTCHAs was to overcome some of the weaknesses of OCR-based CAPTCHAs. Specifically, in some OCR-based CAPTCHAs the text is too distorted or significantly deformed, and therefore, it is cumbersome for users to read and recognise. In

that case, they usually put the audio version of the same text as an alternative way of solving the CAPTCHA. This version of the program randomly picks a word or a sequence of numbers. It then transforms the word or sequence of numbers into a short audio clip and mixes it with a reasonable level of background noise. It then presents the distorted audio clip to the user, who must then recognise and type in the content of the sound clip (Haichang Gao, et al., 2010). This method is based on the auditory ability of humans to realise the distorted words or numbers with a moderate level of background noise present; it is very difficult for most voice recognition software to distinguish between a central voice and background noise (Haichang Gao, et al., 2010).

2.3. Related Works

In the last decade, as the number of online threats have been increasing, much attention has been paid to the development of CAPTCHA technologies. However, the vast majority of research in this field concentrates on the text-based CAPTCHA because these kinds of CAPTCHA are the most popular, and are widely used across the web. EZ-GIMPY is one of the most popular dictionary-based CAPTCHA models, which challenges clients to read distorted or corrupted characters (Ahn, 2000-2003). It was originally built for Yahoo! in order to prevent bots from entering their chatrooms, and to prevent computer programs from harvesting a large number of email accounts. However, the model introduced by Mori and Malik (2003) could break this CAPTCHA type with a 33% success rate. In addition, Yan and Ahmad (2007) could also show that most of the current text-based CAPTCHAs can be broken using pixel-count attacks. The Microsoft MSN CAPTCHA was believed to be segmentation resistant.

However, as the paper written by Jeff Yan (Jeff Yan, 2008) proves, a segmentation success rate of 92% can be achieved and this type of text-based CAPTCHA could be broken with a success rate of 60%. The 3D-CAPTCHA is another kind of text-based CAPTCHA, which was introduced to defeat OCR recognition attacks. However, as it has been pointed out by (Vu Duc Nguyen, 2014), they could break this kind of CAPTCHA with a high success rate. Additionally, audio-based CAPTCHAs were designed in order to be robust against computer sound recognition programs compared to text-based CAPTCHAs. Yet, as the results presented in (Yannis Soupionis, 2010) confirm, current audio-based CAPTCHA models are also highly vulnerable to bot attacks.
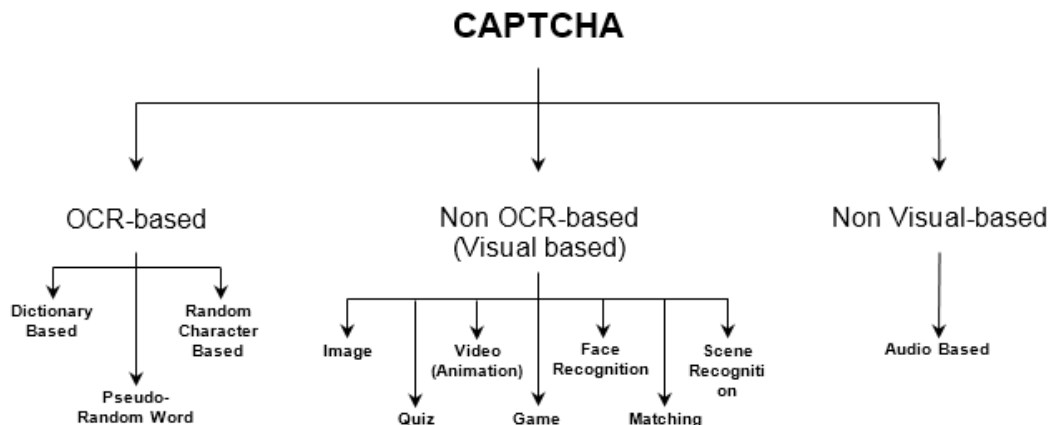
Fig. 3 - A basic categorisation of CAPTCHAs

2.4. Security of CAPTCHAs

Robustness and security of CAPTCHAs has been the centre of attention for researchers in the online and cyber security fields for many years. As the number of online threats is growing at a high rate, the importance of making the CAPTCHA challenge more secure and robust is becoming even more critical. Since OCR-based CAPTCHAs are the most vulnerable types, this section will focus on key aspects of their security (Jeff Yan, 2008). As shown in (Vu Duc Nguyen, 2014), most of the OCR-based CAPTCHAs have been broken using different recognition and segmentation techniques. Optical Character Recognition (OCR) software is the electronic software used widely in order to convert hand written documents or typed glyphs and words into digital format. Most OCR recognition programs are based on pattern recognition techniques using vertical segmentation as a means to separate each glyph and character. This is done in order to recognise every single character by comparing all of the pixels with pre-stored characters on a database (Woodford, 2013).

OCR recognition programs rely on different techniques in order to recognise characters or words. Image preparation or pre-processing techniques are used to convert the image into greyscale or binary format. The pre-processing phase can also include background noise or line removal. Segmentation is a technique to separate each glyph and character in order to analyse the pixel values for each separately, which are then matched with the characters pre-stored on the OCR database or a dictionary (Nicomsoft, 2012). Recognition (or pattern matching) is the final step that OCR programs use in order to match the separated characters to similar characters and glyphs stored on their database (which is dictionary-based). The different processes involved in the recognition of text-based CAPTCHAs are shown in **Fig. 4**.

Segmentation is a very important procedure in any OCR recognition technique, and is needed in order to separate the characters from each other to make recognition possible. According to (Vu Duc Nguyen, 2014), there were a number of 2D-CAPTCHA models believed to be segmentation-resistant. However, as the outcomes of Jeff Yan and Ahmad Salah El Ahmad have shown, most of the 2D-CAPTCHAs can be broken using vertical segmentation techniques (Jeff Yan, 2008). In the same paper, there is also another technique called *pixel count*, which counts the total number of the pixels in each chunk (segment) and by comparing the total number of the pixels with the character information on its database, it would easily map the corresponding character (Jeff Yan, 2009).
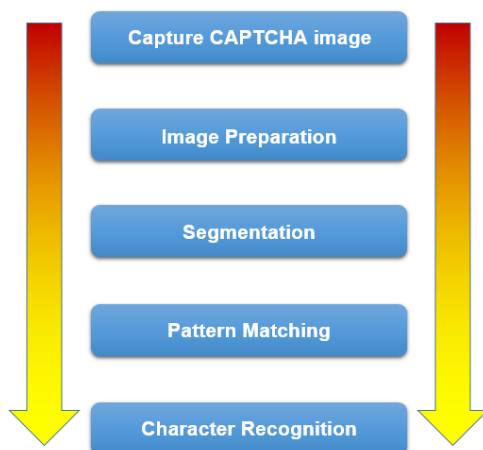
According to (Vu Duc Nguyen, 2014), most of the text-based CAPTCHAs have been broken using different image processing and pattern recognition algorithms. As an example, the famous Google ReCAPTCHA is relevant to mention. This was broken using a holistic approach of recognising the shape context of the words after applying character segmentation and recognition techniques (Vu Duc Nguyen, 2014).

From this evidence, we can conclude that dictionary-based information causes dictionary attacks on these types of CAPTCHA. Therefore, in order to have a more secure and robust model, one of the steps we must consider is avoiding the use of a text string from a dictionary. In other words, if the presented information is random (instead of dictionary-based) it will make it much more difficult for the computer recognition program to decipher and recognise the entire CAPTCHA text (Yan, 2007; Jeff Yan, 2009). In the past, 3D-CAPTCHAs were assumed to be secure and robust because no OCR recognition software was able to break them directly. However, as shown in (Vu Duc Nguyen, 2014), 3D-CAPTCHAs have also been broken using different filters and image-processing techniques with a high recognition success rate. Audio-based CAPTCHAs are also vulnerable to audio recognition techniques, as it has been discussed in (Yannis Soupionis, 2010).

3. Proposed approach

The previous sections of this paper have provided a brief introduction to CAPTCHA and online security, different types of CAPTCHAs in regards to their specifications, and also some of the main security aspects of CAPTCHA challenges. As it can be understood from the definition of CAPTCHA, the challenge should be designed in a way that makes it easy for human users to solve, but not be solvable for computer recognition programs (Luis von Ahn, 2004). Therefore, the usability and security of the CAPTCHA challenge poses a trade-off in CAPTCHA design. In fact, a CAPTCHA could even be very difficult for human users to solve (Saadat Beheshti & Liatsis, 2015). In this section, a novel CAPTCHA model is introduced labelled VICAP (Visual Integration CAPTCHA) based on psychophysics and the properties of the human visual short-term memory (VSTM). This method would superimpose and integrate fleeting frames of visual information captured by the human eye in order to build the final image of an object in the brain. This proposed model is designed to capitalise on a user's sophisticated visual abilities, and therefore, it is logical to conclude that this proposed CAPTCHA model could demonstrate increased security against current computer recognition programs.

3.1. Persistence of vision and CAPTCHA

Currently, neuroscientists and psychologists believe that a key factor enabling us to see the world as integrated and continuous is a phenomenon called persistence of vision. This is the core reason why the world around us does not turn to black with each blink of our eyes (Saadat Beheshti & Liatsis, 2015). The term persistence of vision is the key element in any movie produced by the film industry and is the main reason why a film can be viewed as a smoothly running series of moving images. Every film is made up of a series of individual fleeting images (or frames). By running these in front of the human eye, persistence of vision will cause an illusion so that

**Fig. 4 - Different phases involved in OCR recognition process**

all of the individual frames form of an integrated and uniform shape in our visual system. Since persistence of vision is a unique characteristic of the human eye, we have utilised this distinctive ability in order to distinguish between real human users and automated computer bots.

In order to have a better understanding of how persistence of vision works, we first need to look into the main causes of this phenomenon. *Afterimage* causes our visual system to remember the effect of every single image we see for a very short period of time in our Iconic Memory (IM), following the disappearance of the object from our sight (Clause, 2003-2014). This persistence can last for one tenth to one fifteenth of a second depending on different criteria such as image brightness, colour, and the angle of light (McKinney, 2008).

According to research, Afterimage is the cause of persistence of vision in the brain. Studies dictate that normal and healthy human eyes cannot react or distinguish changes in light frequency in the visual system any faster than a certain period. Thus, the final outcomes will either not be noticeable to the human eye, or the changes in light frequency will be seen in an integrated form (Steven J. Luck, September 2008).

As shown in **Fig. 5** (David E. Irwin & Laura E. Thomas, 2008) when stimulus is present, the human visual system can pick up the most information. However, this visual sensory information will drop gradually after the stimulus disappears from our sight. As it can be seen in **Fig. 5**, the quality of visual sensory information is at the maximum level within 0-50 milliseconds after the stimulus offset and then it decreases rapidly. This quickly decaying function can explain the fundamentals of iconic memory and persistence of vision, as elaborated in (Steven J. Luck, September 2008).
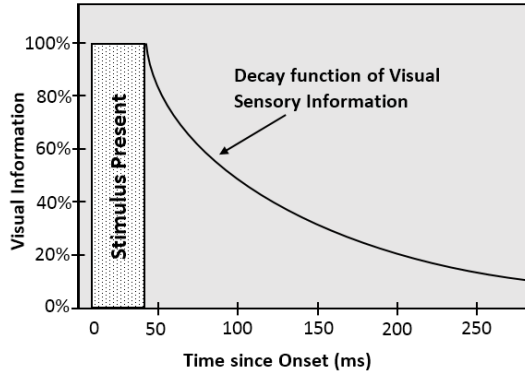

Fig. 5 - The graph represents the rapidly decaying function of visual sensory information following the stimulus offset (David E. Irwin & Laura E. Thomas, 2008)

## 3.2. Persistence of vision and Temporal Integration

Following this brief introduction about the procedure of persistence of vision and the functionality of iconic memory, we would now like to explain how the persistence of vision can cause the human visual system to see the world in an integrated and uniform fashion. To explain this, we firstly need to refer to the definition of persistence of vision. As the light from a stimulus arrives in the retinal part of our eyes, the effect of this light can be retained in our visual system for a brief fraction of a second before disappearing. If a second stimulus is presented whilst the visual information of the first stimulus is still retained, the visual system will perceive the two stimuli as a single stimulus. In psychophysics, this phenomenon is known as Temporal Integration. According to (David E. Irwin &

Laura E. Thomas, 2008), the visual information for each stimulus can persist for about 100-200 milliseconds after its offset. However, this persistency can be affected by a number of factors, such as stimulus intensity, duration, and colour. Temporal Integration is known as the effect of two stimuli appearing with a very short delay, or Inter Stimulus Interval (ISI) from each other. In other words, ISI is defined as the distance between the offset of the first stimulus and the onset of the second stimulus.

If the two stimuli are presented with a long ISI delay, then the visual sensory information for the first stimulus will already be wiped from our sight. Therefore, there will be no integration happening with the second stimulus, as shown in **Fig. 6**. However, when presenting the two stimuli with a very short ISI delay, a person will be able to see the results of two signals as one integrated signal, as shown in **Fig. 7**.
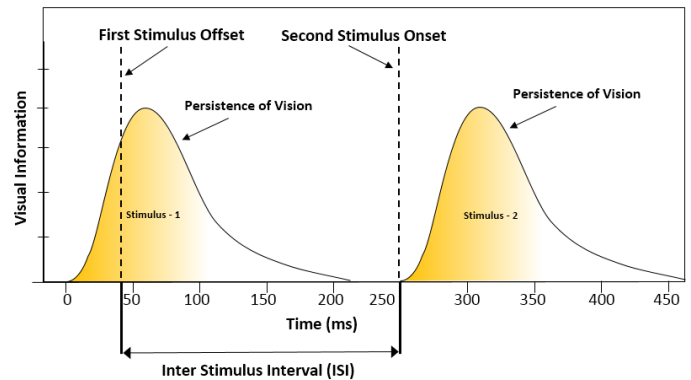
Fig. 6 - Two stimuli are presented with a long ISI delay. As shown in the graph, there is no overlap between the two signals. This means that visual sensory information from the first stimulus is wiped completely before the presentation of the second stimulus.
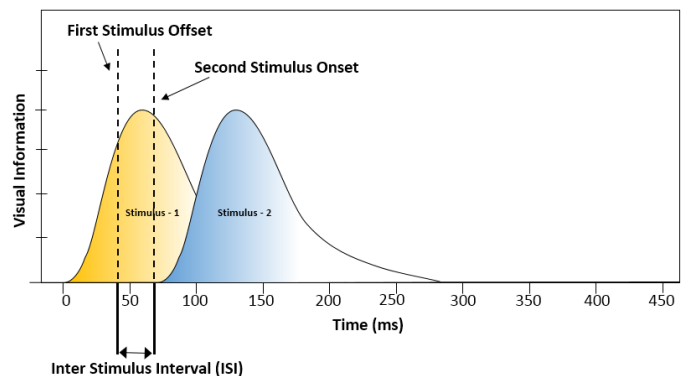
Fig. 7 - Two stimuli are presented with a very short ISI delay, and as a result, there will be some areas where two signals overlap each other. The overlapping areas will have some information from stimulus-1 and some from stimulus-2.

As discussed previously, after the stimulus offset, some of the sensory information will still remain in the visual system for a very brief fraction of a second before completely disappearing (known as persistence of vision). However, if the second stimulus is presented whilst the effects of the first stimulus remain, the human visual system will superimpose the two signals together and perceive them as one integrated image. The new image will contain characteristics from both stimuli (David E. Irwin & Laura E. Thomas, 2008; Steven J. Luck, September 2008). This action is called *temporal*

integration and the properties of this phenomenon are being used for the development of our proposed CAPTCHA model.

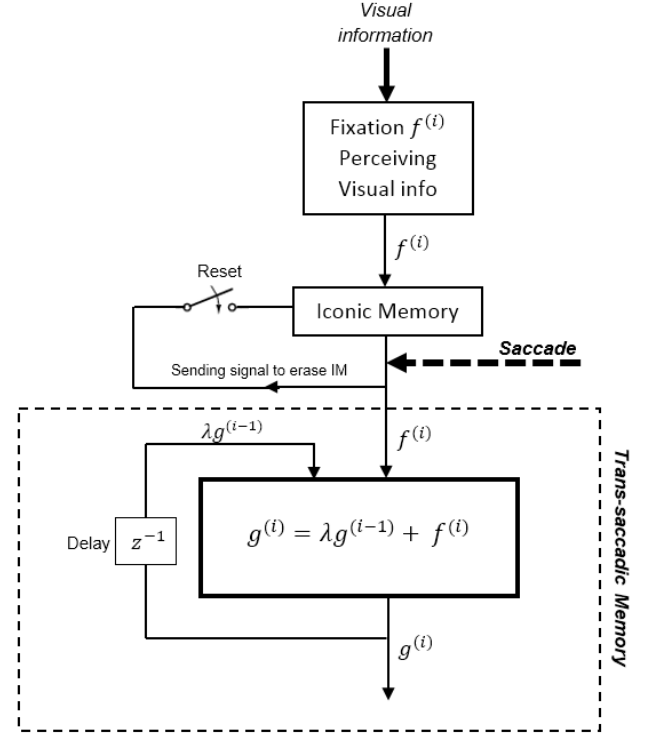## 3.3.   Trans-Saccadic Visual Integration Technique

As has already been acknowledged, persistence of vision causes temporal integration, which subsequently allows us to perceive the surrounding world in an integrated and continuous fashion. In order to appreciate how the mechanism of temporal integration actually works, a novel scheme for temporal integration is proposed in this section, which will help us to understand how a sequence of images (or frames) are combined by the brain using trans-saccadic eye movements. According to research, healthy human eyes are characterised by rapid eye movements, occurring about 3 to 5 times per second, which are known as *saccadic eye movements*. These last for approximately thirty milliseconds on average (Rayner, 1998; Irwin, Jun., 1996). These rapid eye movements are necessary for our visual system in order to perceive a high quality image from the surrounding environment through integration and fusion of visual information. During trans-saccadic eye movements, there are intervals called *fixations*, which last for approximately 300 milliseconds (Irwin, Jun., 1996). During each fixation period, the fovea part of the eye focuses on a particular object and sends the visual information of the object to the brain in order to analyse and process the associated visual information. Fixation periods are separated by rapid eye movements, called *saccades*.

During each saccade, the visual information perceived during the fixation $f^{(i)}$ is combined and superimposed with the visual information from the previous fixation $f^{(i-1)}$. This integration procedure takes place in a part of the memory module known as the *trans-saccadic memory* (Irwin, Jun., 1996) (McKinney, 2008). Yet, in many studies, the trans-saccadic memory has been mentioned to have the same characteristics as the human visual short-term memory (VSTM) (Steven J. Luck, September 2008; D. E. Irwin, 1991). According to (Irwin, Jun., 1996), human working memory is able to process information of 3 to 4 saccades at one time. Therefore, in order to build a stable visual impression of the environment (or a scene), repetition of the visual information is required. In order to better explain how the proposed trans-saccadic integration scheme works, the following section will consider this scheme under two scenarios.

3.3.1.   Single stage scenario

As discussed previously, in order to see a video clip smoothly without flashing images, all the images need to be presented in such a way so as to enable our visual system to integrate and superimpose them. To achieve this goal, the visual information perceived during one fixation period needs to be combined with the visual information perceived during the subsequent fixation period. **Fig. 8** (Saadat Beheshti & Liatsis, 2015) shows the proposed processing scheme for trans-saccadic integration, which takes place in a human's short-term memory. The proposed model starts with perceiving visual information from our environment during a fixation period $f^{(i)}$. This fixation period lasts for about 300 milliseconds and is known as *visual information acquisition*. All the visual information received during each fixation period is stored into the *iconic memory* for a very short period of time before the information is passed to the short-term (or working) memory. With each saccade, the visual information stored in the iconic memory will be passed to the working memory in order to be

integrated with the pre-stored visual information from the previous fixations , $\lambda g^{(i-1)}$. In visual psychophysics, this procedure is called *trans-saccadic integration*. Another important procedure that takes place during each saccade is the erasing of the iconic memory so as to prepare it for receiving new visual information from the eye.

Fig. 8 - The procedure of visual information integration using Trans-Saccadic Memory (Saadat Beheshti & Liatsis, 2015)

3.3.2.   Multi stage scenario

As can be seen in **Fig. 8**, in order to see the world in an integrated and continuous form, the process of temporal integration needs to be repeated. The output image $g^{(i)}$ is the result of superimposing a series of frames, which were perceived during the previous fixation periods, multiplied by a weight, known as *forgetting factor* λ, which can be described as follows:

$$g^{(i)} = \lambda g^{(i-1)} + f^{(i)} \qquad (1)$$

Where, $0 < \lambda \leq 1$ is the forgetting factor which will allocate exponentially less weight to the older samples. The term $\lambda g^{(i-1)}$ refers to the weighted outcome of the previous integration of the fixations. In the model, older samples are allocated a lower weight, and therefore, the influence of their visual information will automatically diminish. By iteratively expanding Equation (1) and substituting the associated terms, we obtain the formula below for a number of n fixations:

$$g^{(n)} = \sum_{i=0}^{n} \lambda^{(n-i)} f^{(i)} \qquad (2)$$

Equation (2) shows how fleeting images will produce the final image in the visual system using the trans-saccadic visual integration technique. As we can see, consecutive frame sequence $A = \{F_1, F_2, F_3, \dots, F_n\}$ is being run at high speed and the human visual system is able to integrate and superimpose all the fleeted frames during different fixation periods in order to produce the final image. Since every single image (or frame) will be retained in our iconic memory for a very short period of

time before being wiped from our memory, it has to be repeated many times in order to for our memory system to remember and memorise the sequence for a very short period of time (in milliseconds). This process of repeating images causes our visual system to distinguish between variations of pixel frequencies. Consequently, the user will recognise the combination image.

Every single frame is made up of a number of random pixels, some of them belonging to the object, while others are background noise. As mentioned earlier, this approach uses a number of consecutive frames $(F_1, F_2, F_3, ...)$ in the sequence "A". These are run fleetingly, and every single image will be retained in the visual memory before disappearing due to the effect of persistence of vision. This will cause the human visual system to combine all seen images and 'reconstruct' the final image, which is the superposition of the previous sequence.

4. Proposed CAPTCHA model

The proposed CAPTCHA model is based on the abilities of the human visual system to retain and superimpose the images of a sequence in order to make up the final image of an object. The aim of this section is to explain the fundamentals of the VICAP security model and the different procedures involved in this model in order to increase the security of websites and online service providers. Two different approaches to this CAPTCHA design will be presented. In the first CAPTCHA model (version 1.0), the application only produces frames that hold key information about the object. This model has some weaknesses, which will be discussed in due course. The second approach (version 2.0) was created in order to increase the robustness of the VICAP model against automated computer attacks, and as will be shown, provided a success rate of over 90%. In other words, it was 90% successful against not being recognizable by current automated computer recognition programs.

4.1. VICAP version 1.0

In order to generate the VICAP output frames sequence, the CAPTCHA-Generator application (CGA) was developed using .NET programming tools, running on a 64-bit Windows operating system with Intel Core-i5 CPU and 3.20 GHz processing power. The main role of the CAPTCHA-Generator application is to render individual VICAP images based on the specific criteria, which will be discussed below, and play them back in a sequence of frames consecutively and smoothly for users in order to produce a film/animation effect. In order to develop the proposed model, a string of characters and numbers are randomly selected from a database. The VICAP model uses a combination of five characters and numbers. Since this combination is selected on a random basis, it will not be possible for the computer programs to guess, as in the case of dictionary-based attacks. The string of characters to be used by the CAPTCHA-Generator application is made up of 18 letters as follows: {A, C, E, F, H, K, L, M, N, P, R, S, T, U, V, W, X, Y} and 6 numbers as follows: {3, 4, 5, 6, 7, 9}. In total, there will be 24 different characters/numbers to be selected randomly by the program.

As can be easily seen, some letters and numbers have been avoided due to the ambiguities that they would cause. For

example, in many cases, the letter "B" can be mistaken with number "8" and vice versa. Another case is letter "O" which can cause ambiguity with the number "0", and also the letter "Z" which can be confused with the number "2". For more information on CAPTCHA usability issues please refer to the work by (Saadat Beheshti & Liatsis, 2015). The procedure of generating the VICAP model consists of the following three steps:

A. Binarization and bitmap conversion:

Binarization is the procedure of converting all pixel values to a binary value (0 or 1), expressed as 1 bit/pixel, which will produce a black and white image. During this process, the grey level of a pixel is compared to a threshold and is allocated the value of zero/one if it is less/greater than the threshold, thus corresponding to black/white, respectively. It is important to convert the pixel values to their binary equivalents because this will simplify the subsequent steps in order to create the final CAPTCHA image. Additionally, the human eye is more sensitive to black and white, rather than colour, stimuli. This is due to the presence of a large number of *rods*, which are photoreceptors in the retina sensitive to shades of grey, rather than *cones*, which are sensitive to colour. Moreover, using colour CAPTCHA images may have a negative effect on the usability and security of the CAPTCHA, as it may increase the risk of CAPTCHA attacks (Saadat Beheshti & Liatsis, 2015).
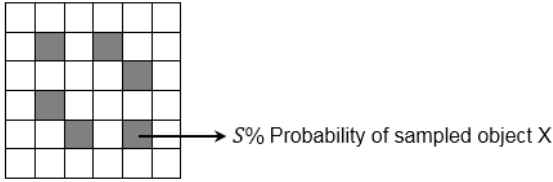
B. Object sampling rate:

Since the proposed VICAP model is based on the ability of the human eye to differentiate between the total number of object pixels and background noise pixels, it is important to choose the correct ratio for the object sampling rate and also the superposition of background noise. In the proposed CAPTCHA model, the object-sampling rate has been chosen in a way that makes it very easy for the human eye to distinguish between the density of object pixels and background noise. However, it is impossible for computer recognition programs to distinguish these two aspects from each other. The sampling rate of the object pixels is random, and therefore, it would be almost impossible for computer programs to predict or learn the behaviour of the pixel elements in terms of appearing or disappearing. **Fig. 9** shows an object corresponding to character "O", made of N^2 number of pixels, where X represents the number of object pixels. In this example, there is a total of $[N^2 - X]$ background pixels.
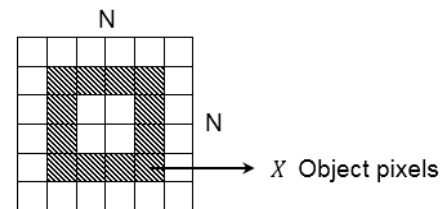
Fig. 9 - An example of a complete object corresponding to the character "O" prior to the application of sampling.

By sampling the object at a rate of $S\%$, there will only be a partial section of the object presented to the user. Since the procedure of the sampling is based on the random generator function, the presentation of the pixels varies from frame to frame. However, overall, the total number of the pixels almost stays the same. For instance, as it is shown in **Fig. 10**, the object "O" has been sampled at a 50% sampling rate, which means the probability of every single pixel appearing in that

frame is almost equal to 50%. Thus, on the single frame scenario there will only be half of the pixels appearing for the object "O" and another 50% of the object pixels will not be shown at all. However, the combination of the pixels can vary from frame to frame, as shown in **Fig. 11**.

S% Probability of sampled object X

Fig. 10 - Object "O" is sampled at 50%
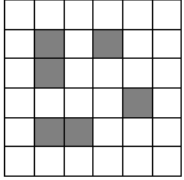
Fig. 11 - Sampling of object "O" at the same sampling rate of 50%, but with a different pixel combination.

This model is based on persistence of vision and the ability of a human's supreme visual system to hold information for a very short period of seconds using iconic memory. Consequently, by presenting only a portion of the object pixels in each single frame and by playing all the frames consecutively, our visual system would be able to combine all the fleeting images and by superimposing the frames together, it will be able to create the final image in the brain.

C. Adding background noise to the CAPTCHA image:

The final step in creating the CAPTCHA images is injecting background noise to the sampled object X. This action will make the CAPTCHA images even harder for optical character recognition (OCR) software to realise or recognise the characters or objects. Since the pixels are randomly selected and presented in each single image it will, in practice, be almost impossible to predict or guess the possible combinations of the pixels in order to extract the final image. As shown in **Fig. 12**, some amount of background noise is added to the sampled object X at a percentage of $n\%$.

n% Probability of the background noise from $(N^2 - X)$
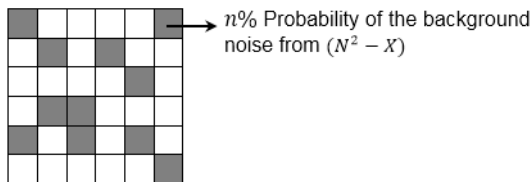
Fig. 12 - Injecting background noise to the sampled object X at the percentage of $n\%$

Fig. 13 shows an example of the three steps involved in generating a single VICAP frame. As it can be seen, step A is producing the binary image of the original data. Step B is the sampling of the object (for instance at a 20% rate in this example), and step C is injecting background noise (at a rate of 15% in this example). After generating the VICAP frames, the next step is presenting the CAPTCHA images at an appropriate animation speed for human users to perceive them as a continuous sequence using the persistence of vision effect. The CAPTCHA-Generator Application is used to render individual images with the required specifications (such as background noise and object sampling rates) and subsequently playback the CAPTCHA sequence for the user.
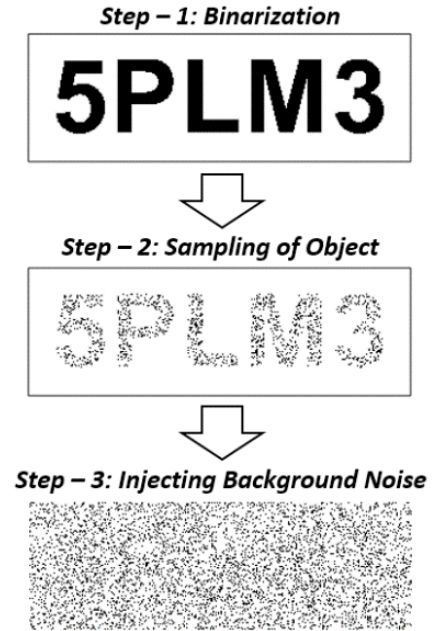
Fig. 13 - Steps involved in generating a single VICAP frame.

As it can be seen from the above example, there is an important issue that needs to be addressed here: namely, the relationship between the object sampling rate (OSR) and the background noise rate (BNR). The key issue here, which we would like to capitalise on, is the appropriate choosing of the density of the object pixels compared to the background noise. By doing this, we could enable the human visual system to distinguish between object pixels and background noise. Therefore, it is absolutely vital to determine the appropriate ratio of OSR/BNR as a means to integrate the noisy information presented in the individual frames to ensure usability, whilst balancing the security aspects of the CAPTCHA in order to maintain robustness to computer program attacks. **Fig. 14** represents 441 experimental results measuring the impact that different combinations of object sampling rate (OSR) and background noise ratio (BNR) have in relation to the computer character recognition rate on the VICAP images. Every value presented in this table is based on the average value for 10 randomly selected CAPTCHA images. These, after being superimposed and rendered using the CAPTCHA-Test Application[1] have been passed on to the OCR recognition program. In total (10 x 441 =) 4410 CAPTCHA experiments have been conducted, and the results are presented as their average value in the above table.

A)   Simulation Results (version-1):

As it can be observed from **Fig. 14**, this research includes over 4000 simulation experiments, which were conducted using a variety of object sampling rates and background noise levels in order to examine the impact on the final output image in terms of computer recognition levels. In order to achieve this, the OSR was selected at a rate of 0% to begin with and then increased by 5% granularity until it reached 100%. Similarly, the background noise levels started at 0% and were increased by 5% until they reached 100%.

---

[1] *CAPTCHA-Test Application* is the state-of-the-art application and was developed as a part of PhD project in order to simulate and render a final superimposed image of the individual VICAP frames in order to test and measure the security and robustness level of the proposed CAPTCHA model.

	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
0	0%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
5	100%	0%	0%	50%	75%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
10	100%	0%	0%	0%	50%	75%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
15	100%	50%	0%	0%	0%	50%	75%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20	100%	75%	50%	0%	0%	0%	50%	75%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
25	100%	100%	75%	50%	0%	0%	0%	50%	75%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
30	100%	100%	100%	75%	50%	0%	0%	0%	50%	75%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
35	100%	100%	100%	100%	75%	50%	0%	0%	0%	50%	75%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
40	100%	100%	100%	100%	100%	75%	50%	0%	0%	0%	50%	75%	100%	100%	100%	100%	100%	100%	100%	100%	100%
45	100%	100%	100%	100%	100%	100%	75%	50%	0%	0%	0%	50%	75%	100%	100%	100%	100%	100%	100%	100%	100%
50	100%	100%	100%	100%	100%	100%	100%	75%	50%	0%	0%	0%	50%	75%	100%	100%	100%	100%	100%	100%	100%
55	100%	100%	100%	100%	100%	100%	100%	100%	75%	50%	0%	0%	0%	50%	75%	100%	100%	100%	100%	100%	100%
60	100%	100%	100%	100%	100%	100%	100%	100%	100%	75%	50%	0%	0%	0%	50%	75%	100%	100%	100%	100%	100%
65	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	75%	50%	0%	0%	0%	50%	75%	100%	100%	100%	100%
70	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	75%	50%	0%	0%	0%	50%	75%	100%	100%	100%
75	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	75%	50%	0%	0%	0%	50%	75%	100%	100%
80	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	75%	50%	0%	0%	0%	50%	75%	100%
85	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	75%	50%	0%	0%	0%	50%	100%
90	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	75%	50%	0%	0%	0%	100%
95	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	75%	50%	0%	0%	100%
100	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%

Fig. 14 - Comparison of different object sampling rates-OSR (horizontal rows) vs. backgrounds noise ratio-BNR (vertical columns) and the output results in terms of computer recognition success rate (in percentage).

All output results were generated using one of the most sophisticated CAPTCHA breaker programs, called GSA CAPTCHA Breaker (GSA, 2014), which uses three different OCR engines in order to recognise every single character and number. It also has one of the most advanced databases, and is able to break the majority of current CAPTCHAs with over a success rate of over 99% (GSA, 2014). By applying GSA Captcha Breaker, the performance of some of the most popular CAPTCHA decoders such as *DeCaptcher*, *DeathByCaptcha*, *Bypass Captcha* is evaluated on the proposed CAPTCHA model (GSA, 2016). The above named CAPTCHA attacks will be discussed more in details in the section 5.1 VICAP security analysis.

Fig. 15 represents examples of 10 different output frames generated by the VICAP-Generator application. Due to space limitations, we are not able to show the final output effect. This has to be experienced in the real world. Interested readers may experience the proposed model on the CAPTCHA evaluation and user experience website at: *http://mrbeheshti2.wixsite.com/captcha/vicap*
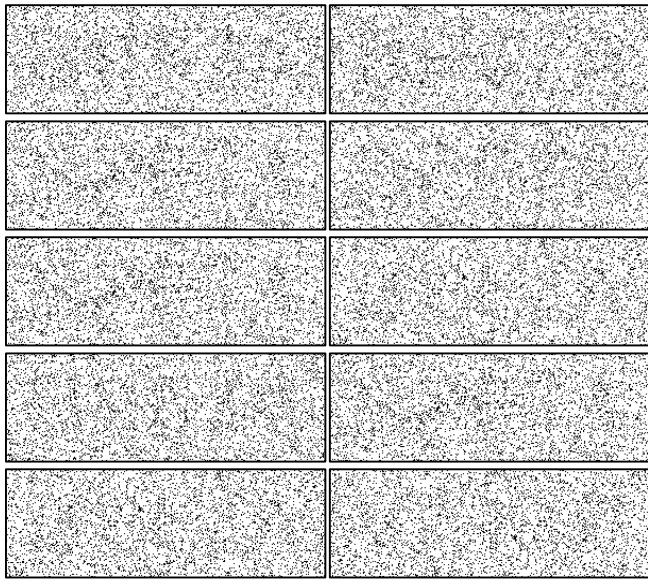

Fig. 15 - Example of ten different output frames generated and rendered by the VICAP-Generator Application with 15% background noise and 25% object sampling rates. The user would be able to see the string of "AE34M" easily, while there is no information that can be seen by analysing individual frames.

Iconic memory is known as a key factor that causes persistence of vision. Forming the final object based on the presentation of a series of frames depends on the difference in the density of the object pixels and background pixels. The way our visual system can distinguish between these two sets of pixels is by analysing the density of all presented pixels in a single frame basis, and then capturing the initial frame in the iconic memory and comparing it with the subsequent frame. As it can be understood from Equation (1) and (2) in the section 3.3.2, in order for the human eye to distinguish between the object pixels and the background noise pixels, the whole process of displaying the CAPTCHA frames needs to be repeated. This process of repeating will cause our visual system to distinguish between pixel density of the object and background noise. Subsequently, based on the density information of the pixels, the human eye would be able to recognise the final object. However, the main question raised here is: what if a computer program also compares the individual pixel frequencies in terms of multiple frames? By mapping these frequency values to the location of the pixels, in practice, the computer program would be able to reconstruct the object too.

To address this problem, the proposed CAPTCHA model has been tested using the CAPTCHA-Test Application. This application would work in a very similar way to the human visual system. It could calculate the pixel occurrence frequencies according to the weight of each pixel, and by mapping the output results against different shades of grey, the application would be able to simulate the final image of the object in greyscale (8-bit/pixel) as well as 1-bit/pixel in a black and white image.

This task will be done by making a graphical representation of the most frequent pixels by analysing the pixel value based on a single frame, and then expanding the calculation to the rest of the frames. In this experiment, we have chosen 10 consecutive frames. Thus, as it can be observed from **Fig. 16** (VICAP version 1.0) the developed application is easily able to retrieve the hidden information from 10 different CAPTCHA frames shown in **Fig. 15**, and can recover the final object image very clearly. This is a significant weakness of the first generation model. For this reason, the second CAPTCHA design (version 2.0) was generated in order to overcome this recognition problem.

Fig. 16 - Final integrated output result from superimposing 10 CAPTCHA frames shown as an 8-bit/pixel greyscale image. String of "AE34M" can be clearly retrieved by OCR recognition software using CAPTCHA v.1.

4.2. VICAP version 2.0

In the second CAPTCHA model, the process of generating CAPTCHA frames is very similar to the first model, with the only difference being that in the second model, there will be two production lines working concurrently. In the production line-1, the CAPTCHA images (which are original frames containing object information, shown using symbol "O") are generated with the same procedure as before, while at the same time in the production line-2, random frames (shown using symbol "R") are generated, containing random pixels at a rate of n. Where n is chosen is the same value as the background noise rate in the CAPTCHA frames of production line-1. Since both pixel frames (background from production line-1 and production line-2) are generated from the same random process (and thus have the same characteristics), they become indistinguishable. This increases the robustness of the model. There is currently no available recognition software that is able to distinguish between these two series. An example of these two series of frame generator engines is given below:

(3) $Production\ line\text{-}1\text{:}\ Original\ frames\ =$
         $\{O_1, O_2, O_3, \dots, O_m\}$

(4) $Production\ line\text{-}2\text{:}\ Random\ frames\ =$
         $\{R_1, R_2, R_3, \dots, R_m\}$

Here, a key parameter is introduced called *Original-to-Random Output* (ORO) frames, which plays a key role in the second CAPTCHA design. This defines the properties of the mixing procedure, by specifying the percentage of random and original frames in the final sequence. A higher percentage of ORO translates to a larger number of original frames and a smaller number of random frames being mixed together in the sequence, and vice versa. An example of how ORO can be used to mix these two series of frames is shown below:

(5) $Output\ frame\ sequence\ =$
         $\{R_1, R_2, R_3, \boldsymbol{O_1}, R_4, \boldsymbol{O_2}, R_5, \dots\}$

As it can be observed from Equation 5, the original frames are being mixed randomly with random frames with $ORO \simeq$ 30%. The mixing procedure is performed on a random basis. Therefore, in practice, it would not be possible to detect the ordering of the frames as a means to separate original frames from random ones. However, the superior properties of the human visual system are able to superimpose the structured information in the presented frames, and thus distinguish the object from the background noise. **Fig. 18** shows an example of the final output frames generated and rendered using an object-sampling rate of 25%, a background noise rate of 15%, and an ORO parameter set to 20%. As it can be seen from **Fig. 18**, by analysing every single frame, no useful information can be observed. Yet, by running all frames at high speed, it is

possible for the human visual system to recognise the hidden object in the frame sequence.

B)   Simulation Results (version 2):

As discussed previously, having different background noise and object sampling rates affects the level of robustness of the proposed CAPTCHA model. Thus, as shown in **Fig. 17**, an object-sampling rate of 25% and a background noise rate of 15% from the VICAP version 1 give a computer recognition success rate of 50%. When introducing the VICAP version 2, the computer recognition success rate drops rapidly to near 0%, as shown in **Fig. 19** and **Fig. 20**. The final simulation output results for the computer recognition success rate for VICAP version 2 is shown in the **Fig. 19** and **Fig. 20** using the CAPTCHA-Test Application for 8-bit/pixel greyscale format and 1-bit/pixel black and white format. As shown in these Figures, no useful information can be retrieved from the series of frames and thus, it is expected that the computer recognition rate is almost close to 0%.

	5	10	15	20	25	30
5	0.0%	50.0%	75.0%	100.0%	100.0%	100.0
10	0.0%	0.0%	0.0%	50.0%	75.0%	100.0
15	50.0%	0.0%	0.0%	0.0%	50.0%	75.0%
20	75.0%	50.0%	0.0%	0.0%	0.0%	50.0%
25	100.0%	75.0%	50.0%	0.0%	0.0%	0.0%
30	100.0%	100.0%	75.0%	50.0%	0.0%	

Fig. 17 - The table shows a 25% object sampling rate and a 15% background noise rate will give rise to a 50% risk of VICAP v.1 being defeated by computer character recognition software.
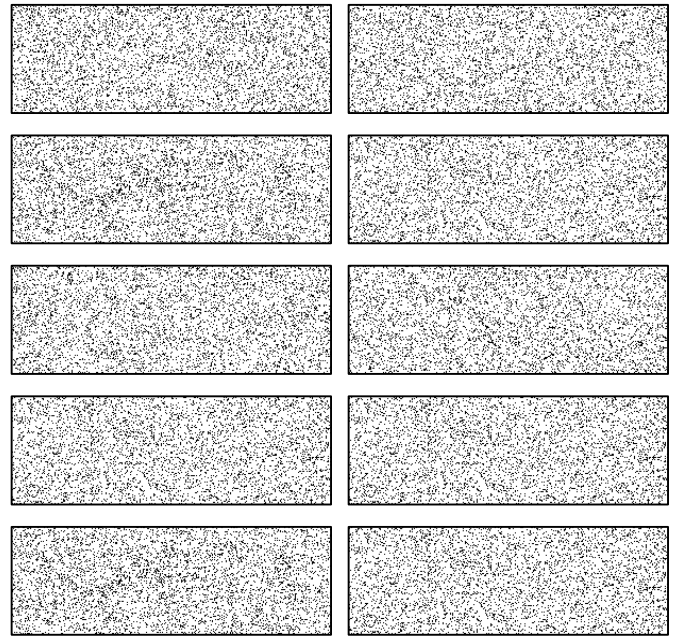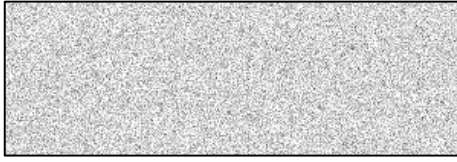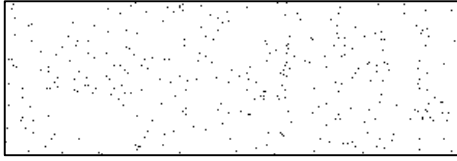

Fig. 18 – An example of 10 individual CAPTCHA frames generated and rendered by the VICAP-generator application with a 15% background noise and a 25% object sampling rate and ORO=20%.

Fig. 19 – The final integrated output result from superimposing 10 CAPTCHA frames shown as an 8-bit/pixel greyscale image. No useful information can be retrieved in the second scenario.

Fig. 20 – The final integrated output result from superimposing 10 CAPTCHA frames shown as a 1-bit/pixel black and white image. No useful information can be retrieved in the second scenario.

5. Evaluation results and discussion

5.1. VICAP security analysis

As discussed previously, the level of distinguishability of the letters and numbers in the proposed CAPTCHA model directly depends on the value of the ORO parameter, as it defines the recognition success rate for both human users and computer recognition software (OCR). In this section, the VICAP v.2 model was tested using different techniques and under different conditions for both human users as well as computer recognition software in order to determine the optimal level for the ORO parameter. This translates to the threshold value for which the CAPTCHA is easily recognizable and visible for the human users, but at the same time it is difficult (if not impossible) for computer attackers to recognize.

5.1.1 Computer recognition success rate experimental results (CRSR)

In order to measure the level of robustness and security of the proposed CAPTCHA model, various experiments were conducted in laboratory conditions using a computer with Intel Core-i5 CPU and 3.20 GHz processor. 200 attacks were simulated using the state of the art GSA-Captcha Breaker (GSA, 2014). The aim of these experiments was to determine the resistance level (or threshold value) of the proposed VICAP model against different computer recognition attacks. The experiments began with ORO = 0% and the computer recognition success rate was measured for every single experiment. Then, the granularity of ORO increased by 10% until it reached 100% to complete the whole test. To generate statistically meaningful results, for every single setting of the ORO parameter, a chunk of 20 randomly generated CAPTCHAs were fed to the GSA-CAPTCHA Breaker application. Thus, a total of 200 experiments were conducted.

The final proposed CAPTCHA model is based on the processes of persistence vision and also the properties of temporal integration. After playing back and analyzing the individual frame information, no useful information could be retrieved. Thus, on an individual frame basis, the model is 100% secure against any current character recognition application for two main reasons. Firstly, the characters and

numbers are completely faded into background noise using sampling techniques (refer to section 4.1.C of this paper). Consequently, every single frame is made up of random dots (or black pixels) which, on their own, have no meaning. Secondly, the way OCR programs recognize text is by firstly segmenting the characters, and then recognizing them. Since the letters and numbers in our model are sampled, only a portion of their pixels appear in each frame, which may not carry sufficient information about the object. So, OCR may not be able to segment the VICAP frames. The only possible way to attack this CAPTCHA model is to superimpose the information in the frames in a very similar way to our own visual system.

For this reason, in order to test the proposed CAPTCHA model, we need to simulate a final output image, which implements the integration rules of Equations (1) and (2). In order to get the final simulation output results, we have used our CAPTCHA-Test Application. As discussed previously, the ORO threshold value is the key parameter that is being tested and evaluated in this section. In order to better understand how the ORO parameter can affect the visibility and distinguishability of the characters and numbers in VICAP v.2, some of the superimposed output results are presented in **Fig. 21** to **Fig. 25**. As it can be observed from the output results, the distinguishability of the characters is at the lowest level when the ORO parameter is about 10% and gradually increases as the value of ORO parameter increases accordingly.
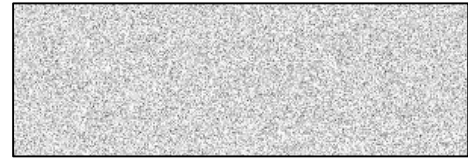
Fig. 21 - Superimposition rendered output image with ORO = 10%.
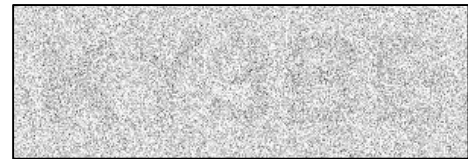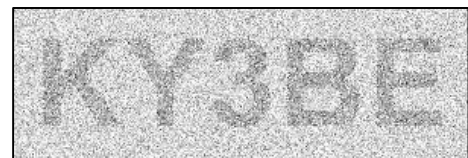

Fig. 22 - Superimposition rendered output image with ORO = 30%.

Fig. 23 - Superimposition rendered output image with ORO = 50%.

Fig. 24 - Superimposition rendered output image with ORO = 70%.

**Fig. 25 - Superimposition rendered output image with ORO = 90%.**

In order to measure the security and robustness levels of the proposed CAPTCHA model v.2, 200 experiments have been conducted in the same laboratory conditions as explained before. Every single experiment consisted of superimposing images of 10 randomly selected VICAP frames with a specific ORO value as shown above. Ten different ration of ORO parameter (from zero to 100%) and 20 randomly rendered and superimposed VICAP images per ratio were tested. Thus, a total of 200 different VICAP security experiments have been conducted. As stated before, individual VICAP frames were rendered and generated using state of the art VICAP-Generator Application and after that, every group of frames were superimposed into a single image using the CAPTCHA-Test Application.

After rendering and generating 200 different superimposed CAPTCHA images using the CAPTCHA-Test Application, these were passed on to the GSA-CAPTCHA Breaker in order to measure the security and robustness levels for each of the experiments individually. As **Fig. 26** shows, the GSA-CAPTCHA Breaker was able to break our proposed model with a success rate of 100%, when ORO = 90% in about 35 seconds. As the value of ORO parameter dropped, the probability of mixing original frames with random frames also decreased. As the number of original frames dropped, less information about the object was presented. Therefore, it would be more difficult for computer recognition software to decipher useful information and subsequently, this would translate to a reduced recognition success rate for automated computer programs. **Fig. 27** shows that computer recognition attacks are not able to break our proposed CAPTCHA model when the ORO parameter is equal to 20% or less.
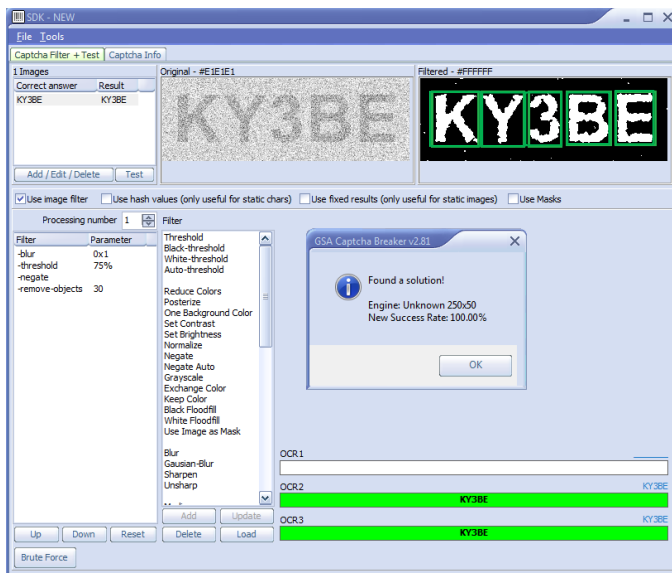
Fig. 26 Screenshot of the GSA-CAPTCHA Breaker Application. VICAP is recognised by the CAPTCHA breaker application with a percentage of Original-to-Random Output frames (ORO) = 90%, with a recognition success rate of 100%.
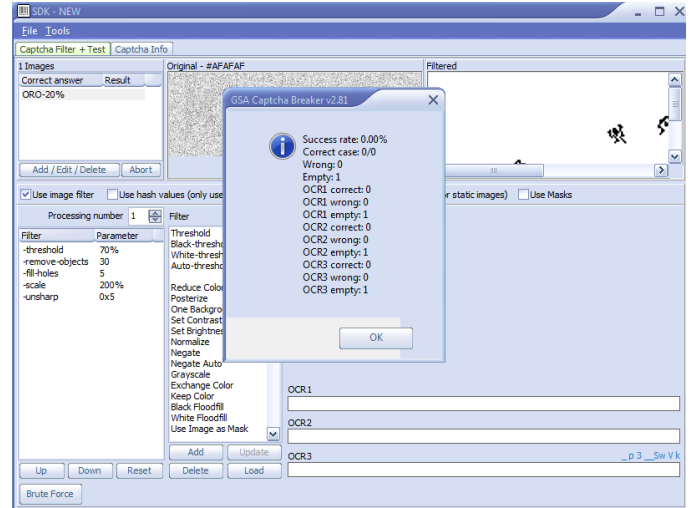
Fig. 27 - Screenshot of the GSA-CAPTCHA Breaker application. The application is not able to retrieve any information from the object and in this case the recognition success rate = 0% at ORO = 20%.

Table 1 represents the comparison of different ORO values and the corresponding computer recognition success rates. As the table shows, the value of ORO parameter can directly affect the computer recognition success rate in a way that by increasing the ORO parameter, the computer recognition rate would also increase. There would be a threshold value (highlighted in the table in green) which is determined to be ORO = 20% in this experiment, where the computer recognition success rate equals 0%. This threshold value plays a key role in this experiment, as for values of ORO > 20% computer recognition attacks will be able to break the proposed CAPTCHA model. Therefore, as we can see from **Table 1**, there is a significant jump in terms of computer recognition success rate, going from 0% to 20% when ORO increases from 20% to 30%.

Table 1. Comparison of computer recognition success rate versus Original-to-Random Output frame (ORO) parameter.	
ORO Parameters	**Computer Recognition Success Rate**
0%	0%
10%	0%
20%	**0%**
30%	20%
40%	50%
50%	80%
60%	90%
70%	100%
80%	100%
90%	100%
100%	100%

In order to find out whether the sharp jump in terms of CRSR in the range of ORO from 20% to 30% is instantaneous or a gradual, we ran an extensive series of new experiments under the same laboratory conditions as explained previously under two scenarios, 1- Individual characters and 2- Entire CAPTCHA string, as it is explained below.

Scenario 1: (Individual Characters)

In this experiment, the recognition success rate for computer character recognition programs was evaluated using 220 randomly generated VICAP images, containing combination of five randomly selected letters or numbers. Each superimposed VICAP image was rendered using the Captcha-Test application and in total 20 randomly generated and superimposed VICAP images were tested for each value of ORO. The percentage increase in ORO values was set to 1%, thus resulting in 220 VICAP rendered images in the ORO range between 20% to 30%. In the first scenario, we measured the output results in regards to providing the correct response for a single character rather than the entire CAPTCHA. As it can be seen from the output results, as the value of ORO increases, so does the CRSR value. It is interesting to note that up to a value of ORO equal to 26%, computer recognition rates are at 0%, however from this value onwards there is a non-linear relationship in terms of the increase of CRSR. This is exemplified by the massive jump from 29% to 30%, where CRSR quadruples. The output results for CRSR for individual characters are shown in Table 2.

Table 2. The change of CRSR per individual characters for variation of ORO parameters

ORO Parameters	CRSR for Individual Character
20%	0%
21%	0%
22%	0%
23%	0%
24%	0%
25%	0%
26%	0%
27%	1%
28%	3%
29%	7%
30%	28%

Scenario 2: (Entire CAPTCHA String)

The purpose of the second experiment is to evaluate computer recognition success rates in the context of deciphering the entire CAPTCHA, rather than simply considering partial recognition of a string of five individual characters and numbers. As in the previous experiment, we set the percentage increase in terms of ORO to 1%, and created a test database of 20 randomly generated and superimposed VICAP images for each ORO value, resulting to a total of 220 images. As expected, computer recognition 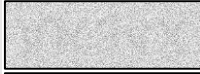rates of a CAPTCHA string are lower than those of individual characters. Specifically, up to an ORO level of 29%, CRSR is at 0%, followed by a step increase to 20% at an ORO rate of 30% as shown in Table 3.

Table 3. The change of CRSR for entire CAPTCHA string for variation of ORO parameters

ORO Parameters	CRSR for Entire CAPTCHA
20%	0%
21%	0%
22%	0%
23%	0%
24%	0%
25%	0%
26%	0%
27%	0%
28%	0%
29%	0%
30%	20%

In order to enhance the accuracy of the final recognition output results in terms of computer attacks, the proposed CAPTCHA model was tested against most popular CAPTCHA attacks. As it was explained previously, "GSA Captcha Breaker" is a software that uses three different OCR engines in order to recognise every single character and number. "Captcha Sniper" is another type of CAPTCHA decoder that works very similarly to "GSA Captcha breaker" and both of them are based on the same concept. By applying "Captcha Sniper" and "GSA Captcha breaker", the performance of some of the most popular CAPTCHA decoders such as DeCaptcher, DeathByCaptcha, Bypass Captcha is evaluated on the proposed CAPTCHA model (Anon., 2016) (GSA, 2016). The table below represents some of the output results made from the experiments using the named CAPTCHA decoders.

Table 4. Recognition output results for different CAPTCHA decoders

CAPTCHA Decoder	ORO	Output Rendered VICAP Image	RSR Recognition Success Rate	Recognised Answer
GSA Captcha Breaker	20%		0%	Null
Captcha Sniper	20%		0%	Null
DeCaptcher	20%		0%	Null
DeathByCaptcha	20%		0%	Null
Bypass Captcha	20%		0%	Null

5.1.2. Human recognition success rate experimental results (HRSR)

In previous experiments, we measured the highest threshold value for the ORO parameter that would give 0% for computer recognition programs to break our proposed CAPTCHA model. Any CAPTCHA model should satisfy two conditions in order to be considered valid: namely, as acknowledged, be too difficult or impossible for computers to break and very easy for humans to solve. Therefore, another set of experiments should be repeated for human users in order to find the threshold value

for the ORO parameter that would enable human users to recognize the CAPTCHA very easily.

The aim of running these series of experiments is to find the optimal level for ORO parameter that gives the best possible distinguishability level of the CAPTCHA for human users, and the minimum level of recognition success rates for computer programs. In order to run the second experiment, the CAPTCHA User Evaluation Website[2] was designed to enable users to participate in VICAP model evaluation and provide their feedback regarding usability and distinguishability. In total, 150 participants from different age groups and backgrounds participated in the experiment through this website. The website is designed in such way that every time a user visits the website, a new set of random characters and numbers is generated and displayed to them by recording the IP address of the user.

Feedback from the 150 participants was collected and the average value for each grade of the ORO parameter was calculated, based on the ability of the users to see and recognise the CAPTCHA challenge. **Table 5** shows a comparison of 10 different ORO parameters and their associated human recognition success rates. As it can be seen from the table, the test is run at ORO = 0%, where there are no original frames presented and therefore, no useful information is presented to the users. Consequently, as we can see from the results, the human recognition success rate is also equal to 0%. By increasing the value of the ORO parameter, the recognition success rate for the users also increases rapidly. However, as the experimental results can confirm, the sophisticated recognition abilities of the human visual system react more sharply and accurately than computer recognition programs. As it can be deduced from **Table 5**, for ORO < 20%, the recognition success rate is almost equal to 0%. This means that for ORO < 20%, there is no sufficient information about the object and it is difficult for human users to see and recognise it. However, as the results from **Table 5** show, the first big jump in terms of the human recognition success rate is at ORO = 20%, which gives a recognition success rate of 65% as highlighted in the table. Surprisingly, in the second experiment similar to the first, the threshold value is measured at ORO = 20%, where there is the first big jump in terms recognition success rate for the human users.

Table 5. Comparison of different Original-to-Random Output (ORO) frame rate parameter and the effect of human recognition success rate.

ORO Parameters	Human Recognition Success Rate
0%	0%
10%	0%
20%	**65%**
30%	80%
40%	100%
50%	100%
60%	100%
70%	100%
80%	100%
90%	100%
100%	100%

---

[2] The CAPTCHA user evaluation website is accessible at *http://mrbeheshti2.wixsite.com/captcha/vicap*

In order to enhance our output results, we would like to conduct another series of user experiments to determine whether the sharp jump in HRSR in the range of ORO values between 10% to 20% is instantaneous or follows a gradual increase. For this reason, the CAPTCHA user evaluation website was used as in the previous experiments. Over 50 participants participated in this CAPTCHA user experience evaluation tests. The feedback provided by the participants was collected and the output results are analysed and discussed in the two scenarios as explained below.

Scenario 1: (Individual Characters)

In this experiment, we are concerned with human recognition success rates under the individual character scenario. We set the percentage increase of ORO to 1%, so as to better capture the nature of the human recognition performance. The range of ORO values for this experiment is between 10% to 20%. Table 6 summarizes the findings of our research.

Table 6. The change of HRSR for individual characters for variation of ORO parameters

ORO Parameters	HRSR for Individual Characters
10%	0%
11%	0%
12%	0%
13%	4%
14%	8%
15%	12%
16%	20%
17%	28%
18%	40%
19%	56%
20%	84%

As the final output results confirm, up to an ORO level of 12%, HRSR is 0%, however following this level, there is a nearly linear increase for an ORO rate up to 17%. For values higher than 18%, human users were able to recognise the single character CAPTCHAs with good accuracy, rising to 84% at ORO of 20%.

Scenario 2: (Entire CAPTCHA String)

By analysing the output feedback received from the participants in Scenario 1, we focused our attention in evaluating human recognition success rates for string CAPTCHAs. As the final output results from over 50 participants confirm, the recognition success rates for ORO values up to 19% were equal to 0%. This means that no participants were able to recognise the entire CAPTCHA string, when the ORO parameter is less than 19%. This is consistent with the results of the previous experiment, where recognition rates of individual characters in this range were quite low for human users. At an ORO rate of 20%, HRSR suddenly jumps to 60%, which translates to some CAPTCHAs being deciphered completely by the majority of participants as shown in Table 7. Nevertheless, this is still lower than the 84% recognition accuracy achieved in the case of individual character scenario, shown in Table 6 of experiment 2.

Table 7. The change of HRSR for entire CAPTCHA string for variation of ORO parameters	
ORO Parameters	**HRSR for Entire CAPTCHA String**
10%	0%
11%	0%
12%	0%
13%	0%
14%	0%
15%	0%
16%	0%
17%	0%
18%	0%
19%	0%
20%	60%

In orde to justify this big jump in term of HRSR, we can compare the output results from indivual recgtion rate and entire CAPTCHA recognition rate. By looking at the analytical results it can be undrestant that in the case of ORO equals and less than 19%, only partial information was readable by the human users and none of the partipants were able to decode the entire CAPTCHA. Thereofe, the HRSR for the entire CAPTCHA string becomes 0%. Hoever, by increasing the value of ORO to 20%, therfore will be sufficient original frames to present to the user inroder to form the final object in the bran. Thus, the entire CAPTCHA becomes more readable for majority of the users. The output results for ORO = 20% was indicating that 3 out of 5 users were able to answer the entire CAPTCHA correcly which produce HRSR = 60%.

5.1.3. ORO analytical comparison

The proposed CAPTCHA model is designed specifically to work in collaboration with the human visual system; therefore, the expectation is to get better and more accurate results for humans rather than computers. **Fig. 28** shows a comparison in terms of recognition success rate for both human users and computer recognition programs versus the different ratios of the ORO parameter. As the graph confirms, the human recognition success rate rises earlier and faster than the computer recognition success rate.
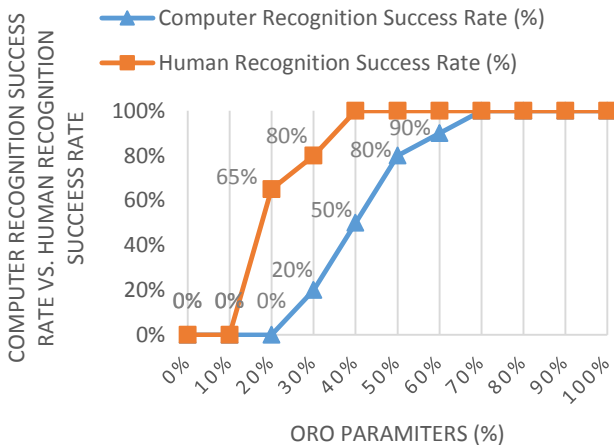
Fig. 28 - Comparison of computer recognition success rate vs. human recognition success rate according to different ORO values.

As it can be observed from **Fig. 28**, the recognition success rate for ORO values is less than 20% for both human users and those for computers is equal to zero. This means there is no sufficient information presented about the object and the users (nor automated computer programs) are not able to perceive any useful information. A critical value for ORO, which has been determined during the experiments of this research is when ORO = 20%, for which there is a big jump from 0% to 65% in terms of human recognition success rate while the computer recognition success rate is still at 0%. The ratio of ORO = 20% is defined as the Optimal Character Recognition rate. At this specific rate, the VICAP model can satisfy the requirements of a CAPTCHA.

As **Fig. 28** shows, by increasing the value of the ORO parameter further, the recognition success rate will improve for both humans and computers. However, looking at the graph, it is clear that the recognition success rate increases more quickly for humans than computers. For example, according to the experimental results, the participants in the evaluation experiments were able to see the characters with 100% success, at the rate of ORO = 40%, while the same level of recognition success rate for the computer programs would be possible at ORO = 70%. This significant difference confirms the strength of the proposed model, which capitalises on the properties of the human visual system. Conclusively, in terms of the recognition success rate, as it can be observed from **Fig. 28**, the optimal setting for the ORO parameter is ORO = 20%, which provides a satisfactory success rate of 65% for human users, while demonstrating robustness to OCR attacks with a recognition rate of 0%.

5.2. Recognition Improvement Level (*RIL*)

In the previous section, different levels of recognition success rates for both human users and computer programs were compared and analysed according to varying values of the ORO parameter. In this section, we focus our attention on the analysis of the Recognition Improvement Level (RIL) in terms of security and robustness of the CAPTCHA challenge. RIL can be simply defined as the difference between the human recognition and the computer recognition levels. This difference may demonstrate the actual impact of the VICAP model in terms of the human recognition success rate and the computer recognition robustness level. The RIL parameter is important in this research because when we study the impact of different choices of the ORO parameter on the test, it is important that we understand what values of ORO will provide the best possible performance for the proposed CAPTCHA model. In other words, we would like to determine the percentage of the ORO parameter where the CAPTCHA challenge would have the highest level in terms of human recognition rate, but at the same time, have the lowest possible level for the computer recognition rate, i.e., the greatest difference between the two. This is possible to achieve by examining the RIL parameter.

When the RIL value is at its maximum, the parameters of the proposed CAPTCHA model are optimally chosen. In other words, the CAPTCHA is very hard for the computer programs to break, while it is very easy for human users to solve. Similarly, as the value of RIL gets smaller, the CAPTCHA is more vulnerable against different computer-based attacks. **Table 8** is a comparison between various choices of the ORO parameter and the associated RIL values.

Table 8. Recognition Improvement Level (*RIL*); Comparison of different values of the ORO parameter and their corresponding *RIL* values.

ORO	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
RIL	0%	65%	60%	50%	20%	10%	0%	0%	0%	0%

As it can be observed from **Table 8**, the value of the RIL parameter in the beginning is equal to 0%, meaning that the CAPTCHA is neither recognisable to humans nor to computer programs. The highest value of RIL parameter occurs when ORO = 20%, where RIL = 65%. As the value of the ORO parameter increases, the value of the RIL parameter decreases, meaning that the security level of the test is reducing rapidly. From ORO = 30% until ORO = 60% the robustness and security level of the proposed CAPTCHA model is decreasing fast until it reaches the level of ORO = 70%. At this rate, the proposed CAPTCHA model will be vulnerable against most of the character recognition techniques. Therefore, the ideal situation is when the RIL parameter is at its highest level, which in this experiment, is measured as RIL = 65%.

5.3. Diversions in Human Perception

Since performance analysis in terms of security and robustness of the proposed VICAP model have been measured for both human users and computer recognition programs, here we would like to demonstrate whether there are any exceptional cases that would affect the human recognition rate rather than computer recognition.

In order to clarify this issue, we have interpreted the definition of "diversion in human perception" in two scenarios. In the first scenario, we have compared the accuracy of the human perception against computer recognition programs and thus, we have presented some samples from our passed experiment results.

Since our experimental results confirm, human visual system is responding to the proposed CAPTCHA model more accurately than current computer recognition programs. As a result, a complete recognition success rate for human users starts from ORO = 40% which will give Human Recognition Success Rate (HRSR) = 100%. Nevertheless, a complete recognition success rate for the computer program starts at the rate of ORO = 70% which will give CRSR = 100%. Since there is such a big gap (around 30%) between the ratio of ORO in HRSR and CRSR, therefore, it can be concluded that the worst recognition score rate for humans was still far better than any computer recognition software.

Since in total over 200 different experiment have been conducted for Computer Recognition Success Rate (CRSR) and over 150 different human experiments have been conducted for Human Recognition Success Rate (HRSR). Table 9 is representing the worst and best case scenario for each CRSR and HRSR for the different ORO ratios, where, 0% represent the worst and 100% represents the best recognition rate. As it can be observed form Table 9, the is nor such a case that the worst case scenario for HRSR is lower than worst case scenario for CRSR and always HRSR was either equal or higher than CRSR. That means the humans recognition rate was always performing better than computer recognition programs.

Table 9. The table represent the best and the worst recognition success rate for both human users and computer programs for different scopes of ORO parameter.

ORO		Worst Recognition Rate (%)	Best Recognition Rate (%)
10%	CRSR	0	0
	HRSR	0	0
15%	CRSR	0	0
	HRSR	0	0
20%	CRSR	0	0
	HRSR	0	100
25%	CRSR	0	0
	HRSR	0	100
30%	CRSR	0	100
	HRSR	0	100
35%	CRSR	0	100
	HRSR	0	100
40%	CRSR	0	100
	HRSR	100	100
45%	CRSR	0	100
	HRSR	100	100
50%	CRSR	0	100
	HRSR	100	100
55%	CRSR	0	100
	HRSR	100	100
60%	CRSR	0	100
	HRSR	100	100
65%	CRSR	0	100
	HRSR	100	100
70%	CRSR	100	100
	HRSR	100	100

The second scenario, would be the case that human perception would be distracted due to the sudden change in the user environment. For instance, somebody suddenly walks into the room or the telephone starts to ring and etc. that requires attention of the end user. Since our proposed CAPTCHA model is based on persistence of vision and superimposing if information using iconic memory, it requires a short attention on the sequence. Therefore, if the user's attention is distracted even for a very short period of time, the effect of persistence of vison would not make affect and therefore, the final image would not make appear on the human's perceptual system.

Since, in our experiments we have assumed the equal environmental conditions for both human users and computer attackers which in this case human users are not affected by any kind of interruptions during the test. However, the effect of these "diversions" would not affect the accuracy of the test, but it would only affect the time taking to solve the CAPTCHA. Since, in our proposed CAPTCHA model, the "original frames" are mixing randomly with "random frames" therefore, by disrupting the user's attention the accuracy of the answer will not be affected and the user requires to focus again on the sequence in order to perceive the sufficient amount of visual information to realise the answer.

Though, this process requires more time from the user in order to recognise the final answer to the test regardless of the quality of the response. In terms of computer recognition, it is the same condition and because the VICAP sequence generator is based on random function, therefore, the recognition process starts from the time that individual frames are being processed regardless of how long it would take to break the CAPTCHA.

5.4. Text-Recognition Algorithm working on Visual Traces

Another important security issues with most of the text-based CAPTCHAs are the dictionary-based attacks or guessing attacks. These types of CAPTCHA attacks are based on recognizing of partial information from a string of characters and based on the preceding and succeeding information, the recognition software would be able to predict the missing elements. Finally, by mapping and comparing the deciphered information against a dictionary database, the program would be able to retrieve decipher the unrecognized word or elements of string.

Since in the proposed CAPTCHA model, the order of the frames is randomly selected and also the way each individual frame is rendered and presented to the user is also random, therefore it is impossible for any text recognition program to predict the future location of the pixels corresponding to the original character. However, in the case of any sequence follows a regular or uniform behaviour, then, it would be possible to track the changes and base on that information it would be possible to predict the final object. For instance, by looking at the example below we can realise that in the first example the symbol "**?**" can be declared as letter "C" based on the preceding and succeeding information.
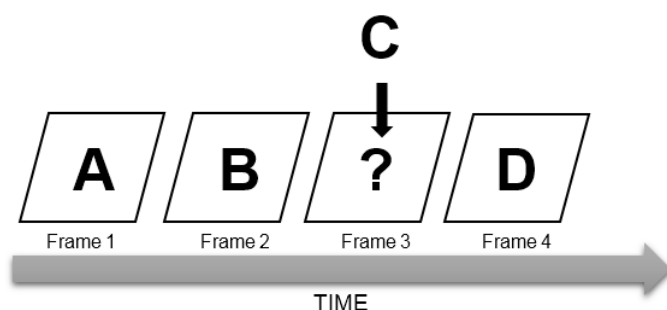
However, in the second example below, it is impossible to have a guess about the value of symbol "**?**". This is because of fact that the presentation of object pixels and background noise pixels in all the preceding and succeeding frames are selected randomly, in the context of a single frame. The lack of uniform and regular patterns in the frame information prevents any predictive algorithms to predict the future position of pixels and as such decipher the original character.
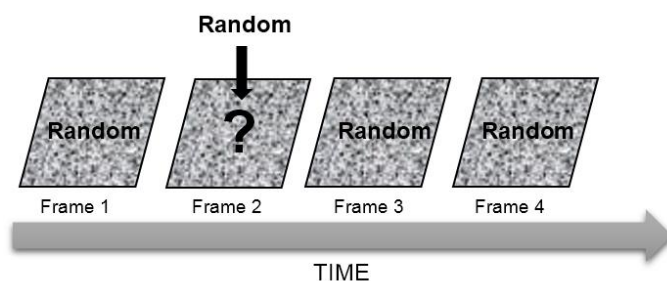
On the other hand, since our proposed CAPTCHA model is not dictionary-based and every individual elements of the CAPTCHA string is selected randomly, therefore, even by decoding one or two elements it would not be possible for the computer recognition programs to guess the entire CAPTCHA string.

6. Conclusions

In this paper, a novel CAPTCHA model called VICAP was introduced in order to increase the level of security and robustness of websites and online service providers. The VICAP model is based on the ability of the human eye to superimpose and build a final image from a sequence of images. Since our proposed CAPTCHA model is designed to work in cooperation with the sophisticated properties of the human visual system, it is believed that none of the current computer recognition programs would be able to break this CAPTCHA model. Consequently, the VICAP model is secure and robust against different computer-based attacks.

The resilience of the proposed CAPTCHA model has been tested and analyzed on both human users and some of the current and most powerful character recognition programs. Using the state of the art CAPTCHA Breaker Application and conducting over 700 experiments on both human users and computer-based attacks, we have achieved a good level of knowledge in terms of security and usability of our proposed CAPTCHA model. As a result of this research, character recognition success rates for human users compared to computer-based recognition programs would ideally increase by 65%.

REFERENCES

Ahn, L. V., Blum, M., Hopper, N. J. & Langford, J., 2003. *CAPTCHA: Using Hard AI Problems For Security.* s.l., Springer-Verlag.

Ahn, L. v. B. M. H. N. a. L. J., 2000-2003. *The CAPTCHA.* [Online]
Available at: http://www.captcha.net/captchas/gimpy/
[Accessed 31 January 2013].

Anon., 2014. *Crimeware: Bots.* [Online]
Available at: http://uk.norton.com/cybercrime-bots
[Accessed July 2014].

Anon., 2016. *Captcha Sniper.* [Online]
Available at: http://www.captchasniper.com/new/index.html
[Accessed July 2016].

Carnegie Mellon University, 2000-2010. *CAPTCHA: Telling Humans and Computers Apart Automatically.* [Online]
Available at: http://www.captcha.net/
[Accessed 2014].

Chellapilla, K. a. L. K. a. S. P. a. C. M., 2005. *Designing Human Friendly Human Interaction Proofs.* Portland, Oregon, USA, ACM, pp. 711-720.

Chew, M. a. B. H. S., 2003. *BaffleText: a Human Interactive Proof.* s.l., s.n.

Clause, C., 2003- 2014. *Iconic Memory: Definition, Examples & Quiz.* [Online]
Available at: http://education-portal.com/academy/lesson/iconic-memory-definition-examples-quiz.html#lesson
[Accessed October 2014].

D. E. Irwin, 1991. Information Integration across Saccadic Eye Movements. *COGNITIVE PSYCHOLOGY,* Volume 23, pp. 420-456.

David E. Irwin & Laura E. Thomas, 2008. Visual Sensory Memory. In: A. H. Steven J. Luck, ed. *Visual Memory.* s.l.:Oxford University Press, pp. 9-42.

Greg Mori, J. M., 2003. *Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA.* s.l., s.n.

GSA, 2014. *GSA Captcha Breaker.* [Online] Available at: http://captcha-breaker.gsa-online.de/ [Accessed December 2015].

GSA, 2016. *GSA-Software development and Analytics.* [Online] Available at: https://captcha-breaker.gsa-online.de/ [Accessed July 2016].

Haichang Gao, et al., 2010. A Novel Image Based CAPTCHA Using Jigsaw Puzzle. In: *Computational Science and Engineering (CSE), 2010 IEEE 13th International Conference on.* s.l.:s.n., pp. 351-356.

Irwin, D. E., Jun., 1996. Integrating Information across Saccadic Eye Movements. *Current Directions in Psychological Science,* 5(3), pp. 94-100.

J. H. J. S. Jeremy Elson, J. R. D., 2007. Asirra: a captcha that exploits interest-aligned manual image categorization. In: *In ACM Conference on Computer and Communications Security.* s.l.:s.n., p. 366–374.

Jeff Yan, A. S. E. A., 2008. *A Low-cost Attack on a Microsoft CAPTCHA.* s.l., s.n.

Jeff Yan, A. S. E. A., 2008. *A Low-cost Attack on a Microsoft CAPTCHA.* s.l., s.n.

Jeff Yan, A. S. E. A., 2009. CAPTCHA Security: A Case Study. *IEEE Security & Privacy,* Volume 7, pp. 22-28.

Jeff Yan, A. S. E. A., 2009. *CAPTCHA Security: A Case Study.* s.l., Security & Privacy, IEEE.

Luis von Ahn, Manuel Blum & John Langford, 2004. Telling Humans and Computers Apart Automatically.. *Communications Of The ACM,* February, Volume 47, No. 2, pp. 57-60.

Luis von Ahn, M. B. J. L., 2004. Telling Humans and Computers Apart Automatically.. *Communications Of The ACM,* February, Volume 47, pp. 57-60.

Luis von Ahn, M. B. N. H. a. J. L., 2010. *CAPTCHA: Telling Humans and Computers Apart Automatically.* [Online] Available at: http://www.captcha.net/

Marc, 2011. *CAPTCHA The Moment.* [Online] Available at: http://mkcohen.com/2011/03 [Accessed June 2015].

McKinney, M., 2008. *The Persistence of Vision.* [Online] Available at: http://www.vision.org/visionmedia/article.aspx?id=136 [Accessed October 2014].

Moradi, M. & Keyvanpour, M., 2015. CAPTCHA and its Alternatives: A Review. *Security and Communication Networks,* 8(12).

Nicomsoft, 2012. *Optical Character Recognition (OCR) – How it works.* [Online] Available at: http://www.nicomsoft.com/optical-character-recognition-ocr-how-it-works/ [Accessed 24 September 2014].

Rayner, K., 1998. Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin,* Volume 124, pp. 372-422.

Saadat Beheshti, S. & Liatsis, P., 2015. *CAPTCHA Usability and Performance; How to Measure the Usability Level of Human Interactive Applications Quantitatively and Qualitatively?.* Dubai, 8th International Conference Developments in eSystems Engineering (DeSE 2015).

Saadat Beheshti, S. & Liatsis, P., 2015. *VICAP: Using the mechanisms of trans-saccadic memory to distinguish between humans and machines.* London, 2015 International Conference on Systems, Signals and Image Processing (IWSSIP).

Shirali-Shahreza, M. & Shirali-Shahreza, S., 2007. Collage CAPTCHA. In: *Signal Processing and Its Applications, 2007. ISSPA 2007. 9th International Symposium on.* s.l.:s.n., pp. 1-4.

Steven J. Luck, A. H., September 2008. Visual Sensory Memory. In: s.l.:Oxford Scholarship Online.

Te-En Wei, Jeng, A.B. & Hahn-Ming Lee, 2012. *GeoCAPTCHA - A novel personalized CAPTCHA using geographic concept to defend against 3rd Party Human Attack.* Taiwan, s.n., pp. 392-399.

Vu Duc Nguyen, Y.-W. C. W. S., 2014. On the security of text-based 3D CAPTCHAs. *Computer & Security,* Volume 45, pp. 84-99.

Woodford, C., 2013. *Optical character recognition (OCR).* [Online] Available at: http://www.explainthatstuff.com/how-ocr-works.html [Accessed 24 September 2013].

Yan, J. a. E. A. A., 2007. *Breaking Visual CAPTCHAs with Naive Pattern Recognition Algorithms.* s.l., Computer Security Applications Conference, 2007. ACSAC 2007. Twenty-Third Annual.

Yannis Soupionis, D., 2010. Audio CAPTCHA:Existingsolutionsassessmentand. *computers & security,* pp. 603-618.