Engagement Identification Using a Dynamic Pattern

OU ID: 14028

Overview

CAPTCHAs ("Completely Automated Public Turing Test to tell Computers and Humans Apart") were first designed to allow the determination of whether or not a user is human using a challenge-response test. We encounter CAPTCHAs on a regular basis, especially when setting up new user accounts or posting comments on a webpage. These tests are put in place to keep spammers from using automated software to set up thousands of user accounts or post comments instantaneously. The most common form of CAPTCHA makes a user type in letters or numbers in a distorted image, the idea being that humans will easily recognize distorted text on an image while computers would not. While CAPTCHAs worked fine when they first appeared in 2000, computing power and OCR programs (both general-purpose ones and those designed specifically to defeat CAPTCHAs) have grown highly sophisticated. Advancements in computing capabilities have led to an arms race where CAPTCHAs continue to become more and more challenging in order to defeat computerized responses. As a consequence, we are now faced with nearly impossible CAPTCHAs.

What is the Problem with CAPTCHAs?

These days, CAPTCHA challenges have become so complex that it takes the average user three or more attempts, which can take over a minute to complete. We have a big problem as even CAPTCHAs that push the limits of human cognition have now been defeated by computer software. Simply, a computer can conduct billions of scanning iterations to extract information from an image within a millisecond, all the way down to the level of the individual pixel. Making matters worse, there are only 26 possible alphabets and 10 possible numbers that have to be detected.

Researchers' Solution

To solve this problem, we have created a dynamic CAPTCHA that leverages the human ability to extract patterns in moving images. Humans need the capacity to perceive dynamics in order to get through the day. Whether it involves kicking a ball or catching one, driving, even sitting down and watching TV requires that we be able to extract continuity and motion patterns from our visual environment, while ignoring unnecessary sources information. For example, the moving images we see on the television are really a series of static frames. Yet, our brains tell us that something is moving on the screen, to the extent that we are able to interact with those images when we play video games the same way we would in the real world.

OHIO UNIVERSITY
1804

Engagement Identification Using a Dynamic Pattern

OU ID: 14028

Biology-Centric Computing

The CAPTCHA system is designed to take advantage of the natural properties of human visual perception, its capacity to aggregate information and extract "gestalt" wholeness from dynamic information. Below you will find two examples of our Dynamic CAPTCHA challenges. First is the letter 'Z' and the second is the shape of a heart. Even though the shapes are being rotated every few seconds, it is still relatively easy to pick out, despite the presence of the background scatter of dots different colors and sizes. This is because the human brain creates continuity by compiling information from many frames.

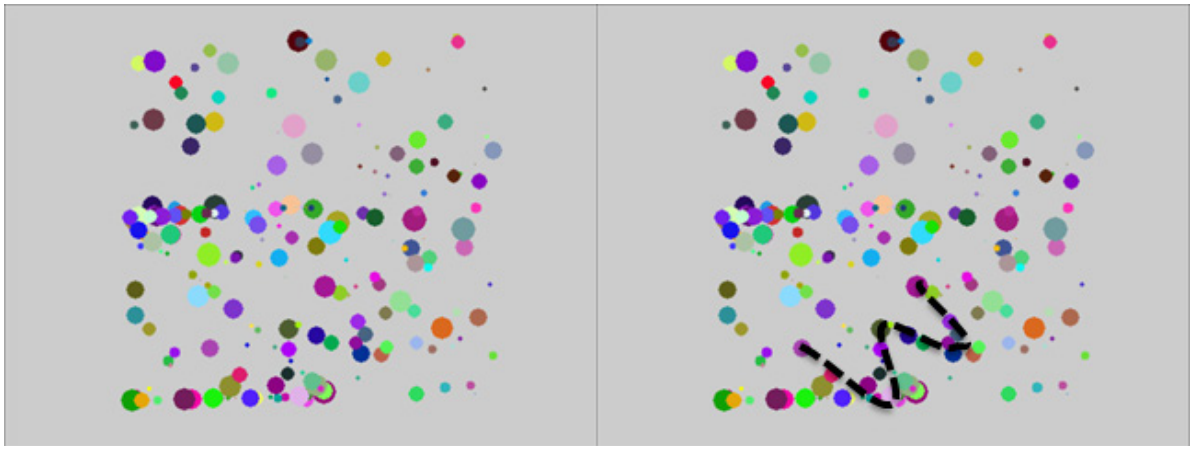

Fig. 1: The left panel is a snapshot from the video containing the letter "Z." Based on the clustering of dots, one would actually guess that the number "3" or letter "M" or "W" is present, but that is incorrect.

Unlike the human eye, the computer sees the same way that it displays images, frame-by-frame. We take advantage of this weakness by breaking up the target image and displaying it in a piecemeal form so that the whole target is never presented in its entirety on a single frame. The actual image is randomly broken up and distributed over a number of frames. The result is a frame such as the one above, taken from the video for the letter Z (Fig. 1). Simply trying to compute the pattern from a single frame is difficult as the random patterns mean that by chance, different meaningful images might form during a single snapshot, images that disappear once the next frame is generated.

Engagement Identification Using a Dynamic Pattern

OU ID: 14028

The random scatter also prevents the accumulation of multiple frames, as the non-target dots will begin to overwhelm the underlying pattern (Fig. 2). While the letter 'Z' is still slightly visible here, it is only because we know that this should be the underlying pattern and that straight line segments need to be detected. Text based CAPTCHAs are limited in the fact that there are only 26 alphabets and 10 numbers, whose orientations have to be maintained in order for them to make sense. These provide a set of rules for the OCR, simplifying the recognition process significantly. But, because the shape (curves vs. straight lines) and orientation (backwards, forwards, upside down) are unknown, detecting the underlying shape in a heavily noisy background is extremely difficult. The randomized rotation and resizing of the target image as well as the background provides an additional benefit in that a single CAPTCHA image can be re-used over and over again without issues.

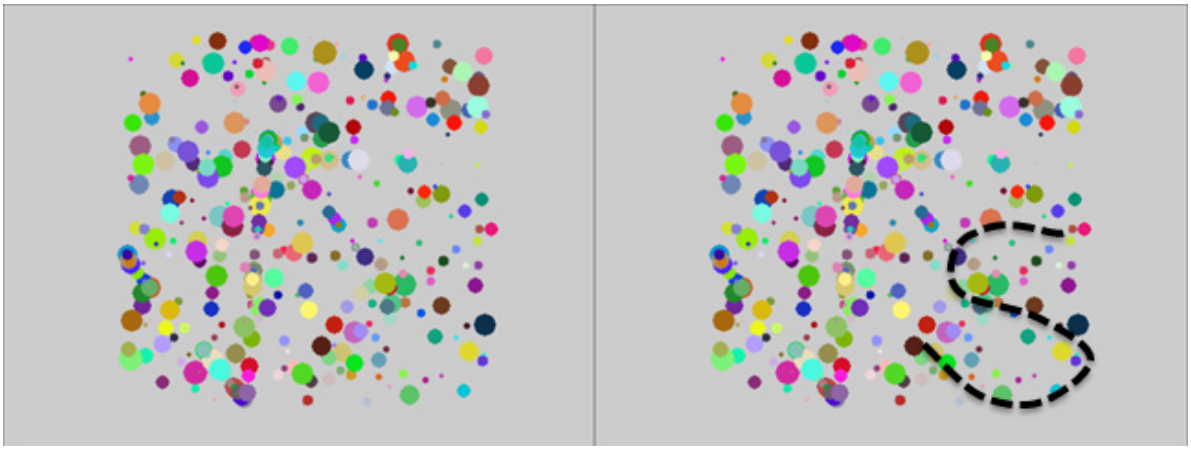

Fig. 2: Needle in a haystack. Above you will see two overlaid frames. While the letter "Z" might appear to be present in the image, the letter "S" is also a possibility.

Engagement Identification Using a Dynamic Pattern

OU ID: 14028

Gesture-Based Input

Our system requires a gesture-based input in response to the CAPTCHA challenge (Fig. 3). The user simply needs to move their finger and trace the target image based on the video movie being displayed on the device. Once the gesture is completed (indicated by the removal of contact with the touchscreen), the coordinates of the gesture via contact with the touchscreen are then transmitted to a central server that can determine if the response is correct. This makes the process extremely intuitive and convenient, where even if the user waits for the image to rotate a few times, it will take merely seconds vs. the conventional CAPTCHA. More importantly, users who are slow typists will no longer be slowed down even more. This technology is particularly suitable for mobile devices that do not come with a physical keyboard. Many of these devices come with a display too small for even a virtual keyboard and would make displaying traditional text-based CAPTCHAs virtually impossible.

Will This Really Work?

The answer is simple. Even the largest and most complex computing system in the world, IBM's WATSON, failed to complete two tasks while playing Jeopardy! First, WATSON did not field any video clues. Second, WATSON was not able to physically push a button to buzz in for an answer. Our CAPTCHA system requires: 1) the ability to decode an image within a video; and 2) generation of a physical response through a gesture.

Fig. 3: Gesture-based response to the dynamic CAPTCHA.

OHIO UNIVERSITY

Engagement Identification Using a Dynamic Pattern

OU ID: 14028

Are There Any Additional Safeguards?

Continued improvements in computing capacity and technology are always to be expected. To provide our new system with greater longevity than the original CAPTCHAs we have added the following safeguards:

(1) Computational Power Awareness

Our system will evaluate the amount of computational power available on any particular device and demand a certain percentage be used in encoding and dynamically changing the "challenge" as a gesture pattern. The percentage can be 50% or more on a device that is known to have a light computational load. The percentage can be lowered if the device if known to be computationally busy. If the required amount of computational power is not allocated, no "response" will be accepted. This threshold is dynamically adjustable.

The more computational power devoted to encoding the "challenge", the more difficult it will be to decode it, and in the same time the less computational power is left for potential hackers to use to decode it and come up with the correct "response." This way, even if in the future, someone invents an algorithm that defeats the gesture-based graphical CAPTCHA, which by itself is still an impossible with today's technology, there will be an insufficient amount of computational power left on a particular computing device for the algorithm to work.

(2) Spoofing and Replay Protection

By forcing the response to be generated using a gesture, we can safeguard against the possibility that an OCR is able to recognize the target image. Let us assume that the hacker now has the capacity to extract the necessary coordinates that are then relayed to the system, bypassing the touchscreen entirely. Again, we leverage the biological properties of the human in order to complicate the spoofing process.

First, we know that the behavior will take time. At the minimum, it will take approximately one second to complete the whole process (about 250 milliseconds to recognize the image and 500 milliseconds to complete the gesture trace). Any response under one second will be rejected. Second, no human will be able to perfectly replicate the target image exactly. Third, there will be dynamics in the behavior, and there will be modulations in speed. A one second delay is unacceptable for large-scale spammers, and additional computation needs to be devoted to generating the necessary artificial behaviors, i.e., velocity modulation and error.

Contact Us

Ohio University Technology Transfer Office
340 W State Street
Athens, OH 45701
740.593.0976
techinfo@ohio.edu
https://www.facebook.com/OUTechTransfer

OHIO UNIVERSITY