COMS 6998 – Computational Photography – Spring 2009

BREAKING AN IMAGE BASED CAPTCHA

PROJECT PROPOSAL

Michele Merler(mm3233)

Jacquilene Jacob (jj2442)

INTRODUCTION

A CAPTCHA is "Completely Automated Public Turing test to tell Computers and Humans Apart". An image of distorted letters is dynamically generated. Since the letters are a part of an image and not text, it is difficult for a spam bot or other computer program to read. A human, in fact, has little trouble reading the letters in a captcha image. Using a captcha test on a website is a great way to ensure, for instance, that a person and not a spambot is filling out a web form.

For example, humans can read distorted text as the one shown in Figure 1, but current computer programs cannot.

Captcha is a so called win-win solution, in that if a bot cannot break it, it provides security, but if it is automatically broken, that means that a difficult task in computer vision or related areas has been solved.

The problem with current visual text based captcha systems is that most of them have proven to be either not robust enough (they have been broken) or they are too complicated or annoying to read even for humans.

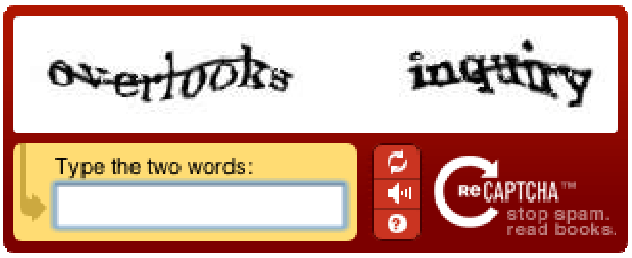

Figure 1 : Example of text based captcha. Source reCaptcha[1]

VidoopCAPTCHA

VidoopCAPTCHA[2] is a verification solution that uses images of objects, animals, people or landscapes, instead of distorted text, to distinguish a human from a computer program. By verifying that users are human, the site and users are protected against malicious bot attacks. VidoopCAPTCHA is more intuitive for the user compared to the more traditional text based CAPTCHAs. It then presents itself as the solution to the current captcha problems.

---

[1] http://recaptcha.net/

[2] http://vidoop.com/captcha/

As shown in Figure 2, a Vidoop challenge image consists in a combination of pictures representing different categories. Each picture is associated with a letter which is embedded in it. In order to pass the challenge, the user is asked to report the letters corresponding to a list of required categories. The robustness of the approach relies in the fact that object recognition is a straightforward and fast to solve task for humans, whereas for a computer it is a fundamentally hard problem. In fact, it has represented for many years and still represents a topic of active research in computer vision. What the authors underestimate, though, is that since a bot can try to access a service thousands of times in a day, recognition rates which are considered quite low by the object recognition community (50% or 60%), still would allow automatic attacks to services protected by the image captcha to be considered fully successful.

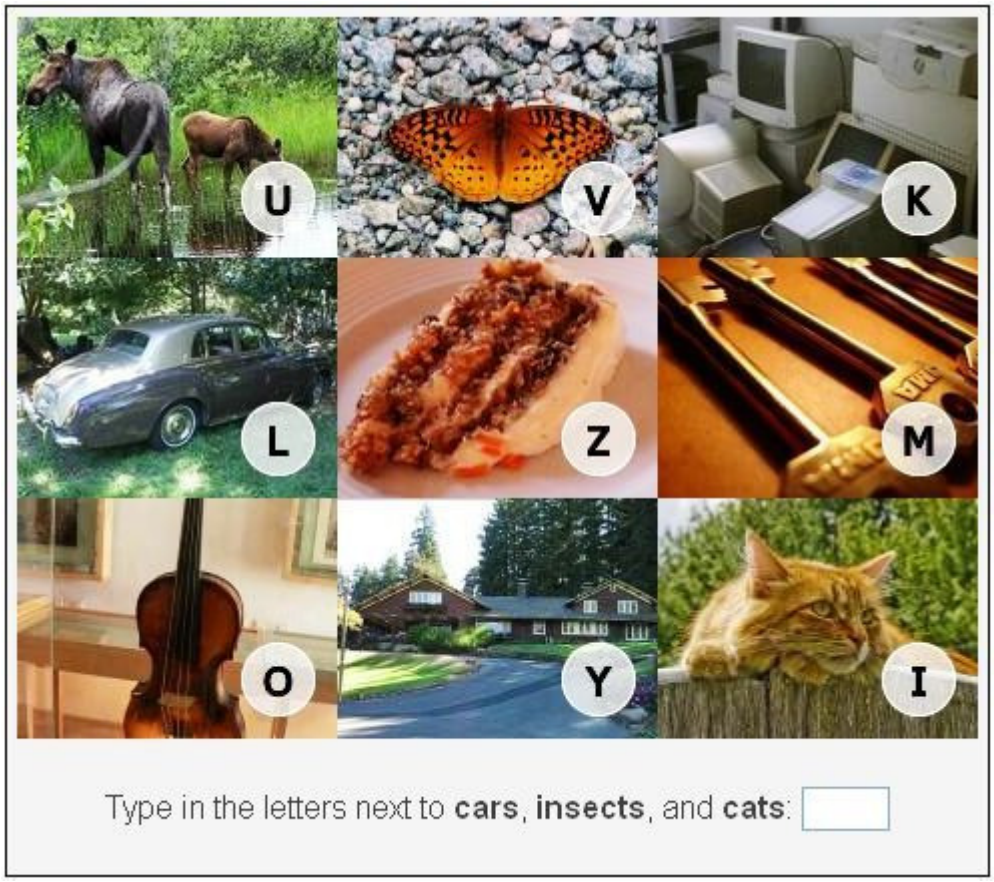

Figure 2 : Example of an image challenge from VidoopCatcha

OBJECTIVE

The core idea of the project consists in trying to break an image based captcha, and in particular VidoopCAPTCHA, following in the line of work initiated by Mori and Malik[3].

The objective of this system is to show that image-based captchas, and in particular the vidoop one, are not as secure as their authors claim. This automatically leads to insecurity for the different applications using the image captchas. We chose this idea in order to show our concerns in today's world where the security methods developed to preserve confidentiality in online systems, of which image based captcha represent of the latest developments, are not only insecure but are prone to attacks by hackers with high success rates.

ALGORITHM

The proposed algorithm consists in the following parts:

- Isolate the single pictures within the challenge image and extract their corresponding characters regions with some simple image processing (line detection in edge images, circle detection or color based methods).
- Classify the images according to the categories required by the test. The classifier could be an SVM (or even a simple K-NN) trained on *n* images downloaded from Flickr or other online image sources, and queried using the keyword provided from the test. We have not thought about what features to use to represent the images yet, but we could start with simple ones such as color histograms and then evaluate their performances.
- Extract the characters corresponding to the images classified as being part of the required categories. This part should be relatively simple, given that they are single characters and all capital letters. SVM trained on more challenging data, such as MNIST[4] database of handwritten characters, reach over 99% accuracy.

---

[3] http://www.cs.sfu.ca/~mori/research/gimpy/

[4] http://yann.lecun.com/exdb/mnist/

- Insert the letters and verify if the test was passed.

MILESTONES

Generally speacking, we plan to divide the project in 2 phases; one offline and one online.

The first consists in creating an offline visual CAPTCHA breaker, by downloading *m* (maybe 100 or 200?) Vidoop image challenge images, and build offline classifiers for their required categories using *n* images downloaded from Flickr (or similar sites). Then test the results on the *m* tests to see if the performances are acceptable.

In the online phase, we will then try to make the system work online, with new tests. We are assuming the number of categories (the tests' taxonomy) Vidoop proposes is limited, so in most of the cases it will be possible to use the classifiers we already built in the offline phase to solve the test. If a test requires a new category, then the system will have to download the images from Flickr and train a new classifier on the fly.

It would be interesting to also implement a baseline online version of the system, with only maybe 3 images as training set for each category and a NN classifier. It would be fast, and even if the accuracy will be low, it might be enough to consider the CAPTCHA to be broken.

We propose the following time map to complete the project, where each date represents the deadline to complete the corresponding task:

- Project Proposal – April 1st
- Offline phase - April 15th :
  - Training data collection : write a Perl or Python script to download images from the web given a keyword
  - Classification: develop preprocessing functions, feature extraction methods and classifiers in Matlab

  - Testing data collection: download around *n* Vidoop challenge images with a Python or Perl script
  - Analysis of results
- Intermediate Milestone Report – April 15th: reporting the results of the offline phase
- Online phase – April 29th :
  - Application of the developed system online on various sites: write an online agent interface that accesses the Vidoop site, reports the challenge image to our system, generates new categories classifiers if necessary, and inserts the proposed answer
  - Analysis of the results
- Final project presentation – April 29th : approach description, report of the offline phase results, possibly present demo of the online system
- Final project report : May 4th: