

# SOUSA: The Sketch-Based Online User Study Application

Emily R. Jacobson, Brandon L. Kaster, and Dr. Tracy Hammond (advisor)

**Abstract**—We have developed an interface for the Sketch-based Online User Study Application (SOUSA), the purpose of which is to create a universal, standardized set of sketch data. This online interface will make sketch data collection more efficient for researchers and more accessible to a general audience. The expected contribution of our work will be an increase in participation of researchers and practitioners in the field of sketch recognition; thus developing a large, robust repository of sketch data. A motivating factor behind our work is to allow sketch recognition researchers to focus on higher-level tasks, rather than data collection. Features of our interface include a standardized collection mechanism and set of sketch data which will allow new sketch recognition algorithms to be more easily compared with existing models. Our new interface will allow researchers to download their own, as well as other publicly available, data gathered from collection and verification studies. This new interface will be hosted by the Sketch Recognition Laboratory at Texas A&M University, providing researchers a single, unified solution for sketch data collection and management.

## I. INTRODUCTION

THE field of sketch recognition currently lacks a large corpus of data with which to compare new algorithms. Without such a dataset, reported accuracies for new algorithms are not comparable to one another. Although there have been attempts at creating such a corpus, these versions have focused on a few common domains, such as circuit diagrams and family trees, and have not been flexible enough for researchers in more specialized areas. We propose, and have created, an online system that allows researchers to easily set up data collection for any domain they choose. Additionally, the system makes it easy to share this data with other researchers, thus creating a repository from which sketch recognition researchers can retrieve sketch data for testing purposes.

## II. PREVIOUS WORK

SOUSA[1] was created in 2007 by Paulson et al. in the Sketch Recognition Laboratory at Texas A&M University.

Paper submitted July 31, 2008. This work was supported in part by the National Science Foundation Research Experience for Undergraduates and the Computing Research Association Distributed Mentor Project.

Emily R. Jacobson is a senior at Grinnell College, Grinnell, IA 50112 (e-mail: jacobso2@grinnell.edu).

Brandon L. Kaster is a junior at Texas A&M University, College Station, TX 77840 (e-mail: wiredgreen@tamu.edu).

Dr. Tracy Hammond is an Assistant Professor of Computer Science at Texas A&M University, College Station, TX 77840 (e-mail: hammond@cs.tamu.edu).

Designed using a Java interface and Java applet for data collection, this version allows for easy creation of collection and verification studies. Collection studies can be created in any domain, with the researcher simply describing the shapes for which she wants to collect samples. Verification studies are created from an associated collection study and can be used to understand human interpretations of images.

Although the original SOUSA provided a simple management system, it lacked security, and data gathered was not intentionally publically accessible. Anyone who knows the name of a study could edit that study or download the data associated with it. Additionally, the editing features were sometimes difficult to use; the description of a shape might be changed, but the description provided to participants often remained the same. Researchers did not have the option to delete studies they no longer wished to be available. Our new system solves these problems by having researchers register with a unique username. This gives researchers sole control over the contents of their studies. Further, researchers can choose to keep their studies and the associated data public or private, on a per study basis. If a researcher wishes to share her study and data, she can easily do so, but she can also easily choose to keep this private. Our system also has special pages for editing studies, thus providing a systematic way to change aspects of a specific study and ensuring these changes take effect immediately. Table I illustrates the updated features in our system.

Other corpora of sketch data are limited to specific domains or focus more on labeling primitives in large diagrams. ETCHA Sketches, designed by Oltmans et al.[2] was comprised of diagrams in four domains: circuit diagrams, floor plans, family trees, and basic geometric shapes. Labels were then assigned to the primitives making up each sketch. This work was more focused on labeling portions of sketches and understanding how these primitive shapes are used to create complex diagrams. In contrast, our work is focused on enabling the creation of a large corpus of sketch data in a wide variety of domains and making this data universally available.

A data collection tool was presented at SBIM 2008 by Blagojevic et al. [3]. While this tool also aims to allow researchers to easily collect data in any domain they define, this system focuses more on gathering and labeling large systems and is thus not as flexible for domains in which desired shapes cannot be described in diagrammatic terms. Additionally, our system is accessible through

TABLE I  
UPDATED FEATURES IN SOUSA

Feature	Original SOUSA[1]	New Version
Easy to access web interface		<input checked="" type="checkbox"/>
Streamlined study management		<input checked="" type="checkbox"/>
Easy to create collection studies	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Dynamically upload (optional) sample images	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Create verification studies with an associated collection study	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Secure user accounts		<input checked="" type="checkbox"/>
Designate studies and associated data as public or private		<input checked="" type="checkbox"/>
View “on the fly” statistics		<input checked="" type="checkbox"/>
Browse or search public studies and data		<input checked="" type="checkbox"/>
Download data	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Download any data designated as public		<input checked="" type="checkbox"/>

a web interface, making its studies and data available to researchers worldwide.

### III. FRAMEWORK AND IMPLEMENTATION

SOUSA was developed to allow researchers the flexibility to create user studies in any specified domain and to afford researchers a large dataset with which to compare computational styles. Our system allows for both collection and verification studies to be created. SOUSA now uses a PHP frontend with a MySQL database; an adapted version of the original Java applet is used for sketch data collection and is launched from the web interface.

#### A. Setting up a collection user study

Once a researcher has registered and logged into our website, she can easily create a new collection user study by accessing the study management page and choosing to create a new study. She will choose a title for the study and choose whether the study itself and the data associated with that study will be published. If she chooses to publish the study, participants browsing or searching for studies will be able to take this study; if she chooses not to publish it, a unique URL for that study will be provided to her for intended participants. Publishing the data will make any data associated with the study available for download to any interested researcher; unpublished data is available to the researcher alone. Our hope is that most data will be published, thus creating a larger, more robust repository of sketch data, but we leave this choice to the researcher.

The researcher specifies any study directions she would like to provide each participant. This is intended to allow for general information of which the researcher would like each participant to be aware. The researcher can then specify questions to ask of each participant prior to completing the study. This information might be used to gauge the participant’s familiarity with the domain or to gather any other contextual information the researcher deems important. The researcher then specifies shapes for which she would like to collect sketch data. These shapes can be in any domain and of any intricacy. A shape is specified by four fields:

1) a description, which will be displayed to each participant;

- 2) any shape-specific instructions, which will appear after the description and are intended to provide any additional necessary clarification;
- 3) the number of times each participant will be asked to sketch this shape;
- 4) a maximum number of strokes that each participant will be allowed to use to sketch this shape.

\* Indicates Required Field

Study Title:*	<input type="text" value="Basic Shapes"/>
Study published?*	<input checked="" type="radio"/> Yes <input type="radio"/> No
Data published?*	<input checked="" type="radio"/> Yes <input type="radio"/> No
Study Directions:	<input type="text"/>
# of Questions:	<input type="text" value="2"/> <input type="button" value="Update"/>
Question #1:*	<input type="text" value="Input type (pen or mouse)"/>
Question #2:*	<input type="text" value="What is your occupation?"/>
# of Shapes:	<input type="text" value="3"/> <input type="button" value="Update"/>
Default Values:	Use default values: <input type="button" value="Disable"/> <input type="button" value="Enable"/>
	Default Draw Count* <input type="text" value="3"/>
	Default Maximum strokes (-1 denotes infinity)* <input type="text" value="-1"/>
Shape #1:	Description* <input type="text" value="square"/>
	Any Special Instructions <input type="text"/>
Shape #2:	Description* <input type="text" value="circle"/>
	Any Special Instructions <input type="text"/>
Shape #3:	Description* <input type="text" value="triangle"/>
	Any Special Instructions <input type="text"/>
# of Tags:	<input type="text" value="1"/> <input type="button" value="Update"/>
Tag #1:*	<input type="text" value="polygons"/>
<input type="button" value="Create New Collection Study"/>	




Fig. 1. Collection user study creation form with sample values.

Finally, the researcher specifies any tags which she would like to associate with this study. These tags will be visible to researchers and participants browsing and searching for studies and will help classify shapes into various categories. To complete collection study creation, the researcher simply clicks “submit,” and the information is written to the database. Fig. 1 shows the collection creation page with sample values.

Once the researcher has created the collection study, she may upload sample images. When a participant takes a collection study, thumbnail images will be displayed at random. These images can provide additional clarification information to a participant, aiding her understanding of the desired shape to be sketched. Each shape can have multiple images associated with it, and a single image will be chosen at random each time a participant is asked to draw that shape. A record is kept of which image was shown to the participant for each sketch. Images are easily added by accessing the management page and choosing to upload images for a particular collection study. Fig. 2 shows the image upload page after one sample image has been uploaded per shape. Sample images are not required, and a researcher may choose to provide none at all, or to only provide sample images for a few shapes. While sample images can certainly be helpful in clarifying the desired shape to be drawn, they may also encourage a participant to draw an image that resembles the sample; for complicated sketches or ones with multiple interpretations, this may skew the data.

All of a researcher’s studies appear on her study management page (Fig. 9 shows the study management page after the example collection study—as well as two verification studies—have been created). From here, a researcher may choose to edit or delete studies, download associated data, or create additional studies.

Your current file status is shown below. Users taking your study will be asked to draw each of these shapes specified by the *number of images to draw* for each shape.

Current File Structure	
Shape: triangle 	Delete
Shape: square 	Delete
Shape: circle 	Delete

Select your local file(s) for upload.

Shape: square	<input type="text"/>	Browse...
Shape: circle	<input type="text"/>	Browse...
Shape: triangle	<input type="text"/>	Browse...
Upload Images		

Fig. 2. Page on which images are added to a collection study; images for example study are shown here.

### B. Setting up a verification user study

Verification studies are created from an associated collection study; each collection study can have multiple verification studies. Verification studies are useful for understanding human interpretations of previously drawn sketches. For instance, in a perception study, a verification study might be set up to understand what participants identify as “near” and as “far” (see Fig. 3 for an example image in such a study).

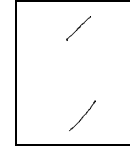


Fig. 3. An example image for a perception study; a participant would be asked to identify whether the intention of this image was two lines that are “near” or “far.”

A researcher creates a new verification study by accessing the study management page and choosing to create a new verification study for a specific collection study. Similarly to a collection study, she will choose a title and decide whether the study and/or its associated data will be published. If a collection study is not published, verification studies associated with that collection study will not be published; likewise, if the data associated with a collection study is not published, any data associated with any related verification studies will also not be published. In future work, we hope to allow verification studies to be independently published. As in a collection study, the researcher may then enter any study directions she wishes to provide each participant; questions are also entered in the same way as a collection study.

In a verification study, however, the researcher does not designate shapes that she wishes the participant to draw. Rather, the researcher sees a list of all shapes in the associated collection study and is able to choose which of these shapes she wishes to include in the verification study. For each shape, the researcher also enters the number of times she would like each participant to verify that shape. This number may not exceed the number of samples collected for this shape; the researcher can enter any number she wishes, but when a participant takes a collection study, the Java applet will first check the number of available samples for each shape and will throw an error if there is not enough collection study data for the verification study. When a participant takes a collection study, she will be shown a previously-drawn image and given a list of labels, one for each shape in the verification study. She is then asked to choose the label which she believes best fits the image shown. When creating a verification study, the researcher may opt to also include a “none of these” option and/or a blank text field, into which the participant may enter her own label. These options may be helpful for understanding why a shape was mislabeled.

Once the researcher is done creating the verification study, she clicks “submit,” and the information is written to the database. Fig. 4 shows the verification creation page with example values. This new verification study will appear on the study management page underneath the collection study with which it is associated (see Fig. 9). The researcher can

\* Indicates Required Field

<b>Collection Study:</b>	Basic Shapes	
<b>Study Title:*</b>	Verify Shapes	
<b>Study published?*</b>	<input checked="" type="radio"/> Yes <input type="radio"/> No	
<b>Data published?*</b>	<input checked="" type="radio"/> Yes <input type="radio"/> No	
<b>Study Directions:</b>		
<b># of Questions:</b>	1	<input type="button" value="Update"/>
<b>Question #1:*</b>	What is your occupation?	
<b>Shapes</b>		
<i>Description</i>	<i>Use Shape?</i>	<i># to Verify</i>
square	<input checked="" type="checkbox"/>	3
circle	<input checked="" type="checkbox"/>	3
triangle	<input type="checkbox"/>	
<b>Add a "None of these" Option?*</b>	<input checked="" type="radio"/> Yes <input type="radio"/> No	
<b>Add a text field?*</b>	<input checked="" type="radio"/> Yes <input type="radio"/> No	
<input type="button" value="Create New Verification Study"/>		

Fig. 4. Verification user study creation form with example values.

then edit or delete the verification study in the same manner as she edits or deletes a collection study.

### C. Performing a collection user study

To take a public collection study, a participant locates the study she wishes to take by searching for keywords in the study's title or by browsing in public studies. To take a private collection study, a participant goes to the unique URL provided to them by the researcher who created the study. In either case, the participant will be taken to a page which will ask them the pre-study questions specified by the researcher. Once the participant has answered these questions, she hits "submit" and the Java applet will load. At this time, the user is added to the database and their questions are stored.

Once the Java applet has loaded, the participant will be asked to draw an example of the first randomly selected shape. If images were added for this shape, one will be picked at random and displayed next to the textual description. The participant will then draw an example of the desired shape. Fig. 5 shows the Java applet after the participant has drawn an example of the shape for which she has been prompted. She has the option to hit "save" and move on, or to clear the drawing panel and re-draw the image. Once the participant hits "save," the data associated with that sketch is saved to a file, `<shape description><user number>-<sketch number>.xml`, located inside that shape's directory in the study's main directory. Data is saved as `<x,y,time>` tuples into this XML document. The XML document is created by having the applet send a POST

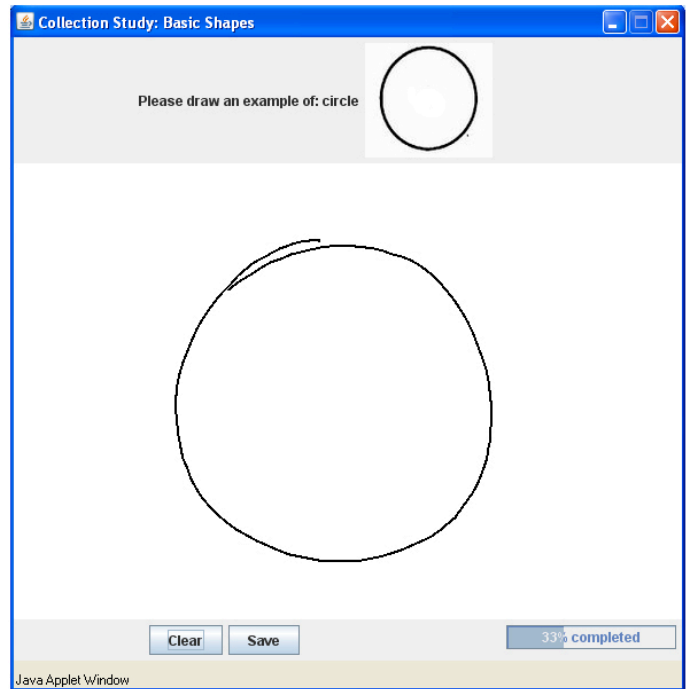


Fig. 5. Example view during collection user study.

command, along with all the necessary data, to a PHP script on our server, which then writes the XML document. Future work on the draw panel will hopefully add additional information, such as pen tilt and pressure.

### D. Performing a verification user study

A participant locates a verification study in the same way as she would locate a collection study. Again, the participant is taken to a page which will ask her the pre-study questions, then to the Java applet. Once the applet has loaded, the participant will be shown an image drawn during a collection study, and asked to choose the label she thinks most accurately describes the image. If the researcher has chosen to include a "none of these" option, this will be listed along with the other labels. If the researcher has chosen to include a text box, the participant can enter her own textual description of the image. The participant may choose multiple labels, including the "none of these" option and the text box. Fig. 6 is an example of what the participant will see during the verification study. Once the participant saves the label, the image shown and the response are written to the database.

### E. Managing Studies

A researcher's study management page includes all studies that she has created. A researcher can easily download the data associated with a specific study. This download (a .zip file) will include data for the collection study as well as for all associated verification studies. For a collection study, an xml file, `<study ID>-collectionData.xml`, details information gathered from participants taking the study, including question answers as well as what images they were shown for each sketch they created. Each sketch is saved in an xml file, and these files are sorted into directories for each shape in the collection study.

For each verification study, two xml files are generated. The first, `<study ID>-verificationData.xml`, contains information gathered from participants taking the study, including questions answered as well as sketches they were shown and the response given. The second, `<study ID>-verificationStatistics.xml`, contains current statistics for the verification study, listed by shape, by sketch, and by user.

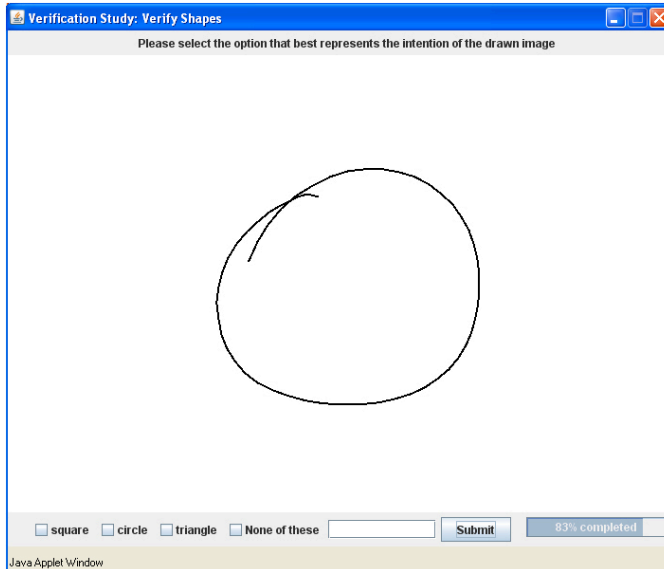


Fig. 6. Example view during verification user study.

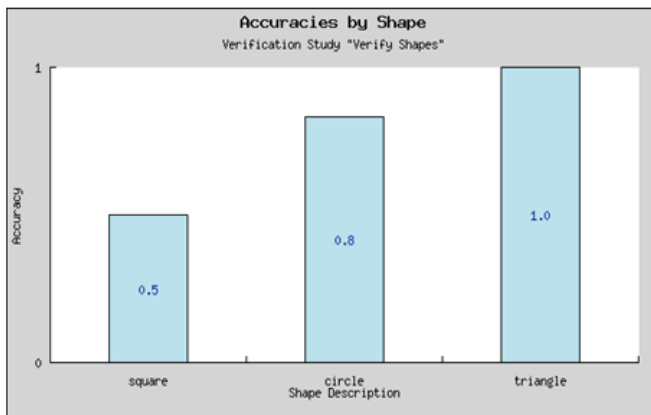


Fig. 7. Example statistics in “view by shape” mode for shape “square” in our example study.

#### Incorrect responses for shape "square"

rectangle(user entered text)  
 long rectangle(user entered text)  
 rectangle(user entered text)

#### Incorrect responses for shape "circle"

spiral(user entered text)

Fig. 8. Alternate responses recorded for shape “square” in our example study; in the web interface, this text appears directly below the graph shown in Fig. 7.

In addition to downloading current data, the researcher can also easily view current statistics for each verification study by clicking on the “View Statistics” link for each verification study. Again, there are three possible views, by shape, by sketch, and by user. Each view will produce a graph or multiple graphs detailing accuracies and will print out a list of incorrect responses. See Fig. 7 for an example graph and Fig. 8 for an example print-out of incorrect responses.

Because study security was a large factor in our work, user management provides the only access to a researcher’s studies. From here, researchers have the ability to create, edit, and delete studies, as well as control whether these studies are published to the public. If a researcher chooses to keep a study private, she will be provided with a unique URL that she will then pass on any intended participants. If a researcher chooses to keep data associated with a study private, she is the only one with access to that data.

## IV. EVALUATION

The study creation and management features of SOUSA are intended for use by researchers in the field of sketch recognition. However, because we hope to allow researchers to gather data from a wide variety of participants, we would like our system to be intuitive for any user with some knowledge of sketch recognition. A preliminary version of SOUSA was used by two graduate students in the Sketch Recognition Lab at Texas A&M University. These students have experience with the original SOUSA and were able to give us feedback about the new features, particularly in comparison to the previous version.

Our system was then evaluated by seven undergraduate research students at Texas A&M University. These students, who have varying experience with sketch recognition, were asked to create a collection study in any domain they wished, to take that collection study as well as a few others, and then to answer a questionnaire. The home page of SOUSA provides online documentation, and these students found the documentation necessary and helpful. Most of the students found the site easy to navigate and use. One student even said, “It was fun and is a black hole for procrastination” when asked to make any additional comments about taking collection studies. Several students were confused about required fields for study creation, and we have since made changes to clearly indicate which fields are required. The pages will also indicate any fields that have been filled out incorrectly at submission time.

## V. FUTURE WORK

The main drawback to our current system is that we lack the ability to archive data associated with an edited or deleted study. If a study is fundamentally changed during an edit, the previous data is associated with the new study and does not reflect the original study from which it was created. Additionally, if a user chooses to delete a study, this data is removed from the system. We hope, in the future, to archive this data, thus maintaining the integrity of the repository while still giving researchers full control over their studies and data.

It would be useful to have a mechanism to save a set of data at a specific moment, so that algorithms being compared could

truly use the same set of data, rather than simply data gathered from the same study.

Further, we would like to give researchers the ability to add externally collected data. This would have two purposes: first, to add data previously collected to the repository; second, to allow for the creation of verification studies without an associated collection study.

## VI. CONTRIBUTIONS

We expect our work to have a broad positive impact on the field of sketch recognition through the creation of a large, robust sketch data repository. Our system will minimize the time researchers spend collecting sketch data for testing purposes, because study creation and management are streamlined and easy to use. Because published data is universally accessible, a researcher can utilize previously collected data for her testing needs. Most importantly, having

a universal set of sketch data in a wide variety of domains will allow new sketch recognition algorithms to be more accurately compared with one another.

## ACKNOWLEDGMENT

The authors would like to acknowledge the guidance of members of the Sketch Recognition Lab at Texas A&M University and specifically Brandon Paulson, Aaron Wolin, and Joshua Johnston for their work on the original SOUSA.

## REFERENCES

- [1] B. Paulson, A. Wolin, J. Johnson, and T. Hammond, "SOUSA: Sketch-based Online User Study Applet," in Proceedings of SBIM, 2008.
- [2] M. Oltmans, C. Alvarado, and R. Davis, "ETCHA Sketches: Lessons Learned from Collecting Sketch Data," *Making Pen-Based Interaction Intelligent and Natural*, 2004.
- [3] R. Blagojevic, B. Plimmer, J. Grundy, and Y. Wang, "A data collection tool for sketched diagrams," in Proceedings of SBIM, 2008.

**Sketch-based Online User Study Application**  
Texas A&M Sketch Recognition Lab

**SEARCH**

Studies  
 Data

**Search**

**NAVIGATION**

[Home](#)  
[Browse Studies](#)  
[Browse Data](#)

**Management: Studies**

Collection Study	Study Published	Data Published	Visitors	Edit	Delete	Verification	Images
Math Signs	Yes (Public)	Yes (Public)	6	<a href="#">Edit</a>	<a href="#">Delete</a>	<a href="#">Create</a>	<a href="#">Upload</a>
<a href="#">Download data for these studies</a>							

Collection Study	Study Published	Data Published	Visitors	Edit	Delete	Verification	Images
Basic Shapes	Yes (Public)	Yes (Public)	7	<a href="#">Edit</a>	<a href="#">Delete</a>	<a href="#">Create</a>	<a href="#">Upload</a>
Verification Study	Study Published	Data Published	Visitors	Edit	Delete	Verification	
Circle Verification	Yes (Public)	Yes (Public)	3	<a href="#">Edit</a>	<a href="#">Delete</a>	<a href="#">View</a>	
Verify Shapes	Yes (Public)	Yes (Public)	8	<a href="#">Edit</a>	<a href="#">Delete</a>	<a href="#">View</a>	
<a href="#">Download data for these studies</a>							

[Create New Collection Study](#)

**REU**

[Manage User Studies](#)  
[My Information](#)  
[Logout](#)

Copyright © 2008 Sketch Recognition Lab, Texas A&M University. All rights reserved. Generated in 1.9 ms

Fig. 9. Web interface and study management page; our example study appears as the second study in this listing, along with its associated verification studies. Note that a collection study can have multiple associated verification studies, and they simply appear in a list following the collection study.