



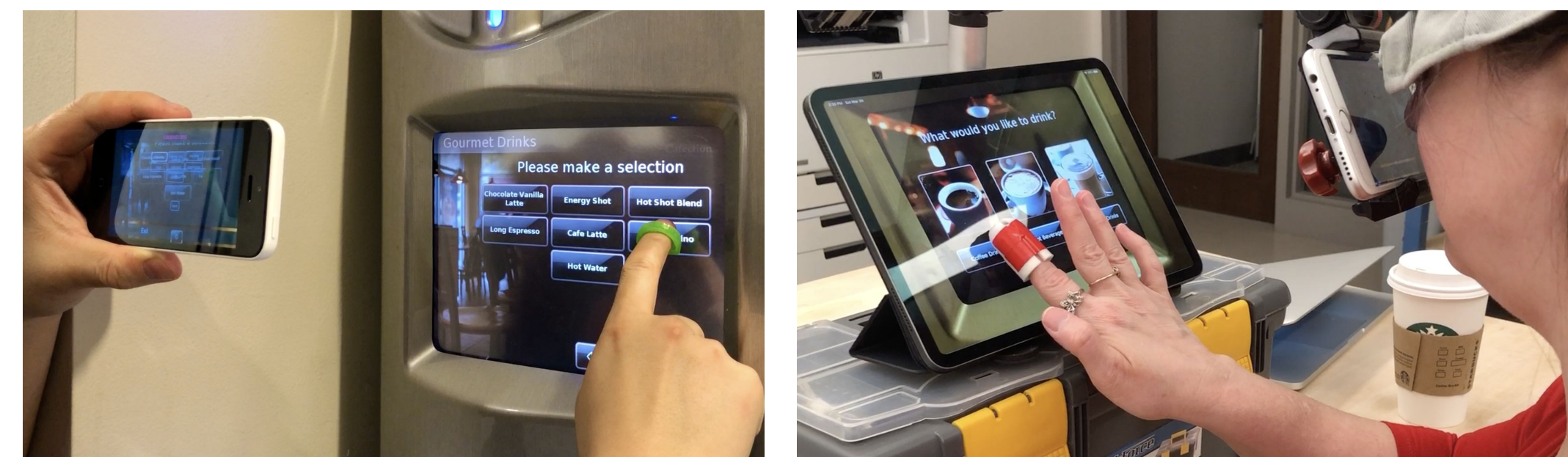
StateLens: A Reverse Engineering Solution for Making Existing Dynamic Touchscreens Accessible



Anhong Guo, Junhan Kong, Michael Rivera, Frank F. Xu, Jeffrey P. Bigham

INTRODUCTION

Blind people frequently encounter inaccessible dynamic touchscreens in their everyday lives. To solve these problems, we introduce StateLens: a reverse engineering solution that makes existing dynamic touchscreens accessible. First, StateLens reverse engineers the underlying state diagrams of existing interfaces from videos using a hybrid crowd-computer vision pipeline. Second, StateLens automatically generates conversational agents to guide blind users through specifying the tasks, allowing the iOS application to provide interactive guidance and feedback for blind users to access the interface. Finally, a set of 3D-printed accessories blind users to explore capacitive touchscreens without risk of triggering accidental touches. We ran both technical evaluations and user study to demonstrate the effectiveness of StateLens.



USER AGENT AND ACCESSORIES

Conversation Agent Example - Coffee Machine

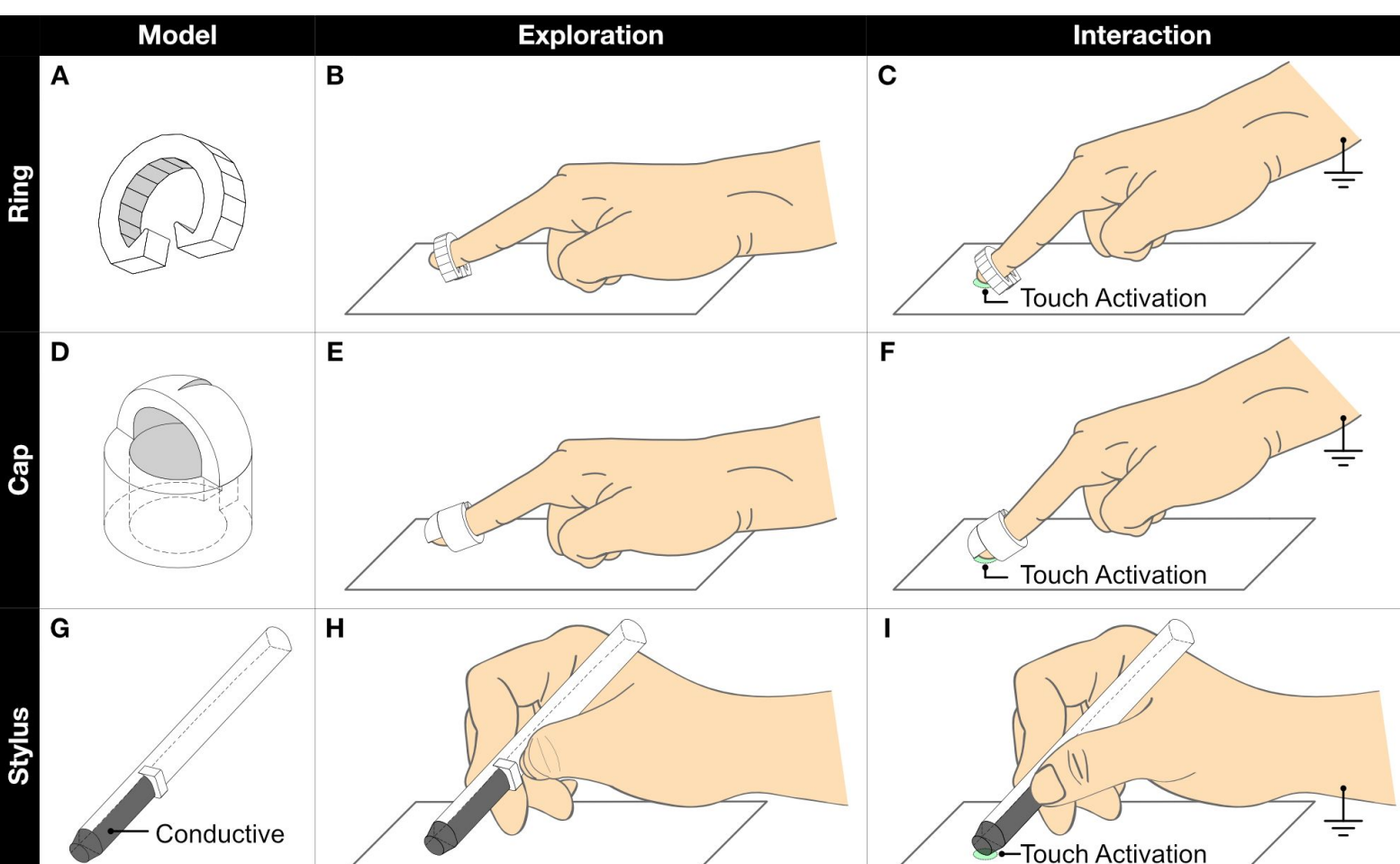
- Select what would you like to drink from coffee drinks, hot beverages, and gourmet drinks.
- Can I get a summary?
- You can say: "I want large cappuccino".
- I want a **large coffee 50-50**.
- Select **strength** from **mild, regular and strong**.
- Strong**.
- You want **large strong coffee 50-50**, is that right?
- Yes**.
- Gotcha. I will help you out!
- 1 Welcome message from the initial state
 - 2 Summary by aggregation
 - 3 Parse required parameters: `size = large`
`coffee_type = coffee 50-50`
 - 4 Prompt missing parameter: `strength = ?`
 - 5 Ask for confirmation
 - 6 Proceed to guidance

Figure 3: Conversational Agent Example

The agent can be automatically generated from the state diagram of the interface and allows users to specify the task to perform.

Figure 4: Design of accessories

The 3D-printed accessories allow users to explore touchscreen interfaces without accidentally triggering touches and then touch when they need to.



SYSTEM WORKFLOW

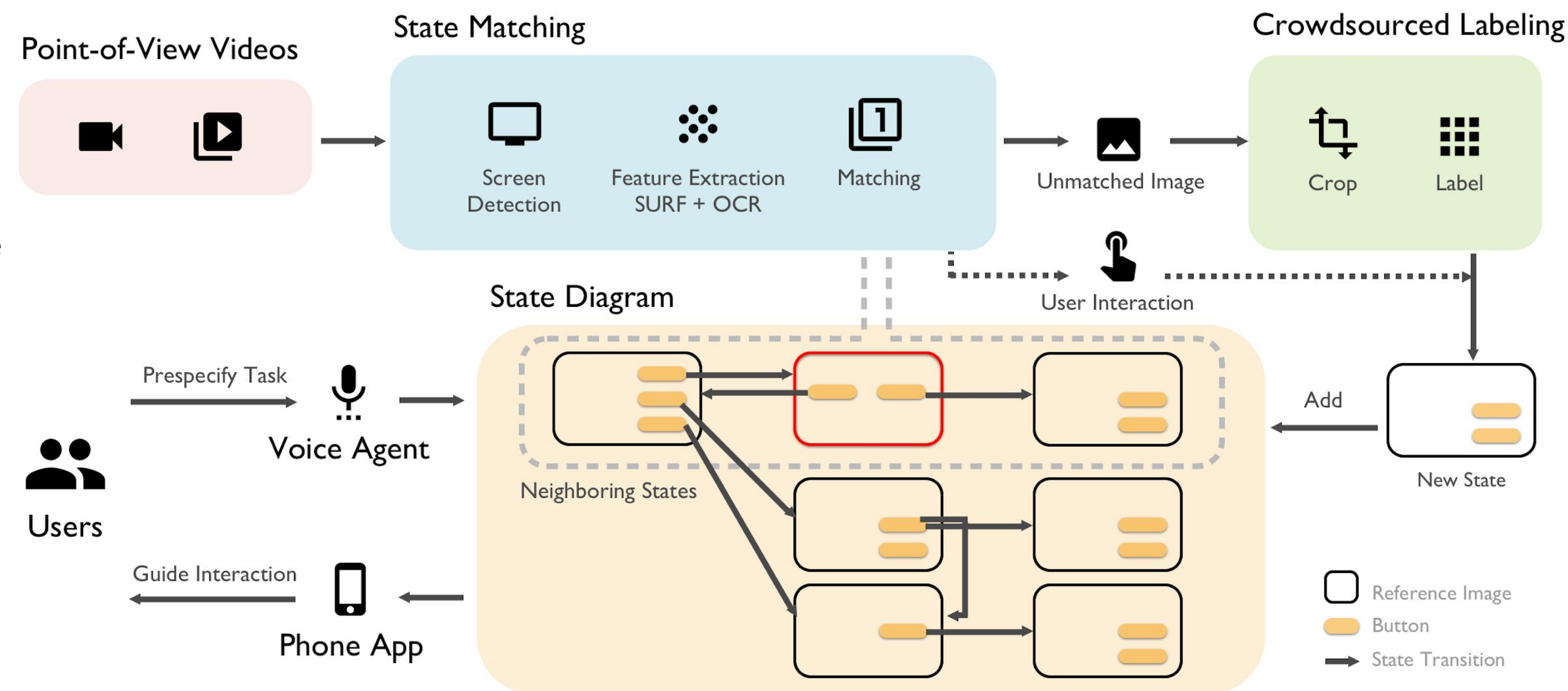


Figure 1: System workflow

Constructing state diagram

StateLens uses a combination of Screen Detection, SURF and OCR to match the current video frame with reference states and decide whether to send to crowdsourcing pipeline and add it as new state.

Generating path from user-specified task

With user-specified task through conversational agent, StateLens generates a target (state, button) sequence to achieve the task.

Identifying current screen

StateLens uses SURF combined with detected user interaction points from camera to identify which state the user is currently on.

Providing guidance to complete task

Based on the detected state, StateLens reads out screen description and guide users to move their finger towards the target button.

Improvements on Speed and Robustness

→ Narrowing down search space to speed up matching

StateLens searches the neighboring states of the detected state before global search and speeds up the matching process for complicated interfaces.

→ Using fingertip location to infer next screens

StateLens uses the detected fingertip location on the previous screen to infer which button the user might have clicked on and help identify the current state.

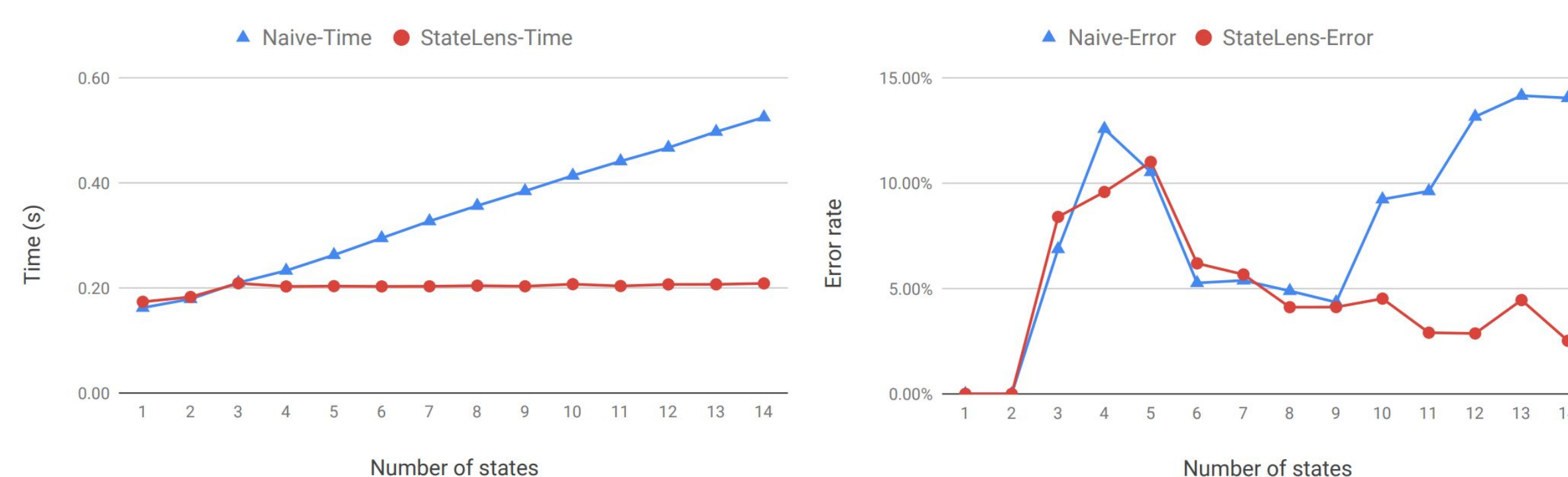


Figure 2: Example state diagram of coffee machine interface

RESULTS

In our technical evaluation, StateLens achieved a relatively stable **error rate of ~5%** in identifying the current state. The matching time is also fairly stable for different sizes of interfaces around **0.2 seconds**.

In our user study, users spent an average of **110.1 seconds** to complete a task and the overall task completion rate was **94.7%**. In subjective rating, users reported the StateLens app to be *easy to learn* (5.5/7), *comfortable to use* (5.6/7), and *very useful* (6.1/7). They also found the audio feedback was *in real-time* and *accurate* (6.1/7).



	Coffee Machine (G)		Coffee Machine (T)		ATM		Printer		Projector Control											
	Stationary	Hand-held	Stationary	Hand-held	Stationary	Hand-held	Stationary	Hand-held	Stationary	Hand-held										
# of states	14	13	11	10	11	12	10	10	13	9										
# of frames	4,980	2,580	3,060	2,310	2,910	2,340	1,380	1,980	3,540	1,530										
SURF	0.47	0.57	0.67	0.62	0.88	0.64	0.78	0.70	0.85	1.00	0.82	0.90	1.00	0.46	0.50	0.56				
SD	0.58	0.50	0.73	0.62	1.00	0.64	0.80	0.80	0.69	1.00	0.63	1.00	1.00	0.20	0.82	0.90	0.86	0.46	0.63	0.56
+SURF	0.54	0.67	0.78	0.80	0.81	0.77	0.33	0.86	0.60	0.59										
SURF +OCR	0.72	0.93	0.65	0.85	0.73	1.00	0.67	1.00	1.00	1.00	0.75	1.00	1.00	0.40	0.63	1.00	1.00	0.54	0.67	0.67
SD+OCR	0.81	0.73	0.85	0.80	1.00	0.86	0.57	0.77	0.70	0.67										
SD+SURF +OCR	1.00	0.93	0.85	0.85	0.91	0.91	0.77	1.00	1.00	1.00	0.75	1.00	0.86	0.60	0.63	1.00	1.00	0.62	0.70	0.78
	0.96	0.85	0.91	0.87	1.00	0.86	0.71	0.77	0.76	0.74										
	Room Reservation		Treadmill		Ticket Kiosk		Coca-Cola		Subway		Washer		Car Control							
	Stationary	Hand-held	Stationary	Hand-held	Stationary	Hand-held	Web	Web	Web	Web	Web	Web	Web							
# of states	7	8	10	10	11	14	9	16	11	24										
# of frames	1,560	1,260	1,260	4,500	1,470	3,480	2,100	6,630	5,940	17,940										
SURF	0.83	0.71	0.33	1.00	1.00	0.10	1.00	0.80	0.47	0.73	0.53	0.57	0.23	1.00	0.48	0.94	0.46	0.55	0.79	0.46
SD	0.77	0.50	0.18	0.89	0.57	0.37	0.64	0.50	0.58	0.58										
+SURF	0.86	0.86	0.50	0.75	1.00	0.10	1.00	0.50	0.58	0.64	0.64	0.50	0.33	0.67	0.71	0.94	0.78	0.64	0.73	0.67
SURF +OCR	0.54	1.00	0.39	0.88	0.75	0.60	0.83	0.50	0.50	0.73	0.60	0.64	0.20	0.89	0.47	1.00	0.53	0.82	0.65	0.83
SD+OCR	0.70	0.54	0.67	0.63	0.59	0.62	0.33	0.64	0.64	0.73										
SD+SURF +OCR	0.78	1.00	0.47	1.00	0.68	0.70	0.83	0.50	0.50	0.73	0.65	0.79	0.40	0.67	0.65	0.94	0.75	0.82	0.73	0.92
	0.88	0.64	0.64	0.63	0.59	0.71	0.50	0.77	0.78	0.81										

Figure 5: State diagram construction accuracy by interface

Figure 6: Matching scalability and robustness