Games For Sketch Data Collection

Gabe Johnson¹ and Ellen Yi-Luen Do²

¹Computational Design Lab, Carnegie Mellon University
²College of Architecture & College of Computing, Georgia Institute of Technology

Abstract

This article describes sketching games made for the purpose of collecting data about how people make and describe hand-made drawings. The approach leverages human computation, whereby players provide information about drawings in exchange for entertainment. The games facilitate the collection of raw sketch input and associates it with human-provided text descriptions. Researchers may browse and download this data for their own purposes such as training sketch recognizers. Two systems with distinct game mechanics are described: Picturephone and Stellasketch. The system architectures are briefly presented, followed by a discussion of our initial results using sketching games as a research platform for sketch recognition and interaction.

1. Introduction

Current calligraphic systems based on sketch recognition typically work in one domain at a time, and often are sensitive to the drawing styles of different people. Ideally, sketch recognition systems would identify input regardless of who drew it, what domain it is in, or how it is made.

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 to a single domain we can build prototypes to explore topics such as segmenting, symbol training, domain modeling, recognition methods and interaction techniques. However in practice, people sketch in many different domains, sometimes in several notation types on the same page. It is common for people to draw back-of-the-envelope diagrams mixed with TODO lists and simple arithmetic calculations. For example, the floor plan in Figure 1 is drawn with conventional architectural notation with iconic figures of furniture like a piano, couches, chairs, and tables. It also includes non-architecture elements like numbers and text. Most of the numbers represent dimensions, but the encircled 10 indicates the drawing is the tenth in a series of sketches.

A frequently cited motivation for developing sketch-based interfaces is the fluid, informal interaction that sketching allows [LHK∗02]. If SkRUIs are to retain the usability of pencil-and-paper, users must not be forced to tell the system which domain they are working in.

Calligraphic systems should be tolerant of different user’s drawing styles. Fortunately for many iconic figures, there is remarkably little variation in the way that people draw. Sezgin found that even though there are 720 possible ways to construct a stick figure (with six distinct components), most are drawn in one of five stroke orders [SD07]. Often, abstract elements such as “wind” or “sunlight” are also drawn consistently. Sunlight, for example, is drawn as a circle (or partial circle) with several short lines extending outward from its edge [Do05].

Figure 1: An architect’s sketched floor plan with several types of notation including text, numeric dimensions and symbols for furniture.
People use a variety of drawing styles when the subject matter is uncommon or complicated. For example, Figure 2 shows drill presses sketched by five people. They are made from different perspectives, emphasizing different features while omitting others. While these sketches are to some degree recognizable as drill presses, test participants mistook them as other things such as chairs, monsters, or robots. If any one of these drawings is used to train a recognition system the other examples would not be identified. However, these drawings do have common features (such as the drill bit) that let humans identify them as depictions of drill presses.

Most current work on sketch recognition is focused on making sense of diagrammatic drawings using restricted visual vocabularies. But such drawings often contain rare but important elements that make the sketch expressive (such as Figure 1’s piano or potted plants). Humans have skill and experience at interpreting such sketches that could be leveraged by sketch recognition systems.

This paper describes our efforts developing multi-player sketching games to capture a data corpus of hand-drawn sketches and player-provided descriptions from many users on a wide range of subjects. We present related work, followed by an introduction the two games, Picturephone and Stellasketch. We then consider how the game design affects the type and quality of that data, and present initial findings from playtesting both systems. Finally, we discuss several possible application areas for the collected data.

2. Related work

People spend countless hours playing games every day. Readers may be familiar with parlour games such as Pictionary [Has08], where players take turns 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 Broken Picture Telephone [SK08, Nov09].

‘Human computation’ programs leverage the ability of people to perform recognition tasks—often in an entertainment environment—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 [vAD04]. 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 [Goo08]. Other projects such as the Open Mind Commons [SHL08] and LEARNER2 [Chk05] depend on many untrained volunteers to provide data about ‘common sense’ knowledge, helping to build libraries of how words are commonly used.

Many sketch recognition strategies use machine learning to form models of what is to be identified. Some approaches require only a single example (e.g. the $1 Recognizer [WWL07]), while others use several. Various machine learning approaches are used by the sketch recognition research community, including Bayesian Networks and Hidden Markov Models, Neural Networks [UFCA25], Linear Discriminant Analysis [Rub91], and visual pattern matching techniques [KS05, NS79]. While these approaches work differently, they generally require several training examples. For a detailed review of sketch recognition techniques, see [JGHD09].

A common problem with many such approaches is training bias—examples made in an idiosyncratic fashion or with too little variation to capture the range of how an element could be drawn in practical situations [HD06]. Sketching games collect data from many people in a variety of contexts, yielding a fuller breadth of styles to record.

Many systems use constraint languages to facilitate sketch recognition, indicating geometric elements and their relative sizes and positions [GD96, HD05, PN93]. 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 constraint systems automatically [VD04] or interactively [HD06]. These approaches might be bolstered in the current work by using the associated text descriptions.

The Caltech 256 dataset includes tens of thousands of categorized photographic images [GHP07]; the MIT LabelMe tool has collected a corpus of hundreds of thousands of labeled objects in photographs [RTMF08]. Both data sets are used by computer vision researchers. Sketching researchers have collected and made available smaller data sets. For example, the ETCHASketches corpus contains hundreds of sketches made in a few diagram languages like electronic circuit design or family trees [OAD04].

Once digital ink has been acquired, portions can be labeled according to their purpose. Such sketch data collection tools have recently been developed to more easily collect and analyze domain-specific sketching data. Blagojevic et. al describe a tool that collects sketch data in specific diagrammatic domains [BPGW08]. The tool also supports
3. Sketching Games

Picturephone and Stellasketch are web-based data collection games designed to give people an entertaining way for researchers to gather data about how people make and describe sketches. The games are implemented as Java applets, which communicate with a server component also written in Java. We have successfully played the games on Windows, Mac OS X, and Ubuntu Linux. 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 the sketching games to reach beyond the laboratory, enabling use for many people.†

The client and server software directly pertaining to capturing, rendering, and sharing sketch data are part of the open-source Olive Sketching Framework. Olive allows many people to concurrently sketch on a shared canvas, and is intended to work in any modern web browser with Java 1.5 or higher installed.

3.1. Picturephone

The first game, Picturephone [Job09], is inspired by the children’s game called Telephone. In Telephone, a player privately describes something to the person to the left. That person 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 fat man is eating lunch.”
Player B: “The big man is eating lunch.”
Player C: “The tall man is eating lunch.”

While the children’s game forgives (or encourages) creative elaboration, Picturephone rewards accurate reconstruction of an object description. Referring to Figure 3, game play might progress as follows: Player A is given a text description and must make a drawing that accurately captures its essence. Player B receives the drawing and endeavors to describe it. Player C is given Player B’s description and draws it. Unrelated players are asked to judge how closely Player A and C’s drawings match, which assigns a score to players A, B, and C.

Picturephone has three primary game modes: draw, describe, and rate. Players are randomly assigned one of these modes. In Draw mode (Figure 4(a)), players are given a text description and are asked to draw it using the sketching surface at the right.

Figure 4(b) shows the Describe mode interface. The system shows a sketch, and users must describe it using the provided text area. The best descriptions are clear and unambiguous, because this text serves as the basis for other player sketches.

Last, the player can be asked to judge how well drawings match using the Rate interface, shown in Figure 4(c). 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 C’s sketch as me-
Figure 4: The three Picturephone playing modes: draw, describe, and rate. The initial description is “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 5: The Picturephone browsing UI, displaying sketches from several games.

In addition to the Java applet, the Picturephone web site gives players additional abilities. Users can suggest additional initial text descriptions, which is necessary to give players new material. Figure 5 shows Picturephone’s web-based sketch browser displaying tiled thumbnails past game drawings. In addition to providing entertainment value to players, researchers can use the browsing interface to find and download sketch data.

3.2. Stellasketch

Stellasketch is a synchronous, multi-player sketching game similar to the parlor game Pictionary. One player is asked to make a drawing based on a secret clue (as shown in Figure 6). The other players see the drawing unfold as it is made and privately label the drawing. While Picturephone’s descriptions are meant to be used to recreate a drawing, Stellasketch’s labels simply state what the sketch depicts. Labels are timestamped, so they can be associated with sketches at various stages of completion.

To play Stellasketch, players join a game room of their choosing. A game begins by giving players a chance to vote for that game’s theme (such as ‘Household Objects’). A game consists of a number of rounds. At the beginning of a round, one of the players is randomly chosen to be the sketcher (person drawing), and is given a clue associated with the current theme. All other players are label-
ers. The sketcher proceeds to draw the clue and the labelers give short descriptions of the drawing. During the sketching phase, players do not see each other’s labels; however when the sketching phase is done, players see all the other labels in the order they were given. Figure 7 shows an example sketch and labels for the clue ‘Horse Racing’.

After the sketching phase, all players are allowed to draw on the sketching canvas. While this data is not recorded and doesn’t directly offer a research benefit, it is entertaining to draw on the shared surface, and helps keep people involved in the game if they haven’t sketched in a while.

Stellasketch has web pages enabling players to suggest new themes and clues. Like Picturephone, there is a web-based browsing interface. Users may view sketch data by theme, clue, artist, or game. Raw sketch data is also available for download, which includes (x, y) points, timestamps of when each point or label was created according to the originating user’s clock and received according to the server’s clock.

4. Playtesting Results

People only play games if they are engaging. Therefore the quality of game play is a serious concern. An early pilot study on sketching games indicated 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 or rectifiers. For this reason, the drawing surface in both games support only freehand ink input without the ability to undo or erase.

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, as users may come and go as they please without affecting others. The synchronous nature of Stellasketch encourages spontaneity: users draw things differently in order to entertain others, as everybody can see what is happening at the same time. However, a few participants in the playtesting reported an uncomfortable sense of stage fright when it was their turn to sketch.

Picturephone players can identify various named elements (e.g., house, tree, path, sun, birds) in a drawing. However, there are objects and relational constraints that were not explicitly stated in the original description. For example in Figure 3, 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 Figure 3(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.

Picturephone encourages users to make complete drawings and describe them in great detail. While some players enjoyed the challenge of giving highly detailed descriptions, many players did not like it. One player described this mode of gameplay as “clinical”; another said it was “like doing homework”. In general, Picturephone users preferred to create drawings and browse other people’s sketches.

The drawings in Figure 3 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.

The characteristics of the two games’ data differ. While Picturephone’s sketches are complete at the time others describe them, a Stellasketch drawing is labeled as it is made. Furthermore, Picturephone descriptions are generally longer and in approximately complete sentences, but Stellasketch labels are often short noun-phrases. Because a Stellasketch drawing is labeled as it is made, players usually furnish multiple interpretations, and there is often significant agreement among players. Agreement indicates those interpretations are more ‘correct’. Sometimes labels cluster into more than one group (e.g., Figure 7 has more than one participant labeling the sketch as ‘dog’ and ‘horse’). This might provide the basis for forming confusion matrices.

Because these tools are based on participant entertainment, players frequently draw or write things to amuse their friends. There is no clear method for automatically discerning which data is valid and which is not. For example, Figure 6 shows a drawing of a Squid with the irrelevant handwritten word Disco. Obviously invalid data should not be used to train recognizers.

Figure 7: Web interface showing results of a single Stellasketch round of play with four people providing labels.
The playtesting sessions for Picturephone involved a total of 40 users, who provided 423 descriptions, 1703 judgments, and 486 sketches. On average, a Picturephone sketch took approximately 30 seconds to make. When describing, participants mostly took less than 10 seconds, though the best descriptions take 20 to 30 seconds. Players can rate a sketch pair quite quickly, averaging only three seconds per judgment.

While Picturephone supports people to play at their own rate, a game of Stellasketch requires several people to play at the same rate. A game of Stellasketch takes just over two minutes, during which three sketches are labeled. Stellasketch playtesting involved 35 participants playing 42 games, producing 105 sketches with 543 labels.

5. Future Directions

Using the current work as a point of departure, there are two likely veins of future research: exploring games as an effective method of sketch data collection, and developing techniques that use the collected data.

The games presented here gather sketches and descriptions with different characteristics: Picturephone asynchronously collects long sentences that describe fully-formed sketches; Stellasketch synchronously gathers short noun-phrases that label sketches as they are made. Subsequent games might be structured to gather labels about particular elements within a sketch, much like the LabelMe system asks users to identify object boundaries in photographs.

Sketches might be effective as the subject of CAPTCHA systems. A CAPTCHA is a small puzzle used by many web sites to determine if a user is a human or a software agent. The puzzle should be easily solved by humans while presenting a challenge to an AI program. Users solve most current CAPTCHAs by typing the letters and numbers contained in an image of distorted text. As automated character recognition techniques improve, textual CAPTCHAs are giving way to other types of puzzles such as rotating an image to its proper orientation [GKB09]. A sketch-based CAPTCHA could ask users to properly label a sketch or draw a common object.

There are several application areas that stand to benefit from the collected data. Researchers have recognized that the technique people use to draw an object are somewhat consistent (e.g., people will draw a garden rake from top to bottom, but cigarette smoke from bottom to top) [vS84]. This insight has been used in sketch recognition techniques that leverage probabilistic models of drawing strategies [SD07]. But before we can employ knowledge of consistent drawing patterns, we must first have a corpus of data to identify such patterns. Sketching games could provide that data.

Developers of sketch recognizers could use sketching games to gather labeled training examples. It is clear that there is more noise in game-collected sketches than in some other contexts. For example, players often embellish an object (such as a house) with unnecessary ink (such as a horizon). However, extra strokes can give human players additional context, easing the human task of recognizing the drawing. Due to such noise, current sketch recognizer training strategies might not benefit the gathered data unless it has been filtered to exclude spurious ink. Fortunately, the proposed data collection technique is designed to gather a lot of data, from which researchers can pick a subset of examples.

Many calligraphic systems perform rectification or beautification by straightening lines, smoothing arcs, sharpening corners, or maintaining perceptual properties like parallelism. Commonly, developers of rectification techniques test their algorithms on their own sketch input. This introduces a form of testing bias because the rectifier might not work well on other people’s sketches. It is a good development practice to test on a wide variety of sketches made by many people. The current work is well-suited to support that development and testing practice.

6. Conclusion

Development of interactive calligraphic systems commonly require access to a pool of examples made by many people in many domains. This paper has presented Picturephone and Stellasketch, two sketching games 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 collected data. While previous sketch data collection tools have been successful in gathering data from tens of users, we suggest that games might be an appropriate method to collect sketch data from many more people than would otherwise be possible.

7. Acknowledgments

We thank the students at Georgia Institute of Technology for their participation in playtesting, Shaun Moon for providing the floor plan sketch in Figure 1, and ReadyTalk (www.readytalk.com) for donating the account used in an early pilot study.

References


