

# BREAKING TICKETMASTER'S VISUAL



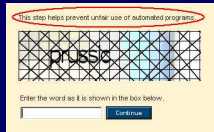
Maksim Rapoport

Yuriy Vasilyev

Advisor: Dr. C. J. Taylor

## Abstract:

Implementing and developing techniques for recognizing text in adversarial clutter  
i.e., Using a computer to defeat TicketMaster's Completely Automated Public Turing-test to tell Computers and Humans Apart, (CAPTCHA).



## Removing Lines:

- Run over the edges detecting likely endpoints.
- Remove horizontal and vertical lines.



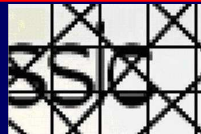
- Remove the cores of the central lines. This allows us to keep the majority of the letters' pixels.



- Remove the lines edges, based on the cores, and find a bounding box.

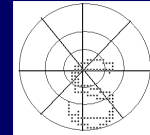
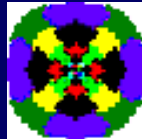


- Observe that our method avoids removing the letter 'i', which falls directly under a vertical line.



## Recognition:

The possible windows containing characters are determined by finding local minimums.  
A pre-created template context is used to characterize each point in the image.



## Formulating guesses:

- A context is computed for each point in the guess-window.
- Each point is matched to a point in three images that all together contain all the letters of the alphabet.
- One guess is made from each image for each slide, forming a guess array.

Image 1:	m	e	v	g	j	n	f	x	
Image 2:	a	k	c	y	i	p	q	d	t
Image 3:	r	u	l	w	s	b	o	h	z

## Dictionary Attack:

- Each word in TicketMaster's CAPTCHA is 5-7 letters long. Using this we use a 65,000 word dictionary to look up possible words based on the guess array.

The guess array for prussic: ['dpo', '5ir', '55u', '55s', 'e5s', 'jil', 'dco']

The guesses after using the dictionary: "Drused", "Prussic"

## Results:

The following table summarizes recognition results when we ran our program on 32 images of the same font as our alphabet. Whenever several words were returned as possible guesses, one was selected at random as the final guess.

Single, correct word identified	8
No correct word found	15
Correct word was one of the several dictionary guesses	9
Ratio of found/total using random guess when several choices are available	<b>36.5%</b>

## Conclusion:

By achieving a significant (36.5 %) recognition rate on a single font we have demonstrated that current state-of-the-art visual CAPTCHAs are not reliable. We believe that with slight modifications and addition of extra fonts our method will be more than capable of defeating Ticketmaster's Bot-prevention techniques.

We believe that the concept of Visual CAPTCHAs is unreliable in principle. We urge commercial websites to reconsider the use of visual CAPTCHAs, and to switch to more reliable human identification protocols.

## References:

1. TicketMaster!@ CAPTCHA (EZ-Gimpy). Source: TicketMaster!@Retrieved April 13, 2005.
2. Mori, Greg. "Re: questions about pattern recognition" E-mail to Maksim Rapoport, September 22, 2004.
3. Matt May, "Inaccessibility of Visually-Oriented Anti-Robot Tests", location: <http://www.w3.org/TR/turingtest/>
4. S. Belongie, J. Malik, J. Puzicha. "Shape context: A new descriptor for shape matching and object recognition". In NIPS, November 2000.

