

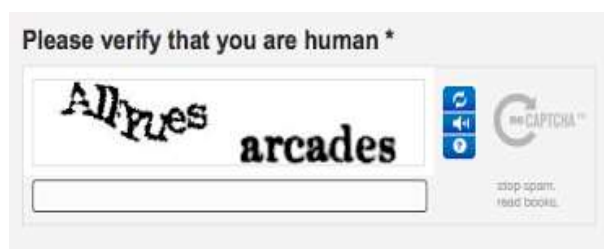
**ABSTRACT**

The science of making computers humanly to an extent in which it is capable of decision making on its own, visual cognizance, speech recognition, etc. is artificial intelligence. In a very less time span, this field has impacted several sectors of life such as medical, engineering, business, forecasting, etc. "Completely Automated Public Turing Test to Tell Computers and Humans Apart" acronym to CAPTCHA is a test to differentiate humans from normal robots. A robot is not capable of cracking a CAPTCHA. In this paper we are focusing on how an artificial intelligence robot can solve a CAPTCHA. It is not a big deal to crack a CAPTCHA but it is a small step to a big thing i.e. consciousness in A

**KEYWORDS:** Artificial Intelligence, consciousness, neural networks, CAPTCHA, fussy logics.

**I. INTRODUCTION**

Completely Automated Public Turing Test to Tell Computers and Humans Apart (CAPTCHA) came into existence in 1997. CAPTCHA is a collection of distorted or randomly stretched characters or numbers sometimes with a noisy image in the background which user is supposed to type in a box mentioned. It is also known as Reverse Turing test as it is a test supervised by a computer. CAPTCHA is a Turing test which is fully automated which requires minimal human attention and involvement to administer, which is also cost efficient and reliable. CAPTCHA's are mainly used by websites which has online polls and registration forms involved. EX: Gmail, Yahoo, etc.



*Image 1 (captcha in use)*

There are basically two types of CAPTCHA, namely: (1) Audio CAPTCHA (2) visual CAPTCHA. Audio CAPTCHA is a type of visual CAPTCHA which also reads the characters out loud to help the physically disabled person or in case the character is hard to recognize.

CAPTCHA has many applications in security, few of which are:

**a. Email services**

Email service providers like Microsoft, rediffmail, etc. who provide free email services have been facing bots attack almost every second. To overcome this problem CAPTCHA test is imposed.

**b. Spams**

CAPTCHA allows mails from computers who have human at the back end. Hence it keeps a check on spams.

### c. Prevent password attacks

CAPTCHA's does not allow computer's to iterate through the entire space of password making human to do it themselves.

## II. METHODOLOGY

A person who wishes to crack CAPTCHA should approach this problem in steps. He should start with writing an algorithm – a self contained set of steps of operations to be performed, to get the desired results. First step in this problem must be to get the image in grayscale i.e. to remove all the colors from the image, taking away one of the difficulty level CAPTCHS employs.

Next step of algorithm should be to detect patterns in the image so as to compare them with standard letters, to find a match. If the program could only match a few letters, it cross checks them with a database of English words. Then result will be plugged into submit field. It may not have 100% accuracy but it is still efficient enough.

A question might strike, what if it is a more complex CAPTCHA! Like for ex. Gimpy CAPTCHA. A gimpy CAPTCHA uses 8-10 words taken from a dictionary, in warped, corrupted and distorted form in an image and user is supposed to type them. It is one of the most reliable system and is presently in use by Yahoo. Here the words are arranged in pairs which then overlap one another. A user needs to get 3 correct words to go further. One thing is for sure; with correct algorithm no CAPTCHA can be reliable. The gimpy CAPTCHA was also cracked by Mori and Malik. Although they had an accuracy of just 33% but still it is a significant number in A.I.

## III. IMPLEMENTING ALGORITHM

In this section we will be discussing different steps that are involved to crack a CAPTCHA in detail. The steps are as follows:

- a) Preprocessing
- b) Segmentation
- c) Feature extraction
- d) Character recognition
- e) Post processing

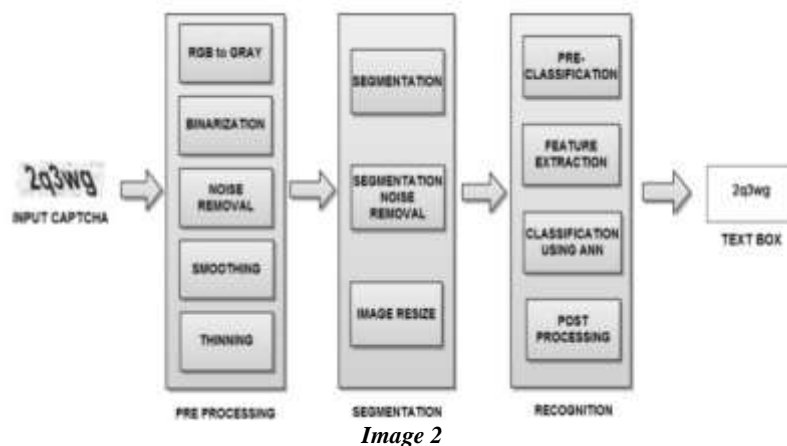


Image 2

### 1. Preprocessing

In this very first step, we clear out the image by converting it into gray scale, then into binary matrix of black and white, followed by removal of horizontal and vertical lines and dots, if any. The only disadvantage here is that pixel density is lost.

#### Gray Scaling

bmp format image is taken as input which is then converted to gray scale image. This is done so as to simplify the work as different colors have a big range of intensities which results in more complexity.

### ***Binarization***

Image displayed in only two columns black and white are treated as binary images. They have two intensity values equal to 0 for black and 1 or 255 for white. This is done to separate an object color from background color. Captcha images have noise which includes horizontal and vertical lines, dots which are also needed to be cleaned.

### ***Line Cleansing***

Sometimes horizontal and vertical lines are also added to CAPTCHA so as to add to its difficulty level. In such cases number of continuous black pixels in row/column is taken into account where if count exceeds more than 80% of total height or width, then it is considered as a line and hence cleaned that is make it white.

### ***Dot Cleaning***

Set of black pixels is dots. Sometimes after binarization certain amount of black pixels can be present. Scanning followed by counting of pixels is done for its cleansing. Suppose on scanning you get a black dot, its next 8 pixels are checked. If white then make the black dot white but if black dots are greater than one pixel but less than 40 then again make it white as no letter can have 40 or less pixels hence it comes in the category of unwanted and therefore should be removed.

## **2. Segmentation**

After processing of image, segmentation is performed because it is difficult for a computer to read joint character. There are several techniques for this

### (a) Color filling segmentation

Here paint bucket flood filling algorithm is used which requires 3 parameters – a start node, a target color and replacement color. Queue or stack is used to structure this algorithm this technique is used to segment tilted letters but fails on contiguous characters.

### (b) By checking continues black pixels & separate them as character. Once black characters are seen, programs mark them red so it can understand later that this character is already reviewed.

## **3. Feature Extraction**

As we know every character has a unique feature so the features taken into consideration are

### (1) Holes

Character such as a,b,d,o,p,g,q have holes in them while rest do not have any.

### (2) Character height

a,o,c are categorized in small letters whereas b,d,q are taken in large & hence a threshold categories are characters

### (3) Interactions (vertically)

A line is drawn such that it passes through the character & most possible white black transition is noted down.

Ex: 3 for a.

## **4. Recogniton Of Character**

There are two modes to this step:

- Training mode
- Testing mode

In training mode classifier is taught what and how each character looks like after segmentation process. In testing mode a predictive model is created to recognize characters.

## **5. Post Processing**

In this minor improvements such as spell check are done. It is only done if it is known that CAPTCHA uses dictionary words. This step increases precision by 20% - 30 % in a single go.



Image 3

#### IV. RESULT

Two things are mainly considered for the efficiency of classification algorithm namely character accuracy and CAPTCHA accuracy. If the character is guessed correctly, it improves character accuracy, otherwise reduces. A testing set 60 images to check the correctness of an algorithm and above are the results captured.

Here CAPTCHA accuracy is percentage of CAPTCHA to which it was correctly recognized.

#### V. CONCLUSION

In this paper we explored methods of solving the problems of identifying words. In some cases we came across clutters involving other real objects which make the problem even more complicated but with our algorithm we were successfully able to crack it. Algorithm used here one instance of general framework that is reusable in different cases.

#### VI. REFERENCES

- [1] Algorithm to break visual CAPTCHA (research paper ICETET).
- [2] Machine Consciousness by O Holland.
- [3] An Introduction to Neural Computing by IAleksander and H Morton.
- [4] Assessing Artificial Consciousness I Aleksander, S Stuart, T Ziemke.
- [5] Artificial Intelligence: Structures and Strategies for Complex Problem solving (5th edition). Benjamin Cummings.
- [6] Flash Flood fills Implementation by Emanuele Feronato.
- [7] Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA by Greg Mori and Jitendra Malik.

#### CITE AN ARTICLE

**Kher, V., Yadav, M., Naramdeo, S., Deolase, V., & Gupta, S. (2017). BREAKING CAPTCHA USING ARTIFICIAL INTELLIGENCE. INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, 6(9), 342-345. Retrieved September 15, 2017.**