Picturephone: A Game for Sketch Data Capture

Gabe Johnson
Carnegie Mellon University
johnsogg@cmu.edu

ABSTRACT
This article introduces *Picturephone*, a sketch-based game for collecting data on how people make and describe sketches. The approach is based on the technique of human computation, where players provide information about drawings in exchange for entertainment. The system collects raw sketch input and associates it with human-provided text descriptions. Researchers may browse and download this data for their own purposes. The proposed sketching game has users describe drawings textually in one phase followed by another phase of drawing based on a text description. To score points, users must reconstruct a drawing based on a text description or vice-versa. A third phase asks users to judge the work of other players, which awards points appropriately. The Picturephone game system architecture is briefly described, followed by a discussion of future work on sketching games as a platform for researching sketch recognition and interaction.

Author Keywords
Sketch recognition, human computation, games, calligraphic interfaces

ACM Classification Keywords
I.2.6 Learning: Knowledge acquisition; H.5.2 Information Interfaces and Presentation: User Interfaces

INTRODUCTION
Ideal sketch-recognition systems would identify input regardless of who drew it, what domain it is in, or how it is made. Current calligraphic interfaces are typically limited to operate in one domain at a time, and often are sensitive to different drawing styles. A robust recognizer need not be explicitly told which domain the sketch is in, and would tolerate variations in the way people make drawings.

Many sketch recognition user interfaces (SkRUIs) achieve acceptable error rates by limiting vocabulary size or constraining the way people must draw. If the vocabulary is restricted we can build prototypes that facilitate exploration of topics such as segmenting, symbol training, domain modeling, recognition methods and interaction techniques. However, in order for SkRUIs to move beyond research prototypes into "real world" usage we must improve recognizer robustness.

IUI 2009 Workshop on Sketch Recognition
February 8, 2009, Sanibel Island, Florida
Chair: Tracy Hammond

This paper proposes a multi-player sketching game called Picturephone. Its purpose is to capture hand-drawn sketches and player-provided descriptions which can be used by other researchers to develop or test sketch recognition systems. Picturephone is not a new recognition system—it is a tool for capturing hand-made drawings in many domains by many people, along with human-classified descriptions.

Picturephone is inspired by the children's game called *Telephone*. In Telephone, a player privately describes something to the person to the left. That person then conveys the message to the person to their left, and so on. Over time the message may change drastically (and usually entertainingly). For example, consider players giving a good faith effort to convey messages:

Player A:	"The tall man is eating lunch."
Player B:	"The big man is eating lunch."
Player C:	"The fat man is eating lunch."

While the children's game forgives (or encourages) creative elaboration, Picturephone rewards accurate reconstruction of an object description. Game play might progress as follows: *Player A* is given a text description, and they must make a drawing that captures that description as accurately as possible. *Player B* receives the drawing and endeavors to describe it. *Player C* is given Player B's description and draws it. An unrelated player *Player D* is asked to judge how closely Player A and C's drawings match, which assigns a score to players A, B, and C.

The point can be made more strongly with reader involvement. Assume the role of Player A by making a sketch based on the description below in Figure 1. Compare your drawing with that in Figure 2.

"A blocky looking house with a window on the left and a door on the right, with a curvy path extending towards you. There is a tree next to the house, and the sun and some birds are in the sky."

Figure 1. Sample text description on which a sketch is based.

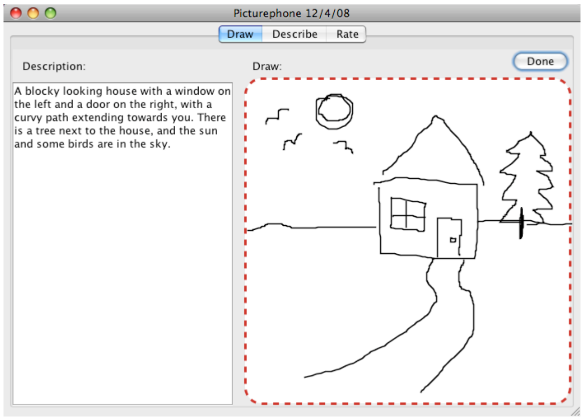

Figure 2. Picturephone's 'draw mode'. The player is given a text description (displayed at left), and is challenged to translate that description into a drawing (on the right).

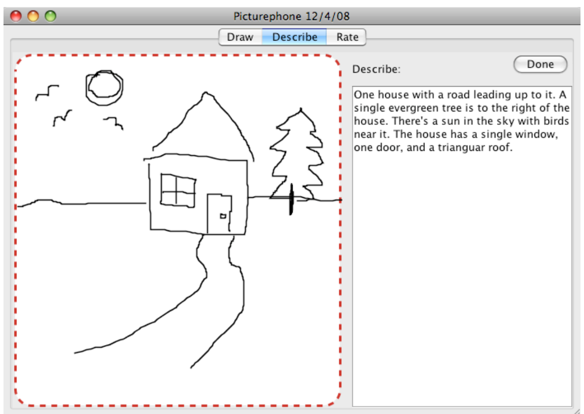

Figure 3. Picturephone's 'describe mode'. Players are asked to accurately describe the sketch so another player can redraw it as closely as possible based on their description.

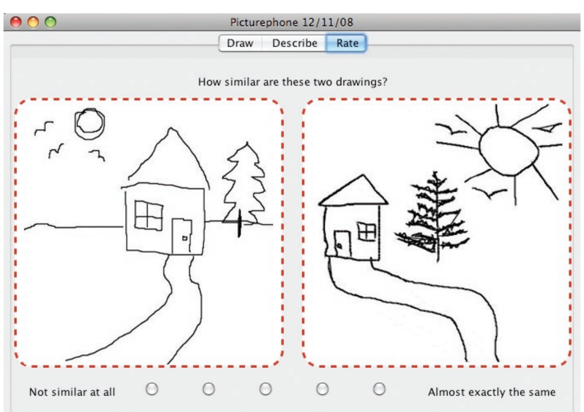

Figure 4. Picturephone's 'rate mode'. Players rate the work of others, which awards points.

RELATED WORK

People spend countless hours playing games every day. Readers may be familiar with parlor games such as Pictionary [10], where players take turn drawing objects, actions, or concepts, and others must guess what the drawing is. A non-commercial parlor game, 'Telephone Pictionary' has players passing notes to each other, alternately drawing or writing clues based on what the previous player created. There are many online computer games that similarly involve drawing pictures and guessing what they depict, such as *iSketch* and *Draw My Thing* [4, 11].

'Human computation' programs entertain users while generating useful data for researchers. von Ahn's ESP game is arguably the best known example, where pairs of players are shown the same picture [21]. Each player provides text labels and are awarded points when the entry matches the other player's label. The approach has been adopted by Google Images to label pictures on the world wide web [6]. Other projects such as the Open Mind Commons [18] and LEARNER2 [3] depend on many untrained volunteers to provide data about 'common sense' knowledge, helping to build libraries of how words are commonly used.

The ESP Game, LEARNER2 and the Open Mind Commons acquire data in structured environments, easing the task of forming semantic models. However, Picturephone's text descriptions are full English sentences. Fortunately numerous natural language processing (NLP) tools (such as MontyLingua [14]) are available for performing semantic interpretation. Picturephone has not progressed to the point where NLP is required, however it will be necessary for captured data to be useful.

Many sketch recognition strategies use machine learning to form models of what is to be identified. Some approaches call for one example (e.g. the $1 Recognizer [22]), while others use several examples. Various machine learning approaches are used in the sketch recognition research community, including Bayes Networks and variants [1, 2, 5], Hidden Markov Models, Neural Networks [19], Linear Discriminant Analysis [17], and myriad visual pattern matching techniques specifically developed for recognizing hand-made drawings and gestures [13, 15]. While the techniques work differently, they all require training examples. For a detailed review of sketch recognition techniques, please see [12].

A common problem with these approaches is training bias—examples may be made idiosyncratically or with too little variation to capture the range of ways something could be drawn [9]. The current work collects data from many people in a variety of contexts, yielding a fuller breadth of styles to record.

Many systems use structured text descriptions to facilitate sketch recognition, indicating geometric elements and their relative sizes and positions [7, 8, 16]. Such approaches can be useful because elements can be described in general rather than particular terms. For example, a triangle is *generally* described as a polygon with three unique vertices, while a *particular* triangle may have vertices $(0, 0)$, $(1, 0)$, and $(0, 1)$. Some have developed ways to translate sketches into rule-based systems automatically [20] or interactively [9]. These approaches might be bolstered in the current work by using the associated text descriptions.

IMPLEMENTATION

Picturephone uses a web oriented client/server architecture and is known to run on Windows, Mac OS X, and Ubuntu Linux. Both client and server are written in Java. Communication is done with the standard HTTP protocol using the host web browser's network connection, allowing the game to work unimpeded by firewall or router restrictions. This allows Picturephone to reach beyond the laboratory, enabling use for many people[1].

Messages are formatted in a simple extensible text protocol. While the current system could easily operate with batch-mode communication (e.g. uploading sketches when finished), subsequent work will require streaming, synchronous data. For this reason, the messaging protocol is lightweight, and individual messages can easily be mixed and routed by the server to any number of clients.

There are three primary game modes: draw, describe, and rate. Players are randomly assigned one of these modes. In *Draw* mode (Figure 2), players are given a text description and are asked to draw it using the sketching surface at the right. A time limit is enforced to encourage simplicity.

Figure 3 shows the *Describe* mode interface. The system provides a sketch on the left, and users must describe it using the text area at right. This mode has both a time limit as well as a maximum word limit.

Last, the player can be asked to judge how well drawings match using the *Rate* interface, shown in Figure 4. The system finds two drawings the player was not involved in making. Each pair of sketches was mediated by a text description which is not shown. Therefore, the rating describes how well Player A's sketch matches Player B's sketch as mediated by Player C's description. The rating given by Player D factors into a score applied to players A, B, and C. The higher the rating, the more points that are awarded to A, B, and C. An individual player's score accumulates from making drawings, descriptions and (when other players rate their work) from ratings.

DISCUSSION AND FUTURE WORK

Picturephone is the first instance of a class of planned sketching games that could provide researchers with a method to acquire data about drawings. This includes the physical act of sketching as well as how people describe those drawn elements. While historically researchers have been confined to collecting data from a limited number of users, this approach enables data acquisition from a vast pool of users on a wide range of topics. In order for this tool to be widely useful to other researchers, it should be possible to submit initial text descriptions. This will allow others to collect data about particular domains. All data will be available to researchers.

Users will only play Picturephone if it is engaging. Therefore, game play is a serious concern. The current generation of popular sketching games like those mentioned above can serve as useful measures of what is (and what is not) fun. An earlier pilot study on sketching games indicates users enjoy synchronously drawing on the same shared surface, and spend more time playing when the game involves a chat component. Alternate drawing tools and colors were requested by several users. However, care must be taken to not erode the purpose of the tool: if structured drawing tools and colors are available, the data may not be appropriate for use in training sketch recognizers. However, if hybrid tools involving structured and calligraphic interaction are useful, there is no reason to limit data collection to sketching.

Game mechanics have consequences for the type of data that is collected. Picturephone is multi-player, but those people are not necessarily playing at the same time. This supports a relaxed playing style, since users can come and go as they please without affecting others. A synchronous multi-player game might encourage users to draw things differently in order to entertain others, as everybody can see what is going on at the same time.

Referring to the the example text from Figure 1, humans can see the various noun phrases (house, tree, path, sun, birds) in the drawing. However, there are objects and relational constraints that were not made explicit in the description. The sun is *above* the house; the tree is to the *right*; the path extends towards *you* (a noun which is not part of the sketch). When translating from one form to another, information changes. For example, players often embellish objects, as in the ironic frowning sun in 5(b). The horizon is never mentioned in the text, yet it appears in two of the four drawings, suggesting that latent, tacit knowledge may be made explicit by others.

The drawings in Figure 5 feature the sun, but each is drawn differently. A recognizer could be made for each individual drawing style, but that strategy would quickly yield too many recognizers to manage. Instead we could use the variety of drawing styles as a basis for learning what is invariant about certain classes of drawn elements, and build recognizers based on those invariants.

CONCLUSION

Sketch recognition approaches commonly require access to a pool of examples made by many people in many domains. This paper has presented Picturephone, a sketching game for collecting data about how people make and describe hand-made drawings. Researchers may suggest drawing topics or domains, and are given complete access to all data collected via this system.

Sketching games like Picturephone might serve as an effective platform for pushing the current state of the art of sketch

[1]Picturephone may be played at http://six11.org/picturephone/

Initial text: A blocky looking house with a window on the left and a door on the right, with a curvy path extending towards you. There is a tree next to the house, and the sun and some birds are in the sky.

Player A:

Player B: One house with a road leading up to it. A single evergreen tree is to the right of the house. There's a sun in the sky with birds near it. The house has a single window, one door, and a trianguar roof.

(a) The system provides an initial text description, which Player A sketches. Player B in turn describes that sketch in words.

(b) Players C, D, and E independently draw their interpretations based on Player B's description.

Figure 5. Several rounds of Picturephone played asynchronously.

recognition based user interfaces. This includes not only recognition but also interaction techniques appropriate for calligraphic interfaces.

REFERENCES

1. C. Alvarado and R. Davis. Dynamically constructed bayes nets for multi-domain sketch understanding. In *International Joint Conference on Artificial Intelligence*, 2005.
2. C. Alvarado, M. Oltmans, and R. Davis. A framework for multi-domain sketch recognition. In *2002 AAAI Spring Symposium on Sketch Understanding*, 2002.
3. T. Chklovski. Designing interfaces for guided collection of knowledge about everyday objects from volunteers. In *Proceedings of Conference on Intelligent User Interfaces (IUI05)*, 2005.
4. FlashNinjaClan.com. Draw my thing. http://www.flashninjaclan.com, 2008.
5. M. Fonseca, C. Pimentel, and J. Jorge. CALI: An online scribble recognizer for calligraphic interfaces. In *AAAI 2002 Spring Symposium (Sketch Understanding Workshop)*, pages 51–58, 2002.
6. Google, Inc. Google Image Labeler. http://images.google.com/imagelabeler/, 2008.
7. M. D. Gross and E. Y.-L. Do. Ambiguous intentions: A paper-like interface for creative design. In *UIST '04: ACM Conference on User Interface Software Technology*, pages 183–192, Seattle, WA, 1996.
8. T. Hammond and R. Davis. LADDER, a sketching language for user interface developers. *Elsevier, Computers and Graphics*, 29:518–532, 2005.
9. T. Hammond and R. Davis. Interactive learning of structural shape descriptions from automatically generated near-miss examples. In *Intelligent User Interfaces (IUI)*, pages 37–40, 2006.
10. Hasbro Corp. Pictionary. http://www.hasbro.com, 2008.
11. iSketch.net. i-sketch. http://www.isketch.net, 2008.
12. G. Johnson, M. D. Gross, J. Hong, and E. Y.-L. Do. Computational support for sketching in design: a review (to appear). *Foundations and Trends in Human-Computer Interaction*, pages 1–101, 2009.
13. L. B. Kara and T. F. Stahovich. An image-based, trainable symbol recognizer for hand-drawn sketches. *Computers and Graphics*, 29(4):501–517, 2005.
14. H. Liu. MontyLingua: An end-to-end natural language processor with common sense. http://web.media.mit.edu/ hugo/montylingua/, 2004.
15. W. Newman and R. Sproull. *Principles of Interactive Computer Graphics, 2nd ed.* McGraw-Hill, 1979.
16. B. Pasternak and B. Neumann. Adaptable drawing interpretation using object-oriented and constraint-based graphic specification. In *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR '93)*, 1993.
17. D. Rubine. Specifying gestures by example. *SIGGRAPH Computer Graphics*, 25(4):329–337, 1991.
18. R. Speer. Open mind commons: An inquisitive approach to learning common sense. In *Workshop on Common Sense and Intelligent User Interfaces*, 2007.
19. F. Ulgen, A. Flavell C., and N. Akamatsu. Geometric shape recognition with fuzzy filtered input to a backpropagation neural network. *IEICE transactions on information and systems*, 78(2):174–183, 19950225.
20. O. Veselova and R. Davis. Perceptually based learning of shape descriptions. In *AAAI '04: Proceedings of the National Conference on Artificial Intelligence*, pages 482–487, San Jose, California, 2004.
21. L. von Ahn and L. Dabbish. Labeling images with a computer game. In *CHI '04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 319–326, New York, NY, USA, 2004. ACM.
22. J. O. Wobbrock, A. D. Wilson, and Y. Li. Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In *UIST '07: Proceedings of ACM Symposium on User Interface Software and Technology*, pages 159–168, New York, NY, USA, 2007. ACM.