Instructor-operator Video Dataset for Two Arduino Tasks

Keng-Teck Ma
Artificial Intelligence Initiative, A*STAR, Singapore
makt@scei.a-star.edu.sg

Mark Rice
Institute for Infocomm Research, A*STAR, Singapore
mdrice@i2r.a-star.edu.sg

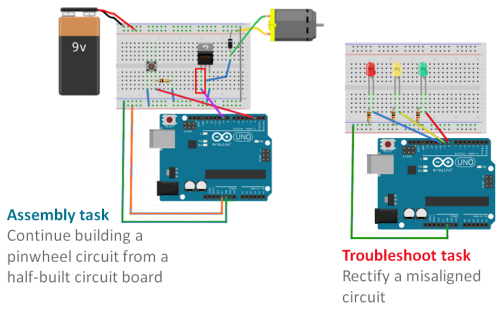

Figure 1. Arduino tasks.

Abstract

We present the first instructor-operator video dataset for two arduino tasks: pinwheel construction and traffic-light troubleshooting. The dataset contains about 50 hours of video recordings from 3 conditions with a total of 90 participants.

1. Introduction

Despite the rising interests in automatic instructional video understanding, there are very few datasets for comparative experiments. One large gap is the scenario where one participant is helping another to complete some task. In this paper, we describe and share a new dataset which has instructor-operator pairs completing two Arduino tasks under varying conditions, see Figure 1. This data was collected for the purpose of studying the human factors affecting remote assistance with mobile devices. Our main results were published in [2, 4]. We now release this dataset to accelerate research for instructional video processing and understanding.

In our first study. we compared the differences between three levels of Augmented Intelligence (low, medium, and high) with the *Wizard-of-Oz* approach [1]. In our follow-up study, we evaluated two scenarios: (1) co-located paper-based assistance vs remote assistance with a mobile device; (2) hand-held vs head-mounted interaction. The dataset contains the video recordings from a stationary camera which record the operator. There are also corresponding video recordings of the instructors in scenarios where there is also participants who were performing the instructor role. See Figure 2 for examples.

1.1. Background

As part of a larger project, REVIVE, we investigated various human factors issues in research and development of Augmented Intelligence on mobile devices. For our study, we used an in-house Remote Assistance Platform (RAP) which was deployed using a Kurento Media Sever, with Web Real-Time Communication (WebRTC) [3]. In our comparative studies, we also had participants perform tasks via a side-by-side scenario.

2. Setup

This section briefly describes the experimental setups. Please refer to our full papers for more details [2, 4].

2.1. Tasks

Arduino platform was chosen for its versatility to combine hardware and software features, as well as relative ease to learn the skills needed to complete simple tasks. Microcontrollers, electrical components and a laptop were used by the participants. Each task was to be completed within 20 minutes.

Pinwheel construction. The operator built a pinwheel circuit. This task took 33 sub-steps, and involved identifying and connecting components to form a complete circuit. If completed successfully, the pinwheel span when laptop software was run.

Condition	Features
Low AI	Video playback in the form of full videos
Medium AI	Video playback segmented into steps Search/retrieval - request for specific videos *AI cannot view live stream from operator's device. *Prompts are used if questions can be answered without viewing the live stream.
High AI	Video playback segmented into steps. Search/retrieval - request for specific videos. Active monitoring/intervention - actively monitors operator's actions and intervenes with prompts whenever a mistake is detected. *AI can view the live stream from operator's device. *Prompts are also used whenever questions are asked by the operator.

Table 1. Conditions in the AmI Scenario.

Traffic light troubleshooting. The operator troubleshooted a misaligned traffic light circuit. This task took 22 sub-steps, and if completed successfully, LEDs lit-up in a traffic light sequence (green, amber, red, repeat).

2.2. Scenarios

The three scenarios are briefly described below.

Co-located Scenario. A total of 20 participants (10 males and 10 females) with a mean age of 21 years were recruited. They were randomly assigned to be either an instructor or operator. Each instructor was given 15 minutes to familiarise themselves with the tasks. Each pair was given 20 minutes to complete each task. Pairs were seated next to one another. The instructor was able to give verbal instructions and pointing gestures to the operator, but was not allowed to physically demonstrate or manipulate the objects.

Remote Scenario. A total of 40 participants (25 males and 15 females) with a mean age of 29 years were recruited. Pairs were randomly assigned to be in the "hand-held display" or "head-mounted display" condition.

AmI Scenario. A total of 30 participants (15 males and 15 females), with a mean age of 21 years were recruited. They were randomly and uniformly assigned to one of the levels (10 participants per condition). See Table 1. None of the participants had prior experience of the tasks. In this scenario, the instructor's role was performed by an experimenter who was very familiar with the tasks (Wizard of Oz).

3. Inventory

There are a total of 169 videos for a total duration of 50 hours and 13 minutes duration. Each video is on average of length 17 minutes and 50 seconds. Some single sessions were split into multiple clips due technical issues during the recording. Ninety-nine of the videos have the resolution of 640 by 480, while the remaining seven-six are of 2160 by 1440 resolution. There are ninety-three operator-only videos, fifty-nine instructor-only videos and seventeen instructor-operator pairs videos.

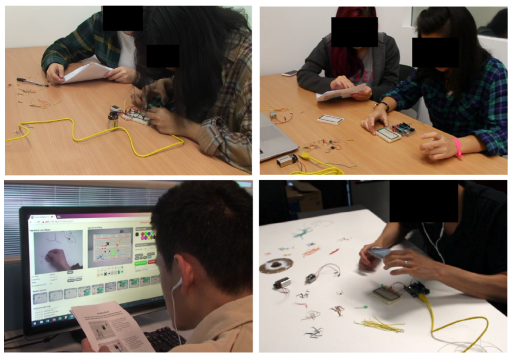

Figure 2. Seperate screenshots from our dataset. Top: Co-located Scenario. Bottom: Remote Scenario.

4. Conclusion and Acknowledgement

In this paper, we present a novel instructor-operator videos dataset for two Arduino tasks. This is the first known dataset to have participants work together for instructional purposes with three different scenarios of the same set of tasks. We hope that this dataset will facilitate discussion on new approaches for instructional videos understanding, including related interaction challanges.

This work was supported by the Reverse Engineering Visual Intelligence for cognitive Enhancement (REVIVE) programme (1335h00098) funded by the Joint Council Office of A*STAR. We also thank Ms Wei Ling Koh, Ms Joyce Kaliappan and Dr Wah Pheow Tan for their contributions in the dataset collection and analysis.

References

[1] N. Dahlbäck, A. Jönsson, and L. Ahrenberg. Wizard of oz studies - why and how. *Knowledge-based systems*, 6(4):258–266, 1993. 1

[2] W. L. Koh, J. Kaliappan, M. Rice, K.-T. Ma, H. H. Tay, and W. P. Tan. Preliminary investigation of augmented intelligence for remote assistance using a wearable display. In *Region 10 Conference, TENCON IEEE*, pages 2093–2098. IEEE, 2017. 1

[3] M. Rice, S. C. Chia, H. H. Tay, M. Wan, L. Li, J. Ng, and J. H. Lim. Exploring the use of visual annotations in a remote assistance platform. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1295–1300. ACM, 2016. 1

[4] M. Rice, K.-T. Ma, H. H. Tay, J. Kaliappan, W. L. Koh, W. P. Tan, and J. Ng. Evaluating an augmented remote assistance platform to support industrial applications. In *4th World Forum on Internet of Things (WF-IoT) IEEE*. IEEE, 2018. 1