Coding Varied Behavior Types Using the Crowd

Jinyeong Yim
Mechanical Engineering
University of Michigan
jyyim@umich.edu

Winnie Leung
HCI Institute
Carnegie Mellon University
winniel@andrew.cmu.edu

Jeel Jasani
HCI Institute
Carnegie Mellon University
jjasani@andrew.cmu.edu

Ellen Lim
HCI Institute
Carnegie Mellon University
eslim@andrew.cmu.edu

Aubrey M. Henderson
HCI Institute
Carnegie Mellon University
aubreyh@cmu.edu

Mitchell Gordon
Computer Science Dept.
University of Rochester
m.gordon@rochester.edu

Danai Koutra
Computer Science & Eng.
University of Michigan
dkoutra@umich.edu

Jeffrey P. Bigham
HCI & LT Institutes
Carnegie Mellon University
jbigham@cs.cmu.edu

Steven P. Dow
Cognitive Science Dept.
UC San Diego
spdow@ucsd.edu

Walter S. Lasecki
Computer Science & Eng.
University of Michigan
wlasecki@umich.edu

CSCW '16 Companion, February 27–March 02, 2016, San Francisco, CA, USA
ACM 978-1-4503-3950-6/16/02.
http://dx.doi.org/10.1145/2818052.2874339

Abstract

Social science researchers spend significant time annotating behavioral events in video data in order to quantitatively assess interactions [2]. These behavioral events may be instantaneous changes, continuous actions that span unbounded periods of time, or behaviors that would be best described by severity or other scalar ratings. The complexity of these judgments, coupled with the time and effort required to meticulously assess video, results in a training and evaluation process that can take days or weeks. Computational analysis of video data is still limited due to the challenges introduced by objective interpretation and varied contexts. Glance [4] introduced a means of leveraging human intelligence by recruiting crowds of paid online workers to accurately analyze hours of video data in a matter of minutes. This approach has been shown to expedite work in human-centered fields, as well as generate training data for automated recognition systems. In this paper, we describe an interactive demonstration of an improved, more expressive version of Glance that expands the initial set of supported annotation formats (e.g. time range, classification, etc.) from one to nine. Worker interfaces for each of these options are dynamically generated, along with tutorials, based on the analyst's question. These new features allow analysts to acquire more specific information about events in video datasets.

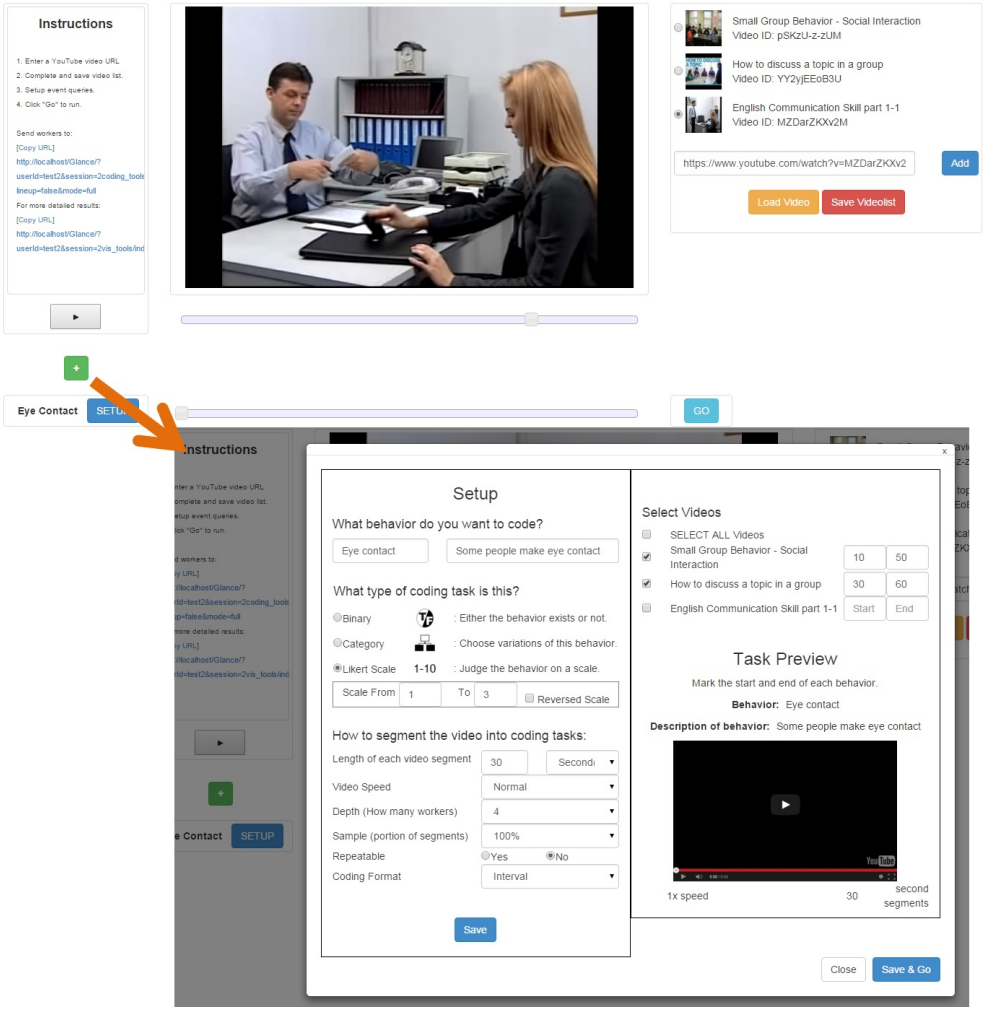

Figure 1: An example screenshot of the analyst user interface. Analysts can add events by loading the modal setup wizard to setup and update events.

Author Keywords

Behavioral video coding; Video analysis; Crowdsourcing

ACM Classification Keywords

H.5.m [Information interfaces and presentation]: Misc.

Glance

Glance is a web-based application that recruits and coordinates online crowds to complete behavioral video coding tasks. It divides video data in order to distribute it to multiple workers in parallel, and then displays their aggregated responses [4]. It was also shown that, in many cases, Glance can preserve identities of participants without significant loss of accuracy, allowing researchers to understand and control privacy risks during analysis [5].

The Glance system consists of three components: the analyst interface, the worker interface, and the merging server. In this work, we present an updated version of the original Glance system, which allows analysts to obtain a more detailed evaluation of their data sets, such as categorizations of event types, or event intensities.

Analyst Interface

The analyst interface (Figure 1) allows analysts to curate a video list to be analyzed, ask a natural language question, define an example, and set parameters for their research query. First, analysts post their video content to YouTube where it can be retrieved and displayed in Glance's embedded video player. Once the video list is complete, analysts can ask about events in the video by defining natural language queries along with simple parameters (e.g. how fine-grained the video division should be), and select the set of videos to include in the analysis. While the original version of Glance was limited to only annotating binary events with definite time spans, our current version allows analysts to request more diverse information along two main axes.

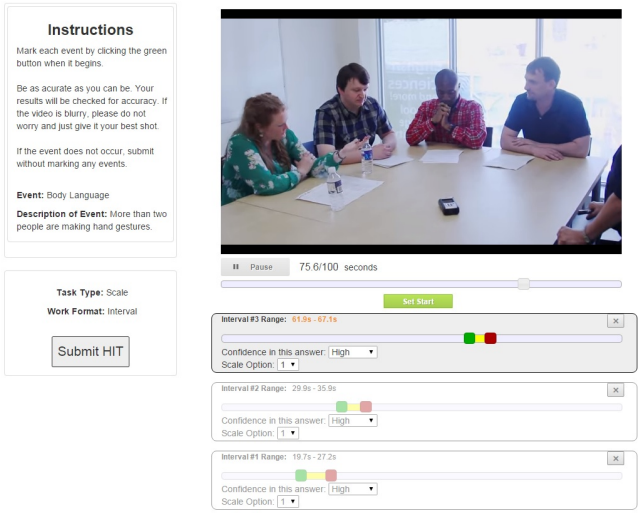

Figure 2: An example of the crowd worker interface. In this example, workers are asked to annotate the 'scale' of an specified event over 'intervals' of time. As new instances of the event are identified, additional intervals are marked and new interval bars are added (in a vertical stack below the video player).

- Task Type
  There are three task types: Binary, Category, and Scale. According to the task type, workers may provide additional information on their markings. "Binary" (the task type used in the original Glance system) is an annotation type that denotes whether or not an event happened. Thus, no additional actions are required from workers after event marking. However, the other two types require workers to provide additional information for each marked event. "Category" describes the type of event witnessed, where event types are taken from a set of analyst-defined options. For example, if a worker is asked to mark when a person in a video showed an emotional response, the category options may include 'angry', 'happy', 'sad', etc. Lastly, "Scale" determines the intensity of an event. For example, if workers are asked to mark laughter, a scalar value of 1 to 5 may be used to determine if people were just smiling, or laughing heartily (kappa scores are shown to indicate worker agreement on subjective events). This more detailed information will help analysts gather more meaningful insights into their data.

- Temporal Format
  The temporal format describes how event times in video clips are marked, allowing more expressive representation of temporal knowledge [1]. Analysts determine which temporal format to use for each event identified. Glance allows three temporal formats: Interval, Occurrence, and Whole Segment. In "Interval" mode, workers define the time span that the event occurred by specifying starting and ending points. In "Occurrence" mode, workers determine the instant that the event occurred in. In "Whole Segment" mode, workers provide a binary judgment of whether or not the event occurred anywhere within a video clip. These different temporal formats allow both analysts and workers to more precisely annotate the temporal location of a specified event.

An analyst's query is sent to multiple workers who mark event occurrences based on the analyst-determined temporal format using the worker interface. As results are returned from workers, they are displayed in an event result bar below the playback controls of the video. Since this result bar is aligned with the embedded video play bar, analysts can easily examine the results in the context of the source video.

Worker Interface
Crowd workers are initially given instructions and a simple tutorial that provides an overview of the worker interface.

Once workers have completed the tutorial, the system places them in a retainer pool until a task is available. When an analyst query arrives, workers are routed to a task page (Figure 2) that is dynamically generated based on the analyst's configuration of the query. Once the workers have completed the video clip, they can submit their results.

Demonstration

When conference attendees arrive at our demonstration, they will be able to select from a predefined set of videos that range in content from dyadic interaction in a controlled setting to videos of informal settings with large groups. All of the videos used will have Creative Commons licenses. When an attendee wants to ask the system a question such as: "When did people in any of the videos laugh?", she selects a query with the matching text, which was set up to annotate whether the event ('binary') occurred ('occurrence') in any 30-second clip from the video. While Glance produces results within minutes instead of days or weeks, this latency may still be too high for attendees who wish to see immediate results. Thus we present a pre-collected set of responses to the available queries, which will displayed in the same manner as if they were newly collected (including an annotation of how long the results took to arrive when first collected).

In our example, the attendee next becomes curious about "how loud and how long people were laughing". She asks the system a follow-up question regarding the 'scale' and 'interval' of laughing events. The follow-up results are displayed as an additional annotation of the same previously-marked segments from her first query, which demonstrates how Glance can save worker effort by scoping subsequent queries. Satisfied that Glance can help her research, our hypothetical attendee makes her way back to the conference with a link to download Glance for use on her own projects.

Conclusion

We will demonstrate a new, more expressive version of Glance that can quickly and reliably code multiple types of events in video data using online crowds. This new functionality provides task type and temporal format options for coding data, which allows sociotechnical researchers to analyze more complex behaviors, and obtain more detailed information about their data. Over time, Glance can also help the train automated visual analytics tool in real settings [3].

References

Allen, J. F. Maintaining knowledge about temporal intervals. *Communications of the ACM 26*, 11 (1983), 832–843.

Heyman, R. E., Lorber, M. F., Eddy, J. M., West, T., Reis, E. H. T., and Judd, C. M. *Handbook of Research Methods in Social and Personality Psychology*. Cambridge University Press, 2014, ch. Behavioral observation and coding.

Laput, G., Lasecki, W. S., Wiese, J., Xiao, R., Bigham, J. P., and Harrison, C. Zensors: Adaptive, rapidly deployable, human-intelligent sensor feeds. In *CHI* (2015), 1935–1944.

Lasecki, W. S., Gordon, M., Koutra, D., Jung, M. F., Dow, S. P., and Bigham, J. P. Glance: Rapidly coding behavioral video with the crowd. In *UIST* (2014), 551–562.

Lasecki, W. S., Gordon, M., Leung, W., Lim, E., Bigham, J. P., and Dow, S. P. Exploring privacy and accuracy trade-offs in crowdsourced behavioral video coding. In *CHI* (2015), 1945–1954.