IMPROVING IMAGE SEARCH WITH PHETCH

Luis von Ahn, Shiry Ginosar, Mihir Kedia, and Manuel Blum

Computer Science Department, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh PA 15213
biglou@cs.cmu.edu, {shiry,majin,mblum}@cmu.edu

ABSTRACT

Online image search engines are hindered by the lack of proper labels for images in their indices. In many cases the labels do not agree with the contents of the image itself, since images are generally indexed by their filename and the surrounding text in a webpage. To overcome this problem we present Phetch, a system for attaching accurate explanatory text captions to arbitrary images on the Web. Phetch is an engaging multiplayer game that entices people to write accurate captions. People play the game because it is fun, and as a side effect we collect valuable information that can be applied towards improving image search engines. In addition, the game can also be used to enhance Web accessibility and to provide other novel applications.

Index Terms— Distributed knowledge acquisition, Accessibility, Web-based games.

1. INTRODUCTION

Current search engines on the Web index images by using textual data such as filenames, image captions and/or adjacent text on the Web page. Unfortunately, such data can be insufficient or even deceptive, making it hard to return accurate search results [5]. In this paper, we address the problem of attaching descriptive captions to images on the Web in order to improve the accuracy of image search.

Rather than attempting to design a computer vision algorithm that generates natural language descriptions for arbitrary images (a feat still far from attainable), we opt for harnessing human brainpower. It is common knowledge that humans have little difficulty describing the contents of images, although they typically would not find this task particularly engaging. On the other hand, many people would spend a considerable amount of time involved in an activity they consider "fun." Therefore, we solve the problem by creating a fun game that produces the data we aim to collect as a side effect of game play.

Our method is similar in spirit to the ESP Game [2] (a.k.a., Google Image Labeler [7]), which encourages users to enter correct labels for images from the Web by turning the process into an enjoyable game. In this paper, we describe how a different game, Phetch, can also be used to improve image search. Phetch was originally introduced in [3] as a possible solution to a major accessibility problem on the Web: the lack of descriptive captions to aid visually impaired users. We now show how Phetch can also be used to improve image search. More specifically, the goals of this paper are:

- **To introduce the concept of "Human Computation" to the signal processing community.** Like the ESP Game, Phetch is an example of a general approach to computational problems, called "Human Computation," where humans are recruited to collectively perform parts of a massive computation [1]. We hope to demonstrate the value of this approach to the signal processing community.
- **To show that Phetch need not rely on the ESP Game.** The original design of Phetch was not well-suited to improve image search, primarily because it was dependent on the output of the ESP Game: an image could not be captioned by Phetch until after it was labeled by the ESP Game [3]. We show how to remove this requirement.
- **To show how Phetch provides additional value beyond previous "human computation" solutions.** In addition to collecting captions for images, the game mechanics of Phetch can be used to further improve image search in many ways. For example, we show that Phetch could perform a service similar to Google Answers [6], where players perform the searches in real time for image search users. In Google Answers, users submit questions to which they cannot find an answer (e.g., "who is the most famous person alive"), and for a monetary fee, other people attempt to find answers for them over a period of several hours or days. Phetch allows a similar service for images except that it would be free and faster: the players would perform the searches as a part of the game.

1.1. Related Work: The ESP Game

The ESP Game [2] is a two-player online game in which players provide meaningful and accurate labels for images on the Web as a side effect of playing. Think of the ESP Game as a slideshow in which players provide the labels. Random images pop up from the Web and players type possible one-word descriptions. If one of their words

matches one typed by their partner, it becomes a label for that image. Among other things, the labels collected by the ESP Game can be used to improve image search. Indeed, the ESP Game has also been implemented online as Google Image Labeler [7].

Although similar in spirit to the ESP Game, Phetch is able to provide additional value. The most obvious difference is that Phetch provides explanatory paragraphs or sentences describing each image rather than just one-word labels (see Figure 1). Even simply mining Phetch data for keywords would yield additional, non-obvious labels; we later show more search-related applications that take advantage of Phetch. Phetch also uses a completely different game mechanic: designed as a multiplayer, competitive game, it can reach an expanded set of players.

Figure 1. Two inherently different images that share the same ESP Game labels: "man" and "woman." The Phetch descriptions are different: "half-man half-woman with black hair" and "an abstract line drawing of a man with a violin and a woman with a flute."

2. GAME MECHANICS

Phetch is designed as an online game played by three to five players. Initially, one of the players is chosen at random as the "Describer" while the others are the "Seekers." The Describer is given an image and helps the Seekers find it by giving a textual description of it. Only the Describer can see the image and communication is one-sided: the Describer can broadcast a description to the Seekers but they cannot communicate back. Given the Describer's paragraph, the Seekers must find the image using an image search engine. The first Seeker to find the image obtains points and becomes the Describer for the next round. The Describer also gains points if the image is found. Intuitively, by observing the Describer's text, we can collect natural language descriptions of arbitrary images. (See Figure 2.)

Each session of the game lasts five minutes, during which time the players go through as many images as they can. The Describer can pass, or opt out, on an image if they believe it is too difficult for the game. In this case, the Describer gets a new image and is penalized by losing a small amount of points.

As mentioned, the Seekers have access to an image search engine, which, given a text query (either a word or a set of words), returns all images related to the query. Once a Seeker believes she has found the right image, she clicks on it. The game server then tells the Seeker whether their guess was correct. The first Seeker to click on the correct image wins and becomes the next Describer. To prevent Seekers from clicking on too many images, each wrong guess carries a strong penalty in points. This penalty also ensures that the text given by the Describer is a reasonably sufficient description of the image, since Seekers will tend to not guess until they are certain.

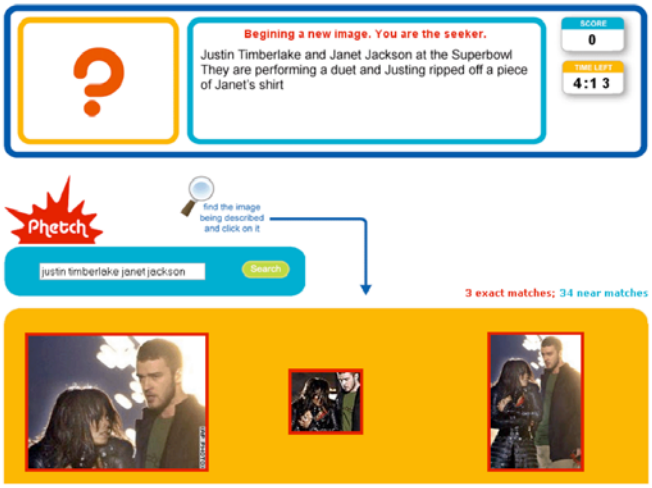

Figure 2. A screenshot of the Seeker's interface.

3. THE IMAGE SEARCH ENGINE

The image search engine given to the Seekers is a crucial component of the game for several reasons. First, the available search space cannot be so large that it requires Seekers to filter through thousands of query results. Second, we must somehow guarantee that the correct image is usually returned given a good query. The original presentation of Phetch contained in [3] achieved both of these properties by using a restricted search engine based on keywords collected from the ESP Game.

In general, any reasonable image search engine can be used, provided that two modifications are in place. First, the search space should be of the right magnitude (roughly 100,000 images). This does not mean that the entire search engine should only contain 100,000 images, but that each session of the game should only be played on a subset of the images. Second, to ensure that the Seekers are able to find the correct image, the search engine should artificially place the image among the results whenever the query is "accurate enough." An "accurate" query is defined by the percentage of query words also located in the Describer's text so far. Note that the Seeker still needs an accurate description in order to locate the correct image.

If a Seeker is able to locate the desired image, we assign all relevant query texts as a set of labels for the returned image. This allows us to easily add new images to our search engine.

4. EMULATING PLAYERS

As stated, Phetch requires three to five players: one Describer plus two to four Seekers. Since the total number of players may not always be split perfectly into games of three to five players (e.g., if there is only one player), we can also pair up players with computerized players, or "bots," when needed.

Bots use previously collected game data to emulate players. When simulating a Describer, the bot merely needs to replay an old description (in fact, we soon show how this is also useful to further ensure accurate descriptions). To simulate Seekers, the bot only guesses the correct image if the Describer's text contains a sufficient number of known keywords associated with the image.

5. DESCRIPTION ACCURACY

5.1. Ensuring Description Accuracy

In the following, we describe some of the strategies used to ensure the accuracy of entered descriptions.

- **Description testing.** Most importantly, we use bots to verify description accuracy. When playing as the Describer, the bot plays back a previously entered description to the Seekers. If a Seeker can still find the correct image, this is a significant indicator that the description is of high quality: two different people chosen at random were able to select the correct image given just this description. Indeed, we can use this strategy more than once ("N people chosen at random were able to find the image") to guarantee description accuracy within a given percentage.
- **Random pairing of the players.** We randomly assign players to sessions. This helps prevent players from colluding: for example, two officemates could play at the same time; one as a Describer entering incorrect descriptions and the other as a Seeker that always finds the images.
- **Success of the Seekers.** We use the amount of time taken by the Seekers to find the correct image as an indicator of the quality of the description. Furthermore, if the Seekers do not find the image, we discard the Describer's text.

5.2. Experimental Results

In [3], experimental data showed that descriptions collected using Phetch are extremely precise and complete, and that they are an improvement over a list of accurate keywords collected from the ESP Game. To determine this, a study was conducted in which participants were assigned to one of two conditions: PHETCH or ESP. Participants in each condition were asked to single out one image among other similar ones, based on either a natural language description collected using Phetch or a set of word labels from the ESP Game. The experimental data showed that 98.5% of the descriptions collected using Phetch were sufficient for the participants to find the correct image, whereas only 73% of the images were found using ESP Game labels. This shows the quality of the data collected by Phetch.

6. USAGE STATISTICS

The Phetch game may theoretically produce data that is useful, but if it is not engaging people will not play and output will not be produced. Hence, it is crucial to test our claim that Phetch is entertaining. To do so, we enlisted test players to interact with the game. These people were obtained by offering random players from another gaming site (www.peekaboom.org) the opportunity to play for a two-month period.

A total of 7,120 people played Phetch during this trial, generating 81,950 captions for images. Each session of the game lasted five minutes and, on average, produced captions for 6.8 images. Roughly 75% of the players returned to play on more than one occasion, and some people played for over 110 hours. We believe this data shows that the game is indeed enjoyable.

Given the average number of captions produced in a single game of Phetch, 5,000 people playing the game simultaneously could associate captions to all images indexed by Google in just ten months. This is striking, since 5,000 is not a large number compared to the number of players of individual games in popular gaming sites [10].

7. OTHER USES FOR PHETCH

7.1. Improving Web Accessibility

One of the major accessibility problems of the Web is the lack of descriptive captions for images. Visually impaired individuals commonly surf the Web using "screen readers," programs that convert the text of a web page into synthesized speech. Although screen readers are helpful, they cannot determine the contents of images on the Web that do not have a descriptive HTML ALT caption. Unfortunately, the vast majority of images are not accompanied by proper captions and therefore are inaccessible to the blind [8, 9].

As mentioned in [3], Phetch was originally designed to improve the accessibility of the Web. Our proposed system would store all of our collected captions on a centralized server. Whenever a visually impaired individual using a screen reader would visit a web site, their browser could contact our server to download all relevant captions for the images on that site. The screen reader would then read the caption aloud based on user preferences.

7.2. Human-Powered Descriptive Search

Phetch can also be used to provide free, real time assistance for people unable to find the images for which they are searching by recruiting others to search for them. When the user of the image search engine types a complete description of the desired image, this description can be fed to a currently active game of Phetch as a "bonus round." In this bonus round, all the players act as Seekers based on the description given by the user. Whenever they find an image that matches the description, they click on it and it is sent to the user. If the image search user finds an image they like, points are given to the Seeker who found it.

In essence, this would provide a service similar to Google Answers [6], but with an almost immediate turnaround time. In Google Answers, users submit questions for which they cannot find an answer (e.g., "who is the most famous person alive"), and for a monetary fee, other people attempt to find answers for them over a period of several hours or days. Phetch could provide a similar service for images, except that it would be free and immediate: the players would perform the searches as a part of the game.

7.3. Turning Image Descriptions into Queries

Thus far, we have described Phetch as a tool to obtain captions for arbitrary images. Although this is the main application of our game, other useful data can also be obtained. The process of Seekers finding the correct image involves conversion of a plain English description into a sequence of appropriate search queries. By recording both the Describer's original text descriptions as well as the search queries entered by the Seekers, we can obtain training data for an algorithm that converts natural language descriptions into successful keyword-based queries. Such an algorithm would have important applications in information retrieval (e.g., [4]). We do not investigate this application here, but remark that the problem of using such data to train an NLP system has been studied before [4].

8. CONCLUSION

This paper describes a novel solution to a well-known problem of image captioning and image search. Instead of developing computer vision algorithms or manually annotating images, we make use of a game in order to entice people to correctly caption images and therefore to improve the state of the art in image search. Even if we could pay people to write captions for images, it would cost hundreds of millions of dollars and require that someone manually verify the accuracy of each and every caption. However, Phetch is able to generate these captions for free. Moreover, the results from this game are guaranteed to be accurate. Phetch has generated thousands of image captions, and if it were placed on a popular gaming site, it could create captions for every single image on the Internet within months. It is our intention to integrate Phetch with a large-scale image search engine and turn these technologies into reality, so people can find exactly what they are looking for.

9. ACKNOWLEDGEMENTS

We thank Laura Dabbish, Susan Hrishenko, and Roy Liu for their insightful comments. This work was partially supported by the National Science Foundation (NSF) grants CCR-0122581 and CCR-0085982 (The ALADDIN Center) and by a generous gift from Google, Inc. Luis von Ahn was also partially supported by a Microsoft Research Fellowship and a MacArthur Foundation Fellowship.

10. REFERENCES

[1] von Ahn**,** L. Games With A Purpose. In *IEEE Computer Magazine*, June 2006, pp. 96-98.

[2] von Ahn, L., and Dabbish, L. Labeling Images with a Computer Game. In *ACM Conference on Human Factors in Computing Systems (CHI)*, 2004, pp. 319-326.

[3] von Ahn, L., Ginosar, S., Kedia, M., Liu, R., and Blum, M. Improving Accessibility of the Web with a Computer Game. In *ACM Conference on Human Factors in Computing Systems (CHI)*, 2006, pp. 79-82.

[4] Brill, E., Dumais, S., and Banko, M. An Analysis of the AskMSR Question-Answering System. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2002.

[5] Carson, C., and Ogle, V. E. Storage and Retrieval of Feature Data for a Very Large Online Image Collection. *IEEE Computer Society Bulletin of the Technical Committee on Data Engineering*, 1996, Vol. 19 No. 4.

[6] Google Answers. http://answers.google.com/answers/

[7] Google Image Labeler. http://images.google.com/imagelabeler/

[8] National Organization on Disability Website. http://www.nod.org/

[9] Watchfire Corporation Website. http://watchfire.com

[10] Yahoo!, Inc. *Yahoo! Games*. http://games.yahoo.com