User-defined Interface Gestures: Dataset and Analysis

Daniela Grijincu
School of Computer Science
University of St Andrews
dg76@st-andrews.ac.uk

Miguel A. Nacenta
School of Computer Science
University of St Andrews
mans@st-andrews.ac.uk

Per Ola Kristensson
Department of Engineering
University of Cambridge
pok21@cam.ac.uk

ABSTRACT

We present a video-based gesture dataset and a methodology for annotating video-based gesture datasets. Our dataset consists of user-defined gestures generated by 18 participants from a previous investigation of gesture memorability. We design and use a crowd-sourced classification task to annotate the videos. The results are made available through a web-based visualization that allows researchers and designers to explore the dataset. Finally, we perform an additional descriptive analysis and quantitative modeling exercise that provide additional insights into the results of the original study. To facilitate the use of the presented methodology by other researchers we share the data, the source of the human intelligence tasks for crowdsourcing, a new taxonomy that integrates previous work, and the source code of the visualization tool.

Author Keywords

Gesture design; user-defined gestures; gesture elicitation; gesture analysis methodology; gesture annotation; gesture memorability; gesture datasets.

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation: User Interfaces—*Input devices and strategies*

INTRODUCTION

Gesture-based interfaces are already common in a variety of devices such as game consoles, mobile phones, TVs and public displays and the study of gesture interfaces is a growing area of research in human-computer interaction. Gesture interface research often requires studies with human participants who perform or invent gestures that are later analyzed by researchers. The data can for instance be used to recommend better gesture sets for control of specific systems, or to allow designers to make better decisions about the gestures that they integrate (e.g., [7, 44]).

Gestures collected in experiments are often video recorded, but most of the time the rich data that was collected (for example, the videos of gestures) are only analyzed once and with the specific focus of the paper that motivated the study. Data reuse represents a lost opportunity in research—other researchers can potentially extract new insights from video, and the accumulation of data from different sources can potentially open up new types of analysis, such as quantitative modeling (which typically demands large amounts of data) and meta-analyses. However, sharing high-quality datasets requires significant amounts of work to annotate it for analysis purposes, and to ensure it is easily accessible to other researchers.

In this work we set out to contribute an example of sharing and annotating gesture data and creating a tool that facilitates explorations and further analyses. We acquired the data from our recently published study on gesture memorability [26]. We then designed and developed a crowdsourcing task to annotate the data. The annotated data is made available through an on-line visualization and exploration tool and further analyzed using machine learning techniques.

The main contributions of this paper are:

- a dataset of user-defined gestures;
- a crowdsourced gesture-annotation procedure;
- an integrative taxonomy for hands-and-arms gesture classification;
- an interactive visualization tool to access the annotated dataset;
- a complementary analysis, using machine learning techniques, of the annotated gesture dataset.

We put special emphasis on sharing the annotated videos, the source files that enable the crowdsourcing of the categorization task, and the exploration web application source in order to reduce the effort required of other researchers to share their data and to enable further analysis and research of existing and future gesture datasets.

RELATED WORK

Gesture-based interfaces were first adopted in 2D input devices (see [46] for a recent review) for a number of applications such as issuing commands (e.g., [4, 5, 18]), and writing text (e.g., [3, 10, 17, 27]). The popularisation of 3D sensors and accelerometers has also spurred extensive research on 3D gestures (e.g., [16, 35, 38, 39]) and commercial applications that are controlled via in-air gestures are now commonplace. Three areas of related research bear direct relevance to our work: (1) how gestures are designed or selected for specific

ITS 2014, November 16–19, 2014, Dresden, Germany.

ACM 978-1-4503-2587-5/14/11 ...$15.00.
http://dx.doi.org/10.1145/2669485.2669511

applications, (2) gesture memorability, and (3) existing gesture sets, taxonomies and classifications.

Gesture Design Process

The variety of distinct gestures that could be recognized by a gesture-based system is very large, even in gesture-based interfaces that only depend on 2D input. A suitable gesture subset must then be defined and selected, for which several research groups have provided tools (e.g. [15, 19, 20, 45]). These tools allow interface designers and implementers to easily create gesture sets that are recognizable by the system, but they do not address the question of how to choose gestures that are appropriate for end-users.

Based on the concept of user elicitation [28], Wobbrock et al. [44] proposed to define gesture sets for a multitouch interface via a systematic process in which a representative sample of the intended end-users are asked to create gestures for specific actions. The final gesture set is the set with maximum user agreement in the sample of the intended end-users. The motivation behind gesture elicitation is that it results in gesture sets that are better related to the actual needs and expectations of the intended users of the system [25]. Gesture elicitation has often been used for on-air gestures [37], multi-touch gestures [7, 8, 22, 28, 44], unistroke gestures [43], pen-gestures [9], foot gestures [1] and mobile motion gestures [35]. For a discussion on the advantages and possible biases of the technique see also [24].

In addition to gesture sets chosen *a priori* by designers (e.g. [23]), or defined *a priori* by a group of participants in a gesture elicitation exercise, gestures can also be defined by users themselves. This approach has several advantages, including better accessibility for people with impairments [2, 26], better memorability [26], familiarity for the access of information [30], and might also offer added expressivity in artistic interfaces [21] (limitations of this approach are discussed in [26] and [29]). This paper focuses on a dataset of gestures that were user-defined and meant to be remembered (a subset of the gestures in [26]).

Gesture Memorability

Researchers studying gesture-based interfaces have focused on the study of several desirable characteristics of gesture sets, including learnability, discoverability, immediate usability [19], and memorability [26].

This paper presents data focused on gesture memorability, which is a critical characteristic of gesture sets. Prior work on gesture memorability includes Cockburn et al. [6], who in the context of single-stroke gestures studied the effect of inducing effort on overall gesture memorability. Their experiments showed that an effort-inducing interface improved recall rates but it was also considered less enjoyable due to the additional effort. A study by Appert and Zhai [4] analyzed how users learn and remember keyboard shortcuts compared to gestures, reporting that the latter category was more memorable. More recently, Jansen [14] investigated how well users can remember interaction gestures for controlling a TV set, based on different teaching methods. Jansen [14] reports that the gestures perceived as most intuitive by the participants were associated with objects from the TV interface or with the gestures task description. Jansen [14] states that users can correctly remember 61-71% of ten (for them unfamiliar) user-elicited gestures depending on the teaching method used to train them. However, there is no information on the specific features that characterize the most memorable gestures.

In this paper we continue Nacenta et al.'s [26] gesture memorability work, which investigated how participants remember three free-form gesture sets designed for three different applications: a word processor, an image editor and a web browser. For each application, the authors chose 22 different actions for which to design the gestures that are currently still representative for most existing activities available on software systems. Nacenta et al.'s [26] study distinguishes between pre-designed gestures (created by designers), stock gestures that are system generic (i.e. not designed for an specific action), and user-defined gestures that are created by the users themselves. A gesture was considered correctly recalled (or memorable) if it was reproduced with the same hands, fingers and overall path and timing. If the gesture was recalled with small articulation differences (e.g. different finger) but maintaining most of the characteristics of the original gesture, then it would be judged as *close*. Otherwise, if too few similarities were found between the original and the reproduced gesture, it would be considered as incorrectly recalled by the participant. The experiments conducted in the study revealed that self-defined gestures were more memorable (up to 44% more gestures recalled than with the pre-designed gestures) and preferred by users. Although these results are important, they offer little guidance regarding the specific characteristics that make gestures memorable. This work focuses on enabling further analysis of memorability through the annotation, sharing and descriptive analysis of the set of gestures that were generated by participants in [26].

Taxonomies and classifications

The first step to understand why some gestures are better than others is to define criteria (dimensions) that separate those gestures into relevant categories. Multiple such categorizations exist. Quek [33] provides a well-cited classification scheme that differentiates acts (where the gesture's shape and movement is directly related to the intended action) from symbols (where the gesture is arbitrary, but linked to its referent by a language agreed-upon in advance). Pavlovic [31] expands on this work by adding manipulative gestures and putting Quek's [33] categorization within the broader context of human movements. Focusing on the quality and morphology of gestures, Laban [41] created a notation system to describe dance performances that can be applied to description of human gestures. Dimensions derived from Laban's include distinctions of speed, continuity and effort of gestures. Vatavu and Pentiuk [40] and Ruiz et al. [35] describe categorizations that distinguish between information contained in posture and movement.

A more recent taxonomy is the one created by Wobbrock et al. [44] for their seminal elicitation study of multi-touch gestures. They manually classified the gestures obtained into different dimensions such as form, nature, binding and flow,

which allowed them to point out that participants of their study preferred single hand gestures to two-hand gestures and that their gestures are influenced by desktop idioms. Simpler categorizations are implicit in the work of Epps et al. [7], who distinguished among common hand shapes and pointing behaviors (e.g. L-shape, C-shape, fist), and Jansen [14], who identified a set of features relevant for interactive TVs.

Available Gesture Sets

In addition to the final gesture set elicited by Wobbrock et al. [44], others have shared proposed gesture sets for specific applications (e.g. multi-touch graph manipulation [36], interaction with omni-directional video [34]), and also stock sets for use in any application [11]. These are generally described through sequences of illustrations, and only occasionally through video. To our knowledge, we are the first to share a full data set of videos as created by participants for specific actions. Moreover, we also provide the labeling (categorization) of each gesture according to a new classification scheme that summarizes existing taxonomies, enabling further research and analysis of the gesture set.

GESTURE CLASSIFICATION METHOD

The starting point of our work is a gesture corpus. We use the self-defined subset of Nacenta et al.'s third study in [26], which consists of 396 user-defined, video recorded 3D gestures performed by 18 participants for 66 different actions (22 actions each). The gestures were created by the participants so that, on request, they could trigger a set of 66 different actions that were selected from common features out of three applications: a web browser, a word processor and an image editor. Each participant only created gestures for one of the three applications. Participants knew that they would be asked to remember the gestures at the time of gesture creation. As a result of the experiment, each gesture was annotated with a label that indicates whether the gesture was remembered in a subsequent test. This memorability label has three possible values: memorable (the gesture was remembered), not-memorable (the gesture was not remembered) or close (the gesture was only partially remembered). Additionally, due to the structure of the experiment, each gesture is annotated with a label that indicates whether or not the gesture was reinforced. Reinforced gestures comprised half of the gesture set that a participant would have to create (11 gestures), but instead of being tested only on the next day, reinforced gestures were also tested (with correctness feedback) immediately after all gestures were created. See their original experiment paper [26] for details.

In most previous studies that classify user-defined gestures the authors themselves classified the gesture sets [44, 14]. This approach does not scale well as the number of gestures go up and is also subject to a possible strong author bias. In this paper we propose an alternative method suitable for annotating large sets of gestures by using a crowdsourcing service to obtain the classifications.

Gesture Taxonomy

The purpose of the categorization (enabling further research) and the chosen methodology imposes a number of constraints that no existing gesture categorization covers completely. This forced us to create a new taxonomy of gestures. Specifically, the new taxonomy has to:

- be descriptive of the type of gestures that people actually perform;
- describe the morphological aspects of the gesture (e.g. number of fingers);
- describe semantic aspects of the gesture (how it relates to the action);
- capture differences in gesture complexity;
- summarize the state of the art (other taxonomies);
- be unambiguous and easy to classify even by non-experts; and
- be succinct.

To achieve these goals we started by compiling a classification schema based on previous gesture taxonomies, with an eye on aspects that could have an impact on gesture memorability. To make sure that the taxonomy was as descriptive as possible of the gestures' features, we manually classified a random sample consisting of 10% of the corpus and used the distribution of the gestures' features to reduce redundancy and ambiguity in the final classification. For example, we noticed from the videos that it was very difficult to discriminate different speeds of movement (aspect proposed by Laban [41]), and as a result, most gestures appeared to have the same, normal (slow) speed. Therefore we decided to omit this dimension from our schema in order to avoid confusing the crowd workers. The final taxonomy, which is displayed in Table 1, is also the result of further modifications after initial piloting of the crowdsourcing process, which is described in the next subsection. Notice that some dimensions are contingent on the category selected in another dimension.

Crowdsourcing

We chose the CrowdFlower platform to classify this corpus because of its large and varied worker base, its sound quality control service, the availability of an infrastructure to train workers using test questions and control their reliability, and the ability to receive feedback from workers on the difficulty and ambiguity of the tasks.

We programmed a web-based application that implemented a gesture classification task based on our taxonomy. The design of the Human Intelligence Task (HIT) interface is critical: if the task unnecessarily hard or confusing it would directly affect the quality of the collected data. Further, a poor HIT design can disincentivize workers to participate [13]. Therefore, we invested a significant amount of effort in designing the interface (see Figure 1). For example, the application was designed to minimize scrolling so that crowd workers could play the gesture in a video any number of times while deciding on each of the classification sub-tasks. Since there are more classification sub-tasks than what can fit on a typical computer screen, these were scrolled horizontally on a button-activated content slider that showed one dimension at

Dimension	Features	Provenience	Choice Type
1. Localization	in air, on surface, mixed	new	single
2. Number of hands	unimanual, bimanual symmetric, bimanual asymmetric	W. et al. [44], P. et al. [32]	single
3. Hand form (HF)	spread, flat, mixed, other	Jansen [14], E. et al. [7]	single
3.1. Hand orientation (if 3 is spread, flat or mixed)	horizontal, vertical, mixed	Jansen [14]	single
4. Additional hand forms	single index finger, single other finger, multiple fingers, fist, grab-release sequence, C-shape, L-shape, other	E. et al. [7]	multiple
5. Hand form and path	same hand form, multiple hand forms, same hand form and path, multiple hand forms and path, mixed	W. et al. [44], V. et al. [40]	single
6. Gesture path	straight, flexible, n/a (no path)	Laban [41]	single
6.1. Gesture path flow (if 6 is straight or flexible)	continuous, segmented	Laban [41]	single
6.2. Gesture path shape (if 6 is flexible)	open, closed, mixed	new	single
7. Relation to gesture action DR - directly related or NR - not related	alphabet letter or number DR, shape of an object DR, arbitrary NR	W. et al. [44], Q. [33], R. et al. [35]	single
8. Relation to gesture workspace	object dependent, workspace dependent, independent	W. et al. [44], R. et al. [35]	single
9. Gesture meaning	metaphorical, symbolical, abstract	W. et al. [44], R. et al. [35]	single

Table 1. Proposed Gesture Taxonomy. 'same hand form' in (5) corresponds to Wobbrock et al.'s static hand pose, but the original formulation proved difficult to understand for crowd workers. 'multiple hand forms' correspond therefore to dynamic hand poses.

a time. Because some of the dimensions involve movement (e.g. additional hand forms, hand form and path in Table 1), we included simple animations that exemplified certain gesture features. The interface forced an answer for all classification subtasks, but tasks that were contingent on other responses only appeared when appropriate (e.g. the hand orientation dimension is shown only if the hand form is spread, flat or mixed).

Some categorizations require semantic judgments of the relationship between the action and the gesture. For this the interface provides, next to the video, the name of the target action and snapshots of the state of the application before the action is executed, and after the action is triggered. This is analogous to how the videos were presented in the original study when the gestures were created.

Test questions are the most important quality control mechanism on the CrowdFlower platform and the best way to ensure that high quality data is acquired. We randomly transformed 30 of the total number of recorded videos into test questions, using CrowdFlower's interface for creating test questions. For each test question, we selected the acceptable answers, which crowd workers would have to answer correctly while working on the task. Special care was taken to ensure an even distribution of test question answers, and also to skip over gesture dimensions which would be too difficult or subjective. CrowdFlower also offers a method called Quiz Mode that ensures that only contributors who understand the task are able to contribute to the task output. By default, and if Quiz Mode is enabled in the settings, CrowdFlower requests that workers complete four test questions before their trust score can be evaluated and their judgements can be incorporated into the task output. For our task on CrowdFlower, the trust threshold was set to 70%. Contributors that were not able to pass the Quiz Mode were permanently disqualified from working on the task. However, even workers that passed the Quiz Mode were still continuously tested as they worked in the task through the method called Work Mode, which keeps the workers trust scores updated and ensures a sustained quality of their answers.

The crowd workers classified five videos per block (for a US$.80 payment), with at least three crowd workers classifying each video. When multiple workers disagreed on the categorization, the chosen answer was the majority vote weighted by the trust scores of the workers.

The task was first piloted on local students, and then piloted again on a small set of crowd workers. This led to final adjustments of the interface and the taxonomy.

DATASET EXPLORATION TOOL

In the spirit of open data and replication in HCI [42], one of the main goals of our work is to enable further analysis of the annotated dataset that we are sharing. Although sharing experimental data is beginning to become an accepted (and encouraged) practice, we find that shared data often requires too much initial work to even consider using it. To ameliorate this problem and reduce the access threshold, we have cre-

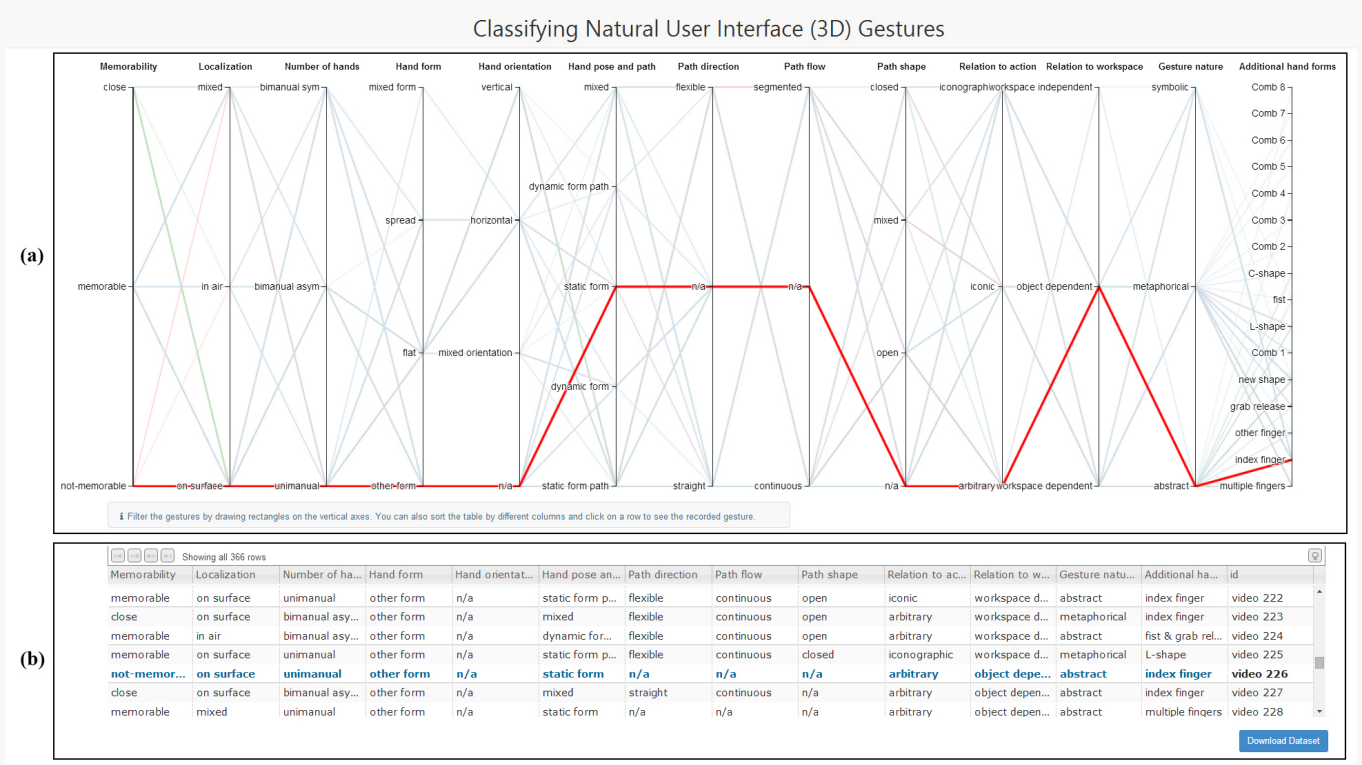

Figure 2. Data exploration tool. (a) presents the parallel coordinates visualization of different gesture dimensions and features; and (b) shows a table containing information about the annotations of the gestures, allowing filtering interaction

ated a web application that provides access to all the videos, represents the classification data, and allows interactive exploration of the dataset.

The tool, which can be accessed online at `http://udigesturesdataset.cs.st-andrews.ac.uk/`, has three main connected components: a parallel coordinates panel, an interactive spreadsheet, and a gesture page. The parallel coordinates visualization [12] represents each gesture as a line that crosses each of the 13 vertical axes at a different height depending on the classification of the gesture (see Figure 2.a). The first axis represents the memorability value (whether the gesture was ultimately remembered or not) and each of the remaining axes represents one of the 12 dimensions of our taxonomy. This visualization is implemented with D3[1] and Parcoords[2]. Gestures can be selected by filtering in one or more dimensions, and dimensions can be reordered via pick-and-drop. Filtering and picking is interactively connected to the spreadsheet section, so that only filtered gestures will appear in the spreadsheet panel.

The spreadsheet (Figure 2.b) allows direct access to the data. Each row represents a gesture and each column corresponds to a dimension. Hovering over a row will highlight the corresponding gesture in the parallel coordinates area above. The spreadsheet widget is implemented using SlickGrid[3]. Clicking on a row opens a gesture page (Figure 3), in which the video of the gesture can be played and all the information about the gesture also appears (classifications, memorability, intended action, and application). Finally, the application contains a button that triggers the download of the raw data.

DESCRIPTIVE ANALYSIS

For all 396 recorded videos from the corpus we received a total of 1783 judgments, of which 1270 were considered trusted and 513 untrusted (this count does not include judgments on test questions). This results on an average of 9.63 crowd workers classifying each gesture. Because all gestures were judged by several workers, we also obtained a measure of inter-rater reliability (IRR) for each of the different dimensions (see the IRR scores in Table 2). These values represent the average across all gestures of the percentage of workers that agreed with the algorithmic classification described at the end of the *Crowdsourcing* subsection above.

Our IRR is a measure of the trustworthiness of the overall classification. Moreover, the independent IRR measures can serve as a proxy measure of how ambiguous/difficult the different dimensions are to categorize. As expected, dimensions that required semantic judgments (8 and 9 from Table 1) had more modest agreement rates. Additionally, dimension 6.1 (whether a gesture's path is open, closed, or contains the two types) was also particularly difficult. Nevertheless, all IRR values are above 70%.

In total,185 crowd workers completed the task and 60 of them sent us feedback. Feedback from crowd workers indicated

[1]http://d3js.org/
[2]http://syntagmatic.github.io/parallel-coordinates/
[3]https://github.com/mleibman/SlickGrid/wiki/

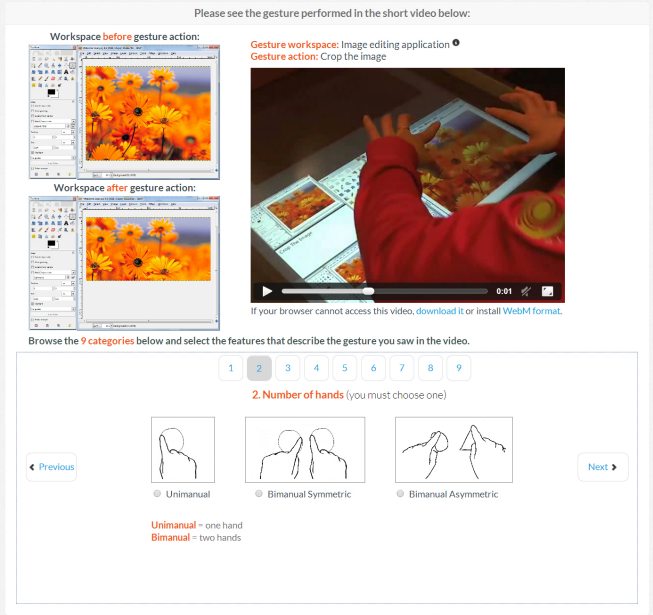

Figure 1. Gesture classification task implemented on CrowdFlower

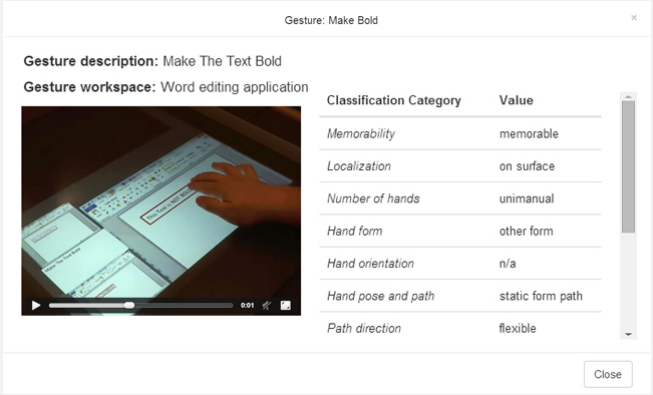

Figure 3. Data exploration tool showing the gesture performed

that they considered the payment fair (4.2 out of 5), that the instructions were clear (3.8 out of 5), that the test questions were fair (3.5 out of 5), and that the task was difficult (2.8 score out of 5 for the ease of the task). The final average worker trust score (computed by CrowdFlower based on their performance on test questions and work in the task) was 0.892.

Gesture Categorization Results

The gestures classified in each dimension for all features are displayed in Table 2. The results indicate that some dimensions are heavily biased towards a particular type of gesture (e.g., most gestures use only the index finger, one hand, or continuous path flows), whereas other dimensions are more balanced (e.g., similar number of gestures are object and workspace dependent, metaphorical and abstract).

Some of these results support previous findings [7, 44, 39]. It is remarkable that even after more than five years since Wobbrock et al's [44] study, with multi-touch interfaces being now pervasive, there is still such a strong bias towards single finger, single hand gestures.

STATISTICAL MODELING

The labeling procedure described above and the discrete nature of the outcome variable of the gestures (memorable, not-memorable, and close) make the dataset amenable to classification algorithms that might be able to provide additional insight into the factors that affect memorability. The starting point is the 43-dimensional feature vector for each gesture, where each value is binary and represents each of the possible gesture classifications of the taxonomy (Table 1). The dependent variable (memorability) has three possible values: memorable, non-memorable, and close, which were heavily imbalanced (312, 39, and 45 cases respectively).

Feature Selection

As some of the features could potentially introduce noise, we decided to pre-select the most relevant subset of features for describing the dataset. We tested combinations of the nine gesture classification dimensions (Table 1), generating 510 possible subsets (the empty and the complete set were excluded). Each generated subset was evaluated based on a 10-fold cross-validation accuracy obtained with Support Vector Machines (SVMs) trained with the Radial Basis Function kernel and parameters C and γ obtained via a grid search. The best results were obtained with a selection of 28 features corresponding to dimensions 2, 4, 6, 6.1, 6.2, 7, 8 and 9 from the classification schema described in Table 1. We generated the feature vectors with the new selection of features and updated the initial training set.

Modeling Approaches

We attempted to predict the memorability label of gestures using different supervised machine learning techniques suitable for a 3-class problem and training using binary feature vectors. We first experimented with Logistic Regression using the Weka[4] open source software for data mining and measured the 10-fold cross-validation results. We report the weighted average scores between the three classes of gestures obtained for different measurements in Table 3. Table 3 reveals that the results for predicting gesture memorability using logistic regression were not satisfactory as the learned model was highly biased towards predicting only memorable gestures, which represented over 70% of the training samples and for which the false positive rate was very high (over 70%).

Decision tree classifiers were another suitable technique for our problem, having the additional benefit of being relatively easy to visualize and interpret. We experimented with a range of different types of decision trees available in Weka such as Best-First Tree, ID3, J48 and also random trees and random forests. The best results were obtained with the ID3 algorithm, which achieved a modest F-score of 0.64. This is only a small improvement with respect to the 10-fold cross-validation scores obtained with logistic regression.

[4] http://www.cs.waikato.ac.nz/ml/weka/

Finally, we tried Support Vector Machines (SVM) in their `libSVM` implementation[5]. The best results were obtained with a Radial Basis Function kernel and with C and γ parameters obtained via grid search (we also tried multiple kernel functions such as linear, polynomial, string based and sigmoid). Table 3 reveals that SVMs performed slightly better, the F-score was 0.68, and overall accuracy was 70%.

DISCUSSION

We set out to achieve a better understanding of the data obtained in [26] through an analysis and categorization of the dataset. To achieve this, we chose to take advantage of crowdsourcing as a tool for the categorization of videos. We believe that this approach has many advantages and the potential to significantly further HCI research on gestures if more researchers analysed (and shared) the obtained gestures in this manner.

The approach is more scalable than expert classification and much faster—once the HIT was prepared it only took 13 hours for the full task to be completed. We were initially concerned that crowd workers might not be able to achieve acceptable consensus about certain dimensions. Specifically, we were concerned about the dimensions that involved semantic judgements. However, the results indicate that the classifications are of a high quality (all IRR $>$70%). We suspect that this is due to the careful setup of the HIT and the simplified taxonomy. It is conceivable that groups of gesture experts can perform categorizations with higher agreement, but this kind of expertise is not easy to identify and recruit, and we conjecture that the results would be largely the same and orders of magnitude more expensive.

We have included our templates in this paper's complementary materials so that other researchers can take advantage of our work and annotate and share their gestures at a lower cost and effort.

Taxonomy

The taxonomy is a critical element of both the research and methodological parts of this work. We designed it to balance generalizability and detail from previous work with the simplicity and conciseness required for a viable human intelligence task. The measures of agreement from the crowd worker judgements show that the taxonomy enables consistent categorizations, although some of the dimensions imply harder distinctions. Specifically, the two dimensions that imply semantic judgements (8, 9) or gesture path shape (6.2) were between 70% and 80% agreement.

Although we created this taxonomy to learn more about memorability and gestures, we believe it captures most of the morphological and semantic characteristics of arm and hand gestures. Therefore we believe that this taxonomy can be useful to researchers in this area as an alternative that is: a) very expressive, b) concise, and c) validated for classification through crowdsourcing. For explicit purposes beyond what the current version can describe and classify, the taxonomy can be extended with new dimensions.

[5]http://www.csie.ntu.edu.tw/ cjlin/libsvm/

Types of gestures

The distributions of gestures that we obtained offer an interesting view of current user expectations regarding gestures on and above surfaces. First, the preference for single hand, single finger gestures suggests that the conceptualization of "finger as cursor" is still strong and common actions would probably benefit from this kind of gesture. Second, in presence of a display surface, people still prefer to design gestures that come in contact with the surface, even when they were explicitly told that this was not a constraint (see [26]). Third, although in-air dynamic gestures can be extremely rich in form, which might be perceived as useful for memorability, the chosen gestures were heavily biased towards simple hand postures that tended to be unchanging (gesture changes are uncommon) and with the hands describing simple paths (i.e. non-segmented and open).

We believe this rich gesture data can support gesture interface designers. For instance, using our exploration tool a designer can get an idea of how naturally users propose certain gestures, or get inspired when creating advanced gestures by exploring the less common gestures created by the participants.

Modeling and its Challenges

The annotation of the data made possible to explore the data using statistical modeling techniques. Our intention was to identify models that would be able to predict memorability based on the characteristics of a gesture. This approach was only partially successful in that, although there is some information in the annotation that allows the best algorithm (SVMs) to classify with performance above chance, the predictions are not sufficiently accurate. Their performance is relatively close to assuming that all gestures will be memorable.

There are several factors that explain why the prediction ability of the models is low. First, although our dataset is substantial in size for this kind of study, it is not large compared to the typical dataset sizes required for accurate statistical modeling. Second, the data is heavily biased towards memorable gestures, which makes modeling difficult. Third, memorability is inherently difficult to predict, with many possible intervening sources of noise, including individual differences and personal experience.

Nevertheless, a detailed look at the more interpretable models (specifically, decision trees), and the results of the feature selection stage (discussed in the *Modeling* Section) suggest that some dimensions contain more information than others towards explaining memorability. Specifically, we suspect that three gesture dimensions deserve particular attention: gesture meaning, gesture path, and gesture flow.

Summary of insights

The following points summarize the main insights obtained from the annotation process and analyses presented in this paper:

- Crowdsourcing is a fast and reliable tool for gesture categorization.

Dimension	Features	#	%
1. Localization (IRR 89%)	on surface	353	89
	in air	22	6
	mixed	21	5
2. Number of hands (IRR 98%)	unimanual	285	72
	bimanual symmetric	56	14
	bimanual asymmetric	55	14
3. Hand form (IRR 89%)	flat	98	25
	spread	25	6
	mixed	6	2
	other	267	67
3.1 Hand orientation (IRR 85%)	horizontal	76	19
	vertical	30	8
	mixed	23	6
4. Additional hand forms (IRR 89%)	single index finger	205	52
	single other finger	3	1
	multiple finger	48	12
	grab-release	12	3
	c-shape	8	2
	l-shape	19	5
	fist	5	1
	other	100	25
5. Hand form and path (IRR 82%)	same HF	41	10
	multiple HF	26	7
	same HF and path	289	73
	multiple HF and path	12	3
	mixed	28	7
6. Gesture path (IRR 87%)	straight	204	52
	flexible	107	27
	n/a	85	21
6.1 Gesture path flow (IRR 88%)	continuous	228	56
	segmented	83	21
6.2 Gesture path shape (IRR 74%)	open	74	19
	closed	17	4
	mixed	16	4
7. Relation to gesture action (IRR 81%)	alph. letter or nr DR	51	13
	shape DR	64	16
	arbitrary NR	281	71
8. Relation to gesture workspace (IRR 78%)	object depend.	173	44
	workspace depend.	219	55
	independ.	4	1
9. Gesture meaning (IRR 72%)	metaphorical	171	43
	symbolical	15	4
	abstract	210	53

Table 2. Distributions of the gesture annotations obtained from CrowdFlower

Method	TP	FP	Prec	Rec	F1	ROC	Acc %
Logistic Regression	0.65	0.61	0.62	0.65	0.63	0.54	65
ID3	0.67	0.6	0.61	0.67	0.64	0.56	68
SVMs	0.70	0.58	0.67	0.70	0.68	0.59	70

Table 3. Performance measures obtained using the best classifiers. From left to right: true positives, false positives, precision, recall, F-score, receiver operating characteristic and accuarcy.

- Crowdsourcing categorization results are reasonably reliable.
- Users still predominantly define single-point and single-hand gestures.
- Users tend to create gestures where the hand form does not change.
- Gesture meaning, path and flow are promising characteristics in the further study of memorability of gestures.

Limitations and Future Work

The work that we present is based on data obtained through a memorability study. Participants created gestures with the specific goal of remembering them afterwards. Although we tried to create a process and taxonomy that generalizes to other desirable features beyond memorability, such as discoverability and performance, some of the findings are inevitably biased by the memorability focus of the experiment that generated the original data. Future annotations of user-defined gestures, including those from elicitation studies, can help build up a consistent knowledge base on the types of gestures that people naturally propose.

Although we provide materials in the additional files that should simplify categorization of other gesture datasets through crowdsourcing, the setup require additional work. Videos and existing data still need to be linked to the new HIT and the interface tested, even if no modifications to the classification task are required. In the future it would be useful to develop a platform that simplifies the classification and storage of gesture collections of several types.

CONCLUSION

Gesture research can benefit from more researchers sharing and annotating their datasets. In this work we show an example of how an existing video gesture dataset can be annotated using crowdsourcing tools and then made available for exploration through a visualization tool. In addition to the methodological contributions, which include a revised integrative taxonomy of arm and hand gestures, we show how further analysis of the annotated data can lead to a better understanding of gestures that go beyond the analysis offered in the original paper. Our dataset can be accessed here: `http://udigesturesdataset.cs.st-andrews.ac.uk/`.

REFERENCES

1. Alexander, J., Han, T., Judd, W., Irani, P., and Subramanian, S. Putting your best foot forward: Investigating real-world mappings for foot-based gestures. In *Proceedings of the 30th International Conference on Human Factors in Computing Systems*, CHI '12 (2012).
2. Anthony, L., Kim, Y., and Findlater, L. Analyzing user-generated youtube videos to understand touchscreen use by people with motor impairments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, ACM (New York, NY, USA, 2013), 1223–1232.

3. Anthony, L., and Wobbrock, J. O. $n-protractor: A fast and accurate multistroke recognizer. In *Proceedings of Graphics Interface 2012*, GI '12, Canadian Information Processing Society (Toronto, Ont., Canada, Canada, 2012), 117–120.
4. Appert, C., and Zhai, S. Using strokes as command shortcuts: Cognitive benefits and toolkit support. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, ACM (New York, NY, USA, 2009), 2289–2298.
5. Bau, O., and Mackay, W. E. Octopocus: A dynamic guide for learning gesture-based command sets. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*, UIST '08, ACM (New York, NY, USA, 2008), 37–46.
6. Cockburn, A., Kristensson, P. O., Alexander, J., and Zhai, S. Hard lessons: Effort-inducing interfaces benefit spatial learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, ACM (New York, NY, USA, 2007), 1571–1580.
7. Epps, J., Lichman, S., and Wu, M. A study of hand shape use in tabletop gesture interaction. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '06, ACM (New York, NY, USA, 2006), 748–753.
8. Findlater, L., Lee, B., and Wobbrock, J. Beyond QWERTY: Augmenting touch screen keyboards with multi-touch gestures for non-alphanumeric input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, ACM (New York, NY, USA, 2012), 2679–2682.
9. Frisch, M., Heydekorn, J., and Dachselt, R. Investigating multi-touch and pen gestures for diagram editing on interactive surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, ACM (New York, NY, USA, 2009), 149–156.
10. Goldberg, D., and Richardson, C. Touch-typing with a stylus. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, ACM (1993), 80–87.
11. Hotelling, S., Strickon, J., Huppi, B., Chaudhri, I., Christie, G., Ording, B., Kerr, D., and Ive, J. Gestures for touch sensitive input devices, July 2 2013. US Patent 8,479,122.
12. Inselberg, A. Multidimensional detective. In *Proceedings of the 1997 IEEE Symposium on Information Visualization (InfoVis '97)*, INFOVIS '97, IEEE Computer Society (Washington, DC, USA, 1997).
13. Jacques, J. J., and Kristensson, P. O. Crowdsourcing a hit: measuring workers' pre-task interactions on microtask markets. In *Proceedings of the 1st AAAI Conference on Human Computation and Crowdsourcing*, HCOMP '13, AAAI Press (Menlo Park, CA, USA, 2013), 86–93.
14. Jansen, E. Teaching users how to interact with gesture-based interfaces; a comparison of teaching-methods. diploma thesis, University of Technology Eindhoven, 2013.
15. Kin, K., Hartmann, B., DeRose, T., and Agrawala, M. Proton: Multitouch gestures as regular expressions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, ACM (New York, NY, USA, 2012), 2885–2894.
16. Kristensson, P. O., Nicholson, T., and Quigley, A. Continuous recognition of one-handed and two-handed gestures using 3D full-body motion tracking sensors. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces*, IUI '12, ACM (New York, NY, USA, 2012), 89–92.
17. Kristensson, P. O., and Zhai, S. SHARK^2: A large vocabulary shorthand writing system for pen-based computers. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, UIST '04, ACM (2004), 43–52.
18. Kristensson, P. O., and Zhai, S. Command strokes with and without preview: Using pen gestures on keyboard for command selection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, ACM (New York, NY, USA, 2007), 1137–1146.
19. Long, Jr., A. C., Landay, J. A., and Rowe, L. A. Implications for a gesture design tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, ACM (New York, NY, USA, 1999), 40–47.
20. Lü, H., and Li, Y. Gesture coder: A tool for programming multi-touch gestures by demonstration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, ACM (New York, NY, USA, 2012), 2875–2884.
21. Merrill, D. J., and Paradiso, J. A. Personalization, expressivity, and learnability of an implicit mapping strategy for physical interfaces. In *In the Extended Abstracts of the Conference on Human Factors in Computing Systems (CHI05* (2005), 2152–2161.
22. Micire, M., Desai, M., Courtemanche, A., Tsui, K. M., and Yanco, H. A. Analysis of natural gestures for controlling robot teams on multi-touch tabletop surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, ACM (New York, NY, USA, 2009), 41–48.
23. Microsoft, Inc. Using gestures. `http://bit.ly/OXOvT4`, June 2014.
24. Morris, M. R., Danielescu, A., Drucker, S., Fisher, D., Lee, B., schraefel, m. c., and Wobbrock, J. O. Reducing legacy bias in gesture elicitation studies. *interactions 21*, 3 (May 2014), 40–45.

25. Morris, M. R., Wobbrock, J. O., and Wilson, A. D. Understanding users' preferences for surface gestures. In *Proceedings of Graphics Interface 2010*, GI '10, Canadian Information Processing Society (Toronto, Ont., Canada, Canada, 2010), 261–268.

26. Nacenta, M. A., Kamber, Y., Qiang, Y., and Kristensson, P. O. Memorability of pre-designed and userdefined gesture sets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, ACM (New York, NY, USA, 2013), 1099–1108.

27. Ni, T., Bowman, D., and North, C. Airstroke: Bringing unistroke text entry to freehand gesture interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, ACM (New York, NY, USA, 2011), 2473–2476.

28. Nielsen, M., Störring, M., Moeslund, B., and Granum, E. A procedure for developing intuitive and ergonomic gesture interfaces for hci. In *Gesture-Based Communication in Human-Computer Interaction*, vol. 2915 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2004, 409–420.

29. Oh, U., and Findlater, L. The challenges and potential of end-user gesture customization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, ACM (New York, NY, USA, 2013), 1129–1138.

30. Ouyang, T., and Li, Y. Bootstrapping personal gesture shortcuts with the wisdom of the crowd and handwriting recognition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, ACM (New York, NY, USA, 2012), 2895–2904.

31. Pavlovic, V. I., Sharma, R., and Huang, T. S. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence 19*, 7 (July 1997), 677–695.

32. Piumsomboon, T., Clark, A., Billinghurst, M., and Cockburn, A. User-defined gestures for augmented reality. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, ACM (New York, NY, USA, 2013), 955–960.

33. Quek, F. K. H. Toward a vision-based hand gesture interface. In *Proceedings of the Conference on Virtual Reality Software and Technology*, VRST '94, World Scientific Publishing Co., Inc. (River Edge, NJ, USA, 1994), 17–31.

34. Rovelo Ruiz, G. A., Vanacken, D., Luyten, K., Abad, F., and Camahort, E. Multi-viewer gesture-based interaction for omni-directional video. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, ACM (New York, NY, USA, 2014), 4077–4086.

35. Ruiz, J., Li, Y., and Lank, E. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, ACM (New York, NY, USA, 2011), 197–206.

36. Schmidt, S., Nacenta, M. A., Dachselt, R., and Carpendale, S. A set of multi-touch graph interaction techniques. In *ACM International Conference on Interactive Tabletops and Surfaces*, ACM (2010), 113–116.

37. Seyed, T., Burns, C., Costa Sousa, M., Maurer, F., and Tang, A. Eliciting usable gestures for multi-display environments. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces*, ITS '12, ACM (New York, NY, USA, 2012), 41–50.

38. Sodhi, R., Benko, H., and Wilson, A. Lightguide: Projected visualizations for hand movement guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, ACM (New York, NY, USA, 2012), 179–188.

39. Vatavu, R.-D. User-defined gestures for free-hand tv control. In *Proceedings of the 10th European Conference on Interactive Tv and Video*, EuroiTV '12, ACM (2012), 45–48.

40. Vatavu, R. D., and Pentiuc, S. G. Multi-level representation of gesture as command for human computer interaction. *Computing and Informatics 27*, 6 (2008), 837–851.

41. von Laban, R. *The mastery of movement*. Dance Books, Binsted, Hampshire, UK, 2011.

42. Wilson, M., Mackay, W., Chi, E., Bernstein, M., and Nichols, J. Replichi sig: From a panel to a new submission venue for replication. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, ACM (New York, NY, USA, 2012), 1185–1188.

43. Wobbrock, J. O., Aung, H. H., Rothrock, B., and Myers, B. A. Maximizing the guessability of symbolic input. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, ACM (New York, NY, USA, 2005), 1869–1872.

44. Wobbrock, J. O., Morris, M. R., and Wilson, A. D. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, ACM (New York, NY, USA, 2009), 1083–1092.

45. Wobbrock, J. O., Wilson, A. D., and Li, Y. Gestures without libraries, toolkits or training: A $1 recognizer for user interface prototypes. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, UIST '07, ACM (2007), 159–168.

46. Zhai, S., Kristensson, P. O., Appert, C., Andersen, T. H., and Cao, X. Foundational issues in touch-surface stroke gesture design—an integrative review. *Foundations and Trends in Human-Computer Interaction 5*, 2 (2012), 97–205.